

Applying Learning Analytics to a Primary School Classroom: Benefits and Barriers.

M.J. Rodríguez-Triana
REACT
Ecole Polytechnique Fédérale
de Lausanne (EPFL)
Lausanne, Switzerland
maria.rodrigueztriana@epfl.ch

A. Martínez-Monés
GSIC-EMIC
Dpto. Informática
Universidad de Valladolid
Valladolid, Spain
amartine@infor.uva.es

S. Villagrà-Sobrino
GSIC-EMIC
Dpto. Pedagogía
Universidad de Valladolid
Valladolid, Spain
sarena@pdg.uva.es

ABSTRACT

There are numerous proposals and solutions that aim to apply Learning Analytics (LA) in authentic scenarios. Many of them have been applied successfully in learning contexts where adults are involved. However, when we try to transfer such proposals to underage students, the application may not be straightforward. This paper reports on a case where a LA approach, which had been already applied to a university context with success, was applied to a primary school classroom. The case shows how the main barriers that had to be faced in the new educational context were legal and ethical issues related to identity and data ownership. The case also serves to illustrate the potential benefits that an apparently simple LA approach can bring to these educational contexts.

Categories and Subject Descriptors

K.4.1 [Computers and Society]: Public Policy Issues—*Ethics, privacy*

General Terms

Experimentation, Legal aspects

Keywords

Learning Analytics, Privacy, Ethics, Primary Education

1. INTRODUCTION

The increasing trend towards the massive data collection in educational settings has raised new ethical and privacy concerns in the Learning Analytics (LA) research community. On the one hand, there is a need for identifying the students across platforms and retrieving as much data as possible to obtain informed analysis about the learning process. On the other hand, multiple constraints should be taken into account such as: who are the owners of these data (ownership), whether they allow to use such data or not (informed

consent and privacy), what are the purposes of the analysis (transparency), who will have access to the results of the analysis (data clients), etc. [2].

In formal education, accomplishing with these two tensions within the technological context provided by the educational institutions, normally in the form of Learning Management Systems (LMS), constitutes a challenge itself [4]. In addition, since students' learning activities increasingly include resources out of the LMS, focusing the analyses on the LMS provides an incomplete view. Therefore, there is an obvious interest for including data from these external sites. However, the inclusion of data from third-party tools which are not under the institutional control, makes the situation even more complex in terms of user identification, ethics and privacy [9].

Our proposal to LA is supported by an infrastructure able to integrate different kinds of data sources, including LMS, tools, and user-generated data [5]. With this integration we aim at applying LA to the existing technological ecosystems to which the users (teachers and students) are familiar with. Our proposal also provides means to let teachers become active actors in the LA process, by including their pedagogical needs in the analysis and by letting them specify when and how receive the information [7]. These proposals are generic and applicable to different learning situations. However every educational context poses its specific requirements, that have to be taken into account when applying a solution to that context [2].

We present in this paper a case study that describes the issues we had to address when applying our LA approach, which had been successfully applied in various university settings, to a primary school classroom. The case illustrates how privacy and legal concerns can become the major issues to solve in the transition to a different educational context.

The structure of the rest of the paper is as follows. Next section describes the main aspects of our approach to LA and our previous experience applying it in several university contexts. Section 3 presents the main issues we encountered when we applied our approach to a primary school classroom, and discusses the main implications. We finish the paper with the main conclusions drawn from this work.

2. STARTING POINT: OUR APPROACH TO LA

We describe in this section the overview of our approach to LA. This description aims at providing the basic elements needed to understand the issues we found when moving to a primary school context, which are described in Section 3.

Our approach to LA is mainly devoted to support teachers in orchestration aspects such as awareness and formative assessment. To achieve that goal, we help the teacher to monitor the students throughout the learning scenario. The solution provided consists of three pillars: First, a *design process* of the learning scenario that takes into account the teacher’s information needs [7]. This means that the teacher decides in advance which information she wants to obtain from the learning context, being able to adapt the design to satisfy her monitoring needs (e.g., choosing the most appropriate tools for both pedagogical and monitoring concerns, or identifying complementary data sources). The second part of the solution is a *monitoring process* guided by the decisions made at design-time [7]. In this monitoring process, the data gathering is focused on those sources chosen by the teacher and then the data analysis pursues to verify whether the current state (the gathered evidence) matches with the desired state (the learning design). Finally, the third component of our proposal is GLUE!-CAS [5], an *architecture* that allows the data gathering in distributed learning environments (DLEs) made up by virtual learning environments and web 2.0 tools.

Even though these proposals were envisioned to be generic, they were iteratively designed and evaluated by their application to seven courses in university contexts [7]. In this section, we describe our approach in terms of the LA framework proposed by Greller and Drachler [2] highlighting some privacy and ethical aspects that emerged in the case studies. Table 2 summarizes the main topics addressed in this description following the framework.

Stakeholders. In our approach, the main stakeholders are the participants of the learning contexts and, in a second place, the researchers involved in the experiments. For example, in the studies carried out at the university, students and teachers were the data providers (data subjects), while teachers and researchers were the ones consuming the outcome from the data analysis (data clients).

Objective. Our LA approach aims at supporting teacher’s reflection, concretely providing monitoring information to be used for regulation, formative assessment, or self-reflection about the learning design and the learning process. To achieve this goal we propose to provide teachers with feedback about the accomplishment of pedagogical decisions made at design-time. Additionally, we push teachers to configure the monitoring process. Following this strategy, they can reflect on their monitoring needs and adapt their learning design to better satisfy such needs. As a consequence, teachers play a crucial role in our proposal: they are not only subjects and clients, but also conductors of the monitoring process. This condition has a clear impact on the transparency and ethics of the purposes of the analysis: the teacher is the one who decides what data are necessary, how to collect them and why.

Table 1: Overview of our approach to LA as it was implemented in university settings in terms of the framework proposed by Greller and Drachler [2].

Dimensions	Values
Stakeholders	<ul style="list-style-type: none"> - Data subjects: learners and teachers. - Data clients: teachers and researchers.
Objective	<ul style="list-style-type: none"> - Reflection: Support orchestration in formal educational settings specially in terms of awareness for regulation, formative assessment, and self-reflection about the learning design and the learning process.
Data	<ul style="list-style-type: none"> - Protected data sets: learning designs; actions registered in the DLEs (e.g., accesses, editions, uploads); ad-hoc evidence provided provided by the participants (e.g., attendance registers or students reflection about their learning process). - Relevant indicators: participation, collaboration, group formation and expected use of resources - Time scale: timeframes for data analyses determined by the teachers at design-time.
Instruments	<ul style="list-style-type: none"> - Pedagogic theory: blended CSCL - Technology: VLEs (Moodle and MediaWiki), Web 2.0 tools (e.g., Google apps) and GLUE! to build the DLEs. GLUE!-CAS and GLIMPSE for the script-aware monitoring - Presentation: tables gathering the evidence of each indicator, warnings and links to resources and data sources
External limitations	<ul style="list-style-type: none"> - Privacy: the data was used by the teacher, and anonymized for research-related publications. - Ethics: the students were informed in advance about the data to be collected and the purpose of the analysis, and they signed an informed consent. - Norms: Spanish laws 15/1999 and 1720/2007. - Time scale: during and after the enactment. The results are still accessible since it may help the teachers refine their learning designs for the next years.
Internal limitations	<p>Required competences:</p> <ul style="list-style-type: none"> - Interpretation: The teachers were able to easily interpret the results of the analysis especially because they were contextualized within the learning design, and the teachers were involved in the design of the monitoring process. - Critical thinking: Since minimalistic indicators were provided, teachers understood the results and triangulated the information obtained from the different data sources.

Data. Regarding the data used for the analysis, we propose to use blended data sources. Frequently, part of the learning process occurs out of the technological context. Besides, the Information and Communication technologies (ICTs) register a limited set of evidence, usually based on user interactions with the platforms, and sometimes it is not possible to authenticate the student identity properly [9]. At a university context, students are adults and have their own virtual identity (if not many). The institution provides them an account, which is known by the teachers. These accounts can be used for handling the automatic opening of accounts in new LMS (such as wikis or Moodle) and setting up the environment where they will interact for their course, and where the data will be collected.

Since restricting the analysis to the data registered by the ICT tools may offer just a partial view of the user activity [1], we recommend to include also ad-hoc information provided by the participants of the learning context. This practice enriches the evidence gathered and allows teachers to triangulate the data coming from the different data sources. For example, this approach helps to alleviate the problem of “enmeshed identities” [2], i.e., fingerprints linked to one user logged in the system but executed by another user or group, a typical problem in collaborative tasks or when students are sharing devices.

From the ethical point of view, this approach entitles teachers and students to rectify the data automatically collected from the ICT tools [8] and to better understand the reasons behind the results obtained [2]. In terms of privacy, this involvement of the stakeholders has also evident benefits, letting them decide about what information they want to share [9].

It should be noticed that, in addition to the data generated on account of the learning process, our proposal also employs the teachers’ learning designs for contextualizing the data analysis. Based on these data and reducing the analyses to the timeframes specified by the teachers at design-time, we obtain simple indicators related to aspects such as participation, collaboration, group formation and expected use of resources.

Instruments. In relation to the pedagogical approach, despite our proposal aims to be generic, our studies at the university involved face-to-face and distance activities that were carried out at different social levels (individual, group and whole class). In other words, blended Computer-Supported Collaborative Learning (CSCL). In addition, the learning scenarios were supported in all cases by DLEs. These DLEs were made up by an institutional VLE (typically Moodle or Mediawiki), Web 2.0 tools (e.g., Google applications), and GLUE!¹, an architecture devoted to integrate third-party tools in learning environments, specially in VLEs. This infrastructure was set up using the resources available in our lab, which gave us full control over the data collected from the servers. An exception to this rule are the third-party tools, which do not always provide data, or do not allow a proper identification of their users [3].

To overcome the problem of the integration of all the aforementioned data sources, we proposed an architecture, GLUE!-CAS [5], which collects the data offered by the platforms and provided ad-hoc by the participants through ICT tools (e.g., on-line questionnaires or spreadsheets). Although the teachers have access to these data and could collect and analyse them by themselves, it would be a very demanding task especially in terms of time. To support teachers in this endeavour, we implemented GLIMPSE [6], a tool that, interacting with GLUE!-CAS, automatizes data gathering, integration and analyses, offering the teacher a comparison between the current and desired state of the learning scenario. The information provided by the different data sources is collected, integrated and stored in the device where GLIMPSE is executed, creating a centralized and integrated copy of the data available. The outcome of this tool is a monitoring report where the information is visualized by means of tables that connect the participants, the data sources, the indicators and the warnings that emerged from the analysis.

External limitations. To ensure transparency in our experiments, the teacher and the research team explained the students what was the purpose of the data analysis and what kind of data was going to be used. Before the experiments started, the students signed an informed consent where, on the one hand, they allow us to collect their data under the aforementioned conditions and, on the other hand, the research team promised to ensure their anonymity out of the course settings.

In terms of privacy, as other proposals also suggest [10], this schema relies on the teacher, establishing a parallelism with real classrooms, where the teacher has access to sensitive data and decides how to use it.

Internal limitations. Dealing with the *internal limitations*, the teachers did not require specific training in order to interpret the monitoring results. The use of very simple indicators facilitated the comprehension. Besides, the fact that the teachers had configured the monitoring process and the results were connected with the learning design helped them to contextualize the results. Additionally, the combination of multiple data sources -technologies, teachers and students- enriched the evidence, allowed the teacher to contrast and compare them, and contributed to better understand the learning process.

3. MEETING POINT: APPLYING OUR APPROACH TO A PRIMARY SCHOOL CONTEXT

The case study in which we are focusing was run at a first grade class (K6-7) with 24 students at a school sited in a rural area in Valladolid, Spain. The teacher in charge of this class was using blogs in his classes, wherein he integrated external Web 2.0 tools (such as Youtube videos or Educaplay games). He employed Google’s Blogger service to edit these blogs. In spite of a general positive experience with this approach, one of the main worries of the teacher was that these blogs did not inform about reading accesses to their pages. This precluded him from knowing whether the students were accessing the blog pages and if so, at which moments of the week.

¹GLUE! - Group Learning Uniform Environment: <http://www.gsic.uva.es/glue>. Last visit: January 2015

In order to face this problem, we proposed to set up a learning analytics module based on the approach described in the previous section.

Following the monitoring-aware design process, the teacher proposed two lesson plans, involving a blog and several external resources, in which the teacher asked the students to carry out some activities from their homes, (e.g., watching a video, read an on-line text, playing games, etc.). We set up the infrastructure that would intercept the blog accesses of the students, so that they could be traced. These accesses were the input to the learning analytics module, which was in charge of providing the teacher with monitoring reports informing him about the activity of his students. These reports would be issued to the teacher at specific moments he had decided during the design phase.

In our way to achieve this goal we faced several problems, which are described below, based on the data collected by an external observer's notes to the face-to-face sessions, two interviews to the teacher (before and after the intervention took place) and the teacher's self-reflections in his diary.

Firstly, and as expected when working with minors, we had to ask for the consent of the principal of the school and of the legal representatives of the children, as they are not adults. Both the principal and the families supported this kind of innovative practices, and obtaining these permissions was not a difficult task. However, we have to note that obtaining permission when working with minors is a very delicate issue and has to be carefully thought and carried out.

The second issue we faced was related to the students' virtual identities. As noted beforehand, the participant teacher was using Blogger to create the blogs. Blogger had suited his needs so far, being widely available, easy to use and free. However, the blog he had created was open for anonymous access, and the teacher was not able to know who had used it or not. In order to get this information, a LA tool needs users to be identified. Blogger can be configured so that only registered users access the blogs, but this requires to have a Gmail account. However, children are not legally allowed to own an email account. In order to overcome this obstacle, we had to look for a turnaround, asking the families to open an email account on behalf of their kids, and access the site using this account. This turned out to be problematic for a number of reasons. As noted by the teacher in the second interview, not all the families knew how to solve some technical difficulties, and some of the families accessed the site using the parents' own accounts.

“When I had the meeting with the families, many of them said to me that they have had a lot of problems to access to the on line activities embedded in the blog. They said that they had read and followed all the indications given by me, but finally it was impossible. Moreover I have noticed when I have reviewed the monitoring report, that there were families that had accessed to the activities with a different e-mail account.” [Teacher interview. 3/07/2014]

As it can be seen, the technical difficulties met by the families resulted in a case of “enmeshed identities” [2]. Aware of this problem, in the final interview, the teacher expressed that if he was to apply a similar activity in the future he would devote more time with the families to provide training to access to the on-line activities with the e-mail logging, as well as supervise their accesses. The implication here is that when working at these educational levels, where kids are not autonomous in their use of technology, families are another actor that has to be taken into account. Issues as privacy, transparency and control over the data are prone to be an issue when the families are aware that the tasks done at home are being analysed and used by the teachers. The experience in this case points to the idea that families should be involved as active participants in order to face these issues satisfactorily.

The third issue we found is related to the expected outcomes of LA. In upper educational levels the input can be very rich, enabling complex analysis. On the contrary, with first graders (6-7 year old children), the kind of data available is prone to be very simple, maybe only consisting of accesses to a page (as these children hardly know how to write with the computer). For this reason, the kind of information being offered to the teacher was based on a rather simple indicator (number of accesses to the blog and to the embedded tools).

In spite of this simplicity and all the restrictions related to the quality of the data, the participant teacher was very positive about the obtained feedback:

“it is important to provide students with other ways to learn, as well as to assess learning. I can say that these teaching and learning practices are innovative for students and we have the opportunity to motivate them and engage the families in the educational process”. [Teacher's reflective diary]

The teacher identified multiple potential usages for the learning analytics module, such as knowing in advance whether the students have done their homework, being able to send reminders, identify more/less attractive resources, analyse the students' evolution against their personal work, etc. Therefore, the teacher was able to understand the results, and was able to identify potential uses of the analysis.

4. CONCLUSIONS

The case described in this paper has illustrated that legal and ethical issues related to data control and data ownership can become the main issues to solve when applying a LA approach to a primary school context.

These results point to the idea that LA providers must be aware of these legal and ethical issues and incorporate them into the system's design. This finding is aligned with the “ethics by design” proposals.

A positive result from the case study is that, once the legal and identity difficulties were solved, the teacher was able to appropriate the ideas and think on different potential

uses of the information provided to him by the LA tool. On the contrary, the cloud-based tools that are becoming widespread at these educational levels, do not cover these information needs, and may pose legal and ethical problems related to data ownership and virtual identity, difficult to solve when working with minors. Not only LA, but also the wider technology-enhanced learning community have a big challenge in providing appropriate tools to these educational levels. As mentioned above, these tools must be able to handle the legal and data issues required when working at this educational level.

5. ACKNOWLEDGMENTS

This research has been partially funded by the European Commission's Lifelong Learning Programme, Project Number 526965-LLP-1-2012-1-GR-COMENIUS-CMP. The authors would also like to thank the rest of GSIC/EMIC Group at the University of Valladolid for their support and ideas, as well as the teacher, Daniel Mantilla Fernández, students and families who participated in the study.

6. REFERENCES

- [1] N. Avouris, G. Fiotakis, G. Kahrmanis, M. Margaritis, and V. Komis. Beyond logging of fingertip actions: analysis of collaborative learning using multiple sources of data. *Journal of Interactive Learning Research*, 18(2):231–250, 2007.
- [2] W. Greller and H. Drachsler. Translating learning into numbers: A generic framework for learning analytics. *Educational Technology and Society*, 15(3):42–57, 2012.
- [3] A. Martínez-Monés, A. Harrer, and Y. Dimitriadis. An interaction-aware design process for the integration of interaction analysis into mainstream CSCL practices. In S. Puntambekar, G. Erkens, and C. E. Hmelo-Silver, editors, *Analyzing Interactions in CSCL*, pages 269–291. Springer US, 2011.
- [4] A. Pardo and G. Siemens. Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3):438–450, 2014.
- [5] M. J. Rodríguez-Triana, A. Martínez-Monés, and J. I. Asensio-Pérez. Monitoring collaboration in flexible and personal learning environments. *Interaction, Design and Architecture(s) Journal, special issue on: Evaluating Educative Experiences of Flexible and Personal Learning Environments.*, 11(12):51–63, 2011.
- [6] M. J. Rodríguez-Triana, A. Martínez-Monés, J. I. Asensio-Pérez, and Y. Dimitriadis. Towards a script-aware monitoring process of computer-supported collaborative learning scenarios. *International Journal of Technology Enhanced Learning*, 5(2):151–167, 2013.
- [7] M. J. Rodríguez-Triana, A. Martínez-Monés, J. I. Asensio-Pérez, and Y. Dimitriadis. Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating cscl situations. *British Journal of Educational Technology*, 2014.
- [8] N. Sclater. Code of practice for learning analytics. A literature review of the ethical and legal issues. Technical report, JISC.
- [9] S. Slade and P. Prinsloo. Learning analytics ethical issues and dilemmas. *American Behavioral Scientist*, 57(10):1510–1529, 2013.
- [10] A. Vozniuk, S. Govaerts, L. Bollen, S. Manske, T. Hecking, and D. Gillet. AngeLA: Putting the Teacher in Control of Student Privacy in the Online Classroom. In *ITHET 2014*, 2014.