# A First Step Towards Learning Analytics: Implementing an Experimental Learning Analytics Tool

JUSTIAN KNOBBOUT & ESTHER VAN DER STAPPEN

**Abstract** The educational domain is momentarily witnessing the emergence of learning analytics – a form of data analytics within educational institutes. Implementation of learning analytics tools, however, is not a trivial process. This research-in-progress focuses on the experimental implementation of a learning analytics tool in the virtual learning environment and educational processes of a case organization – a major Dutch university of applied sciences. The experiment is performed in two phases: the first phase led to insights in the dynamics associated with implementing such tool in a practical setting. The second – yet to be conducted – phase will provide insights in the use of pedagogical interventions based on learning analytics. In the first phase, several technical issues emerged, as well as the need to include more data (sources) in order to get a more complete picture of actual learning behavior. Moreover, self-selection bias is identified as a potential threat to future learning analytics endeavors when data collection and analysis requires learners to opt in.

**Keywords:** • Learning Analytics • Tool • Implementation • Experiment • Education •

---

CORRESPONDENCE ADDRESS: Justian Knobbout, HU University of Applied Sciences Utrecht, Nijenoord 1, 3552 AS Utrecht, The Netherlands, e-mail:justian.knobbout@hu.nl. Esther van der Stappen, HU University of Applied Sciences Utrecht, Nijenoord 1, 3552 AS Utrecht, The Netherlands, e-mail: esther.vanderstappen@hu.nl.

## 1    Introduction

Data analytics is already applied in many industries. The educational domain, however, has only recently started using data to improve its processes (Ferguson, 2012). Analytical activities aimed at improving education at a micro-level of educational institutes is called learning analytics: "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs" (Long et al., 2011). Common research objectives include modelling student behavior in virtual learning environments and prediction of performance (Papamitsiou & Economides, 2014). The learning analytics process comprises four steps: 1) learners generate data, 2) these data are captured, collected and stored, 3) analysis and visualization of the data, and 4) the design and use of data-driven pedagogical interventions (Clow, 2012). The process is a cycle, as the effects of the interventions can again be measured, analyzed, visualized et cetera - see Figure 1.
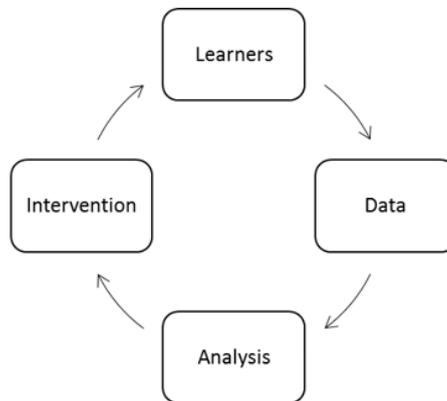


Figure 1: Learning Analytics Cycle (Clow, 2012)

Many examples of small-scale learning analytics initiatives exist (cf. MacFadyen & Dawson, 2010; Romero-Zaldivar et al., 2012; Lonn, Aguilar & Teasley, 2015; Fidalgo-Blanco et al., 2015). Our case organization, however, has only very limited experience with the application of learning analytics. To overcome this lack of experience, a learning analytics experiment is conducted. This study aims at identifying and understanding some of the dynamics associated with the implementation of learning analytics tools within the case organization and in educational institutes in general. Two research questions will be answered: 1) what issues are encountered when implementing an experimental learning analytics tool in the case organization's virtual learning environment, and 2) in what way can the visualizations in the learning analytics teacher dashboard be used to design and perform timely pedagogical interventions? In this study, we will implement a externally developed learning analytics tool and research what barriers need to be overcome in order to apply the visualizations of the tool to education.

30TH BLED eCONFERENCE: DIGITAL TRANSFORMATION – FROM CONNECTING THINGS TO
TRANSFORMING OUR LIVES (JUNE 18 – 21, 2017, BLED, SLOVENIA)
J. Knobbout & E. van der Stappen: A First Step Towards Learning Analytics:
Implementing an Experimental Learning Analytics Tool

325

The remaining of this paper is structured as follows. First, the background of learning analytics processes and our experiment will be provided. We then elaborate on the primarily findings of the experiment conducted thus far. Finally, future work will be described.

## 2        Background

Higher educational institutes implementing learning analytics processes face several difficulties, e.g., changing existing information systems by implementing a learning records store (LRS) and customizing data streams (del Blanco et al., 2013); managing the increase in workload for teachers (Whale, Valenzuela & Fisher, 2013); and making sure all activities are in compliance with privacy legislation (Jisc, 2015). Greller & Drachsler (2012) provide a generic learning analytics framework with six critical dimensions to consider whilst setting up learning analytics services - see Figure 2.
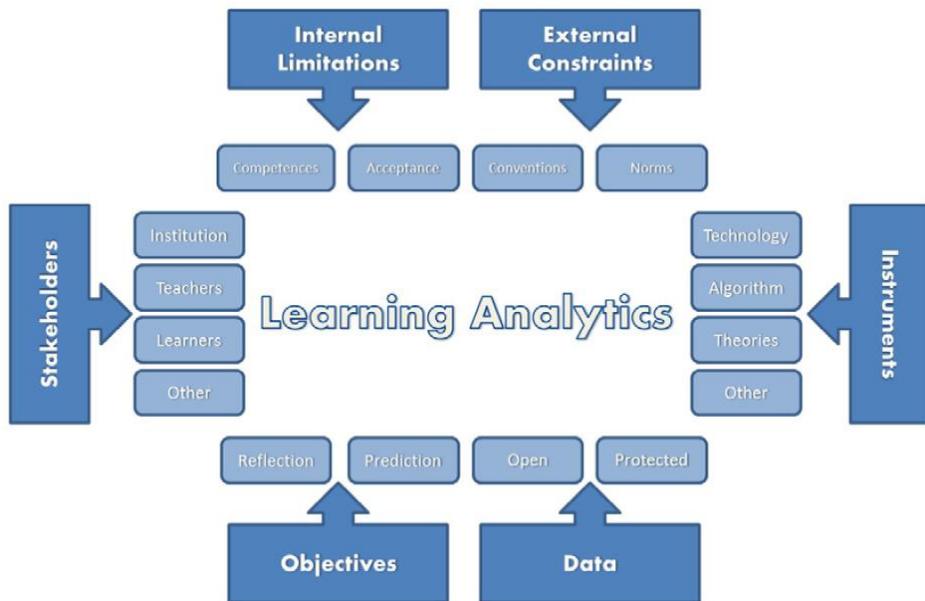


Figure 2: Learning Analytics Framework (Greller & Drachsler, 2012)

In order to help Dutch higher educational institutes overcome the aforementioned difficulties, SURFnet  set up the learning analytics experiment (2016). The experiment provides all instruments – IT-architecture, data standard, algorithms – required to answer five pre-defined pedagogic questions. These questions are sourced from previous research aimed at identifying questions relevant for Dutch educators (Berg et al., 2016). Setting up the tool is relatively easy, as the only necessarily activities involve putting tracking codes on the pages and learning materials in the virtual learning environment.

Since the codes can easily be copy-pasted, only basic computer skills are required. Once placed, the codes allow data to be stored in a learning records store, from which the data can be analyzed and visualized in a dashboard. See Figure 3 for the learning analytics architecture and Figure 4 for the resulting dashboard as seen with a teachers' account. Teachers can see the anonymized activities of all learners in the experiment; learners have their individual dashboard and can view their own activities. To ease compliance with Dutch privacy laws, students must opt-in to allow their data being captured and analyzed.
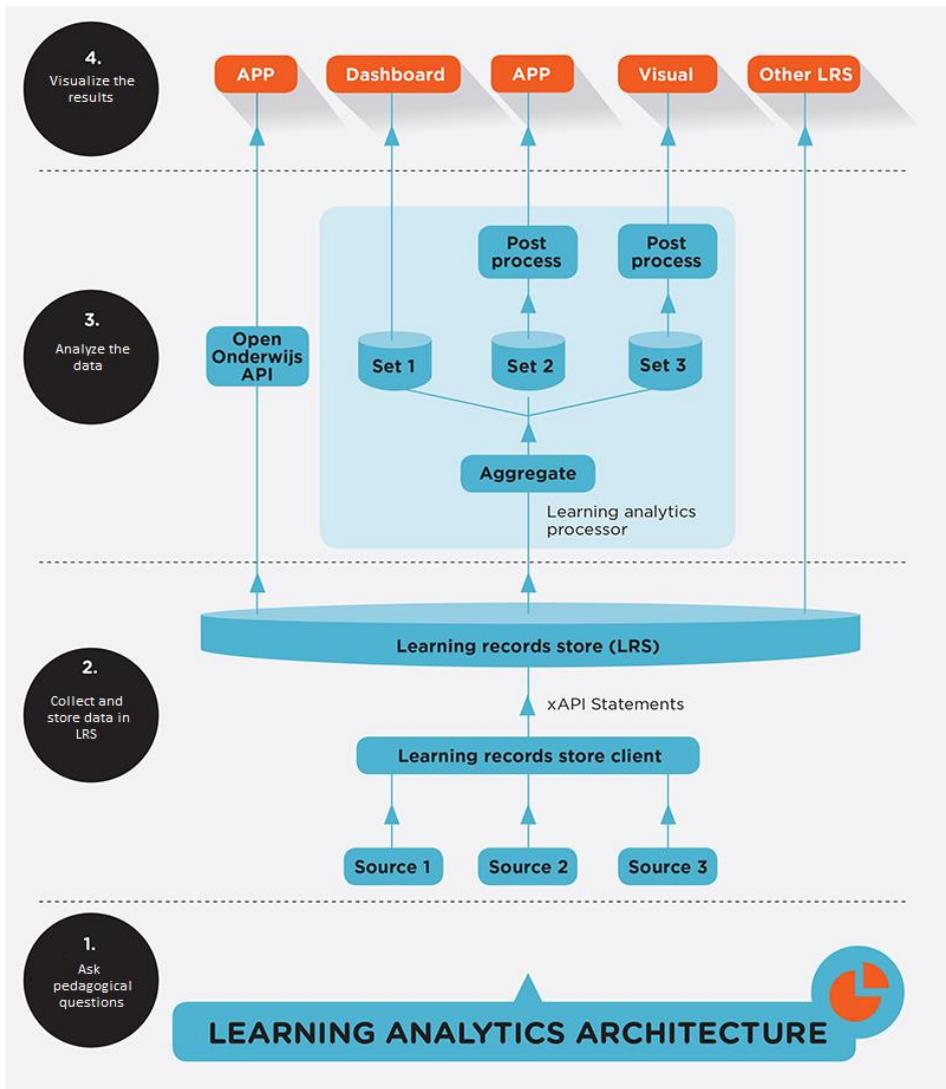


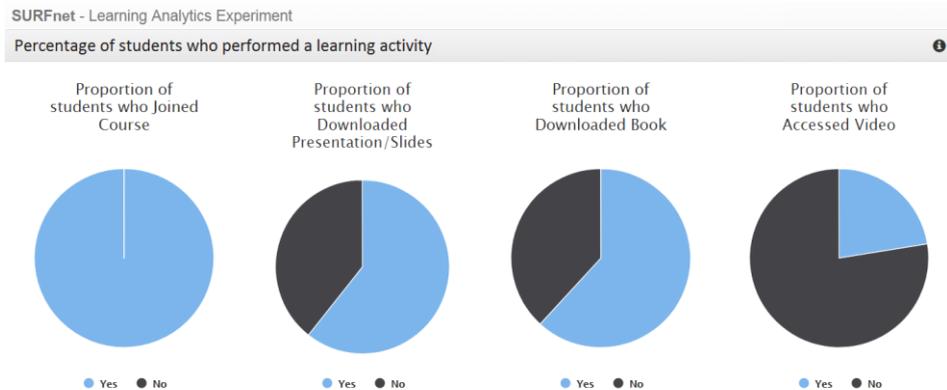Figure 3: Learning Analytics Architecture (source: SURF, 2016)

Figure 4: Dashboard As Seen With Teacher Account

Our case organization is a large university of applied sciences in the Netherlands. At two locations within the organization the Learning Analytics Experiment is conducted; at the Institute of Engineering and Design (IED) and at the Institute of Teacher Education (ITE). At IED, data is collected by tracking activities performed by first-year students of the Business Engineering undergraduate program enrolled in a statistics course. ITE, on the other hand, educates future secondary school English teachers - the experiment is here conducted in the courses Curriculum Design and ICT-rich Education.

## 3 Research objectives and method

The objective of this study is to research to what extent it is possible to perform learning analytics activities in the case organization's virtual learning environment and what barriers are encountered when doing so (first phase of the experiment), and to research in what way pedagogical interventions can be designed and performed during the course by teachers based on the visualizations in the learning analytics dashboard (second phase of the experiment). We do this by answering the following research questions:

1. What issues are encountered when implementing an experimental learning analytics tool in the case organization's virtual learning environment?
2. In what way can the visualizations in the learning analytics teacher dashboard be used to design and perform timely pedagogical interventions?

We use the framework of Greller & Drachsler (2012) to create a shareable description of context parameters for this learning analytics project. Below, we describe the first phase of the learning analytics experiment in Table 1. In the section Future Work of this paper, we will provide a description of the proposed second phase.

Table 1: Learning Analytics Experiment Case Description

| Dimension | Value |
|---|---|
| Stakeholders | *Students* of two institutes of the case organization, participating in one of three courses (n=294). Three *lecturers* teaching the courses and actively involved in the experiment. SURFnet as external *facilitator* of the experiment. IT-department to allow SURFconext connection. |
| Objective | *Implementing* a learning analytics tool integrated in the organization's virtual learning environment. |
| Data | *Activities of students* in virtual learning environment, measured via the experimental tool. |
| Instruments | SURFnet's *Learning Analytics Architecture*, including teacher and student dashboards. *Virtual learning environment* of case organization. |
| External limitations | *Privacy laws* call students to opt-in for the experiment. *Connection* between case organization's and SURFnet's architecture. |
| Internal limitations | *Time* necessary to place tracking code on target pages and learning materials in the virtual learning environment. |

After the first phase of the experiment, all involved teachers and researchers from SURFnet evaluated the process and outcomes during a focus group. Based on open observations, all experiences worth pointing out by the participants were discussed and notated. As all teachers involved encountered the same issues, consensus was reached fast. This led to the description of the preliminary results, as shown in the next section.

## 4 Preliminary results

The first round of the experiment resulted in several experiences and insights. In this section, we will elaborate on the most pressing results. First, an overview of encountered issues is provided. These are then categorized according to the dimensions of the learning analytics framework (Greller & Drachsler, 2012).

## 4.1 Technical issues

Although it is relatively easy to set up the tracking at the virtual learning environment, some technical issues arose. During the experiment, it became clear that not all activities were tracked. Root cause for this anomaly were the necessity to opt-in for the experiment; students who opted-in only provided permission to capture data from the specific device they used to subscribe. This results in incomplete datasets and renders extensive data

analysis useless. The causing issue has been resolved, so this problem will not appear in the second round of the experiment.

Another encountered technical issue relates to the SURFconext connection between the case organization's virtual learning environment and the SURFnet learning analytics dashboard. With SURFconext, it is possible for students to log in the learning analytics dashboard by using their own institutional username and password. This saves the need to create a new account. To establish the connection, the institutional SURFconext contact person must login his administrator account and provide permission to do so. At the beginning of the experiment, however, the connection was not allowed until only a few hours before the start of the courses. This almost led to cancellation of the experiment as the researchers did not want to confuse students with non-working dashboards and connections. This experience shows the need for institutional-broad support and cooperation in order to make learning analytics work.

## 4.2 Self-selection bias

Students need to opt-in for the experiment, allowing their data being captured and analyzed. Of the 234 Business Engineering students who enrolled in the course (both first-year students and students from later years wanting to retake the exam), 89 opted-in – only 38%. Similar proportions were found at the other institute. After the final exam of the course, the difference between the final grade of the students who participated in the experiment and those who did not were calculated  using SPSS. Participants scored an average grade of 5.1 (sd = 2.2) and non-participants an average grade of 4.3 (sd = 2.3). An independent t-test shows there is a significant difference ($p = 0.032$) between the two groups and students who participated in the experiment scored on average better than those who did not. As no interventions were initiated based on the information in the dashboard, it is suspected that self-selection takes place. Future initiatives involving voluntary participation must account for this effect.

## 4.3 Need for more data

In order to perform effective interventions, rich data is required (Tempelaar et al., 2015; Conijn, 2016). In our experience, the current data is too poor to use for interventions. The lack of sufficient data can partly be traced back to the teachers who put too few tracking codes on their course page of the virtual learning environment. As one of the involved teachers puts it: "next time I will track everything: every page, every article and every video." The experiment also relied on data from the virtual learning environment only. This provides only one side of the story - anecdotal observations of students using the online materials showed that they sometimes jointly sit together at a single computer to work on assignments, having the system only registering one student. Similarly, interaction between these students cannot be measured this way (Pardo & Kloos, 2011). In our experiment, data from other resources was not aggregated with the data in the learning records store. For example, the virtual learning environment provides students

330     30TH BLED ECONFERENCE: DIGITAL TRANSFORMATION – FROM CONNECTING THINGS TO
TRANSFORMING OUR LIVES (JUNE 18 – 21, 2017, BLED, SLOVENIA)
J. Knobbout & E. van der Stappen: A First Step Towards Learning Analytics:
Implementing an Experimental Learning Analytics Tool

the possibility to take quizzes. Quiz data (results, number of attempts, required amount of time to finish), however, are yet not stored and processed in the records store. In order to design and perform effective interventions, these data must be aggregated in future experiments.

## 4.4     Problem categorization based on learning analytics framework

In order to analyze which dimensions of the learning analytics framework (Greller & Drachsler, 2012) are causing problems during the implementation of the tool, we mapped the identified issues to the framework – see Table 2. This shows that four out of six dimensions faced difficulties so even though an almost 'plug-and-play' tool is provided, implementing it is a non-trivial endeavor.

Table 2: Problems Encountered per Dimension of Learning Analytics Framework

| Dimension | Issues encountered |
|---|---|
| Stakeholders | Self-selection bias occurs as students need to opt-in for the experiment. IT-department not involved enough to collaborate fast on establishing external connection. |
| Objective | - |
| Data | Data on behavior in virtual learning environment alone do not provide enough insights to design effective pedagogical interventions. More data are needed in order to do so. |
| Instruments | The used tool was too limited in its data capturing. That is, not all activities were measured due to technical issues. |
| External limitations | Connection between case organization's and SURFnet's architecture only established at the very last moment. |
| Internal limitations | Teachers lacked insight in what learning resources to measure by placing tracking codes. |

## 5     Future work

The experiment will be continued in the Fall of 2017. At the Institute for Engineering and Design, two courses will implement the learning analytics tool in their course design. One of the courses is taught to fulltime students, the other one to part-time students. This provides the opportunity to explore behavioral differences between participants of the two programs. Furthermore, to date, no interventions were performed based on the data analysis and visualizations in the experiment. Learning analytics research in general often focuses on data collection, management or how data will help to improve education but designing effective pedagogical interventions becomes a critical element (Wise, 2016). Now we have demonstrated the tool can be implemented in the case organization's virtual

learning environment, we aim to use it for intervention design. This calls for extended requirements, as described in Table 3. The next phase of the experiment will continue the current work and focus on answering the question in what way the visualizations in the learning analytics teacher dashboard can be used to design and perform timely pedagogical interventions. That is, interventions should take place when it is still possible to make changes to the learning behavior, i.e., during the course. Both the practical experiences gained and the to-be performed interventions benefit practitioners from the educational domain as it provides first-hand insights in the dynamics involved with starting learning analytics activities.

Table 3: Future Learning Analytics Experiment Case Description

| Dimension | Value |
|---|---|
| Stakeholders | *Students* participating in one of two courses. Six *lecturers* teaching the courses, two of them actively involved in the experiment. SURFnet as external *facilitator* of the experiment. |
| Objective | *Reflecting* on learning activities of students and performing pedagogical interventions based on (the lack of) these activities. |
| Data | *Activities of students* in virtual learning environment, measured via the experimental tool. *Formative assessment data.* |
| Instruments | SURFnet's *Learning Analytics Architecture*, including teacher and student dashboards. *(External) tools* to capture additional data, e.g., formative assessment data. |
| External limitations | *Privacy laws* call students to opt-in for the experiment. *Connection* between case organization's and SURFnet's architecture. |
| Internal limitations | *Time* necessary to place tracking code on target pages and learning materials in the virtual learning environment. The teachers doing so must *understand* what data are required for the interventions. *Competencies* related to effective intervention design are required, that is, teachers must understand the visualizations and are able to design and perform useful interventions based on it. |

**References**

Berg, A., Bogaard, M. van den., Drachsler, H., Filius, R., Manderveld, J., & Schuwer, R. (2015). Grand challenges learning analytics & open en online onderwijs: een

verkenning        [whitepaper].        Retrieved        from
https://www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2015/rapport-grand-
challenges-la--ooo---een-verkenning.pdf on 11 May 2017.

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. Paper
presented at the Proceedings of the 2nd International Conference on Learning
Analytics and Knowledge, pp. 134-138.

Clow, D. (2013). An overview of learning analytics. Teaching in Higher Education,
18(6), 683-695.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2016). Predicting student
performance from LMS data: A comparison of 17 blended courses using moodle
LMS. IEEE Transactions on Learning Technologies,

Del Blanco, Á., Serrano, Á., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B.
(2013). E-learning standards and learning analytics. can data collection be improved
by using standard data models? Paper presented at the Global Engineering Education
Conference (EDUCON), 2013 IEEE, pp. 1255-1261.

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

Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J., & Conde, M. Á. (2015).
Using learning analytics to improve teamwork assessment. Computers in Human
Behavior, 47, 149-156. doi:http://dx.doi.org/10.1016/j.chb.2014.11.050

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic
framework for learning analytics. Educational Technology & Society, 15(3), 42-57.

Jisc (2015). Code of practice for learning analytics. Retrieved from
http://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics on 11 May
2017.

Long, Philip, George Siemens, Gráinne Conole and Dragan Gašević (2011). Proceedings
of the 1st International Conference on Learning Analytics and Knowledge (LAK11).
Banff, AB, Canada, Feb 27 - Mar 01, 2011. New York: ACM.

Lonn, S., Teasley, S., & Krumm, A. (2009). Investigating undergraduates' perceptions
and use of a learning management system: A tale of two campuses. Paper presented
at the Annual Meeting of the American Educational Research Association (April
16th), San Diego, California. Retrieved June, , 6. pp. 2014.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning
system" for educators: A proof of concept. Computers & Education, 54(2), 588-599.

Papamitsiou, Z. K., & Economides, A. A. (2014). Learning analytics and educational data
mining in practice: A systematic literature review of empirical evidence. Educational
Technology & Society, 17(4), 49-64.

Pardo, A., & Kloos, C. D. (2011). Stepping out of the box: Towards analytics outside the
learning management system. Paper presented at the Proceedings of the 1st
International Conference on Learning Analytics and Knowledge, pp. 163-167.

Romero-Zaldivar, V., Pardo, A., Burgos, D., & Kloos, C. D. (2012). Monitoring student
progress using virtual appliances: A case study. Computers & Education, 58(4), 1058-1067.

SURFnet (2016). Het learning analytics experiment. Retrieved from https://blog.surf.nl/het-
learning-analytics-experiment/ on 11 May 2017.

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.
Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2014). Learning analytics for online discussions: Embedded and extracted approaches. Journal of Learning Analytics, 1(2), 48-71.