

Delft University of Technology

"Learning Analytics is about Learning, not about Analytics" A reflection on the current state of affairs

Van den Bogaard, M.E.D.; De Vries, P.

Publication date 2017 **Document Version** Final published version Published in 45th Annual SEFI Conference

Citation (APA)

Van den Bogaard, M. E. D., & De Vries, P. (2017). "Learning Analytics is about Learning , not about Analytics" A reflection on the current state of affairs. In *45th Annual SEFI Conference* ISEP Lisbon.

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

"Learning Analytics is about Learning, not about Analytics." A reflection on the current state of affairs.

M.E.D. van den Bogaard¹

Senior researcher and consultant Leiden University Leiden, the Netherlands E-mail: m.e.d.van.den.bogaard@iclon.leidenuniv.nl

P. de Vries

Researcher Delft University of Technology Delft, the Netherlands E-mail: pieter.devries@tudelft.nl

ABSTRACT

Learning Analytics is a form of educational datamining that is used on the level of the classroom and the student. This information can be used to predict behaviour and address that, or it can be used for teachers and students to reflect on their learning processes and behaviour. Two examples of LA are presented: one with data from a VLE, one with student progress data and in these examples it is shown where the challenges for the near future of LA lay: there is a need for technology assessment of these technologies, the focus needs to be on learning and appropriate interventions to enhance learning, and on the teachers who are the real key players in the successful application of learning analytics to enhance learning.

Conference Key Areas: open and online engineering education, engineering education research

Keywords: Learning analytics, engineering education, VLE

INTRODUCTION

Big data are presented qualified as the fuel of the future. Clearly it is not something to ignore and with the increasing use of technology, education is also challenged to engage with it. The analysis of big data for education is referred to as educational data mining and when it is applied to the classroom is often referred to as learning analytics (LA). This paper is about the use of learning analytics for the meso- and micro-environment of teaching and learning: the levels of the classroom and the students. The emerging development of digital tools for teaching and learning is

¹ Corresponding author: Maartje van den Bogaard

E-mailaddress: m.e.d.van.den.bogaard@iclon.leidenuniv.nl

creating a playing field of data. Companies very quickly learned how to benefit from the use of data in various ways, think of how Amazon.com makes recommendations for books and products based on previous orders and browsing history and behaviour of customers with similar interests. Fortunately, it has and it will increasingly affect education, but it proves to be a tough challenge to let education benefit from these developments. This paper is about the promises of big data or in this case learning analytics to help improve education and what we as educators can do about it to make it work

In recent years, lots of energy has been put into the development and deployment of virtual learning environments (VLEs) and MOOCs and in research into these developments, assuming that these technologies will have a large impact on education, but these efforts have not yielded major changes or many new insights that have enhanced teaching and learning (Gašević, Kovanović, & Joksimović, 2017).

In this paper, we discuss the state of affairs with learning analytics using two recent examples that should help practitioners understand where the current challenges in research and praxis are and what they can expect of LA in the (near) future.

1 LEARNING ANALYTICS

1.1 What is it?

The history of higher education is littered with imperfect technologies: Slides, overheads, film, video, educational television, multimedia CDs, Second Life, PowerPoint slides, chat rooms, forums, learning management systems. All promised to revolutionize learning and change the face of education. So far only the computer and the internet have lived up to such expectations. With the introduction of the internet and web-based applications like VLEs it became possible to online collect data about learning processes and achievements from students using these technologies and make them available for analysis. The idea behind learning analytics is to use this data to inform/support educational decision-makings. According to SoLAR 'learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs' (Siemens & Gašević, 2012). Gašević, Kovanović and Joksimović (2017) guote Reimann (2016) who notes that, while the use of data in educational research and learning sciences has been present for quite some time, learning analytics differs from the "traditional" data analyses in education as it focuses on the longitudinal collection of a large number of data points from authentic learning environments. "Although the field of learning analytics supports the collection of a wide range of data, a bulk of existing work in learning analytics is dedicated to digital traces collected through the interaction of people with technology, content, and/or other people (Gašević et al., 2017). An important task of learning analytics is the development of measures that can a) offer practical insights into learning processes and outcomes, and b) be theoretically interpreted (Gašević et al, Dawson, & Siemens, 2015). Both are important: if LA is to give practical insights, it means that the variables that are used in the analytics are actionable. For instance, from student success research it is well know that gender has a major impact on success, but just recruiting more women into the population will not aid student success as such. With LA it would be far more interesting to explore in what areas of the curriculum we can identify areas where we can intervene meaningfully building on the body of knowledge on teaching and learning.

1.2 Academic analytics versus learning analytics

The focus of what we do in classrooms often is on the student and the teachers, but these stakeholders operate within the context of an organisation. McKenny and Mor (2015) incorporate these three perspectives in their model on how learning analytics, learning design and teacher inquiry synergistically support a rationale core consisting of aims, context, vision and values underlying education. This rationale should be a consistent and explicit understanding of: teaching and learning aims; context; pedagogical vision; and educational values within an organisation. While many teachers welcome the processes of learning analytics, learning design and inquiry, few have sufficient resources and expertise to engage with each in a sustained and systematic manner. The conceptual model shown in Figure 1 illustrates that each process is guided by, and can contribute to the refinement of the same rationale. In this model learning analytics looks at data, teaching/learning practices and social structures to inform teaching and learning design and as such, forms the input for learning design and teacher inquiry. Again, this would only work if the learning analytics are based on actionable variables. These three perspectives build on and contribute to the same rationale and, additionally, the three core processes can synergise. All of these processes are complex and with every process a teacher is confronted with several challenges, which will require time and expertise to deal with (McKenney & Mor, 2015).



Figure 1: Learning analytics, learning design and teacher inquiry synergistically support core rationale (McKenney & Mor, 2015)

Within the field of learning analytics researchers often discern between academic analytics, which is the analytics used for marketing and administrative purposes (the macro level which is often also referred to as institutional research) and learning analytics, the analytics that could be used by teachers for purposes of digital classroom management (the meso-level) or to improve the students' learning, or could be used by students to reflect on their learning processes and make decisions regarding their learning pathways (the micro-level).

In this paper, we will not discuss this macro-level of academic analytics, but we need to be aware that it is on this organisational level where most decisions are made that influence the meso- and micro-levels of education. Examples include the decision regarding which VLE is used, which applications within this VLE are available to teachers, what data are being collected and if and how they are presented, the facilitation of the teacher to use the data, etc. In this light, learning analytics are very much an institutional endeavour, that need to accommodate the teacher by delivering timely and accurate data to use in a meaningful way in a classroom. By ignoring the institutional level, the teacher is put in an unfavourable position, as there is little this person can do to make the transition from macro- and meso-level to the pedagogical micro-level.

1.3 Issues with adoption and implementation

Challenges with the adoption, implementation and evaluation of learning analytics have been described more extensively by e.g. Sclater (2014) and SURF (2015). In this section, we capture the major outcomes of this ongoing discussion. Learning analytics is an interdisciplinary domain with many disciplines contributing to the research. So far, however, it is mainly the disciplines of data science and to a lesser extent, education that are involved in research on learning analytics. This is reflected in the kind of research that has had a lot of exposure within the learning analytics community. The data scientists have invested a lot of time in exploring the data using datamining techniques. This research has been successful in identifying students at risk for example, but the variables that have discriminatory power are often nonactionable variables, that are difficult to interpret in the light of possible pedagogical interventions and much of the research is affected by problems of overfitting (Gašević et al., 2017). Most of the phenomena predicted in this kind of research are not linear and the relation between cause and effect often is unclear (Forsman, Linder, Moll, Fraser, & Andersson, 2014). From a traditional educational perspective, the approach would be to find a theory that could explain or support mitigation of issues experienced in the classroom and build data collection and analysis from there. With such an approach the theory will guide the teacher or researcher in what kind of data needs to be collected and how the outcomes can be interpreted and used for pedagogical didactical interventions in the classroom or learning process of the students. Most teachers in engineering education, however, are not facilitated nor have the expertise to work in such a structured way, as was also observed by McKenny and Mor (2015).

In recent years, many encompassing models for implementation of analytics have been developed using another perspective. Often these models are generic and address institutions as a whole or implementation as an abstract and strategic process. Colvin at al. (2016) and Greller and Drachsler (2012) observe that learning analytics are not linear or uni-dimensional phenomena, but that (implementations of) learning analytics are complex, shaped by interdependent 'soft and hard' dimensions. Hard dimensions pertain to the availability of data due to technological or privacy issues, tooling and instrumentation. The soft dimensions pertain to stakeholders' knowledge and skills to collect, process, analyse and interpret data, and the conventions and norms as to what is accepted within the context of the institution (Gašević et al., 2017; SURF, 2015). Bos (2016) found in her work on the adoption and effectiveness of using educational technology in higher education that, if implementation of technology is solely left to teachers, chances are that it will not be implemented effectively nor efficiently, while it is the teachers who should play a pivotal role in making the link between the analytics and the pedagogical didactical interventions on the micro-level of teaching and learning. Rienties, Toetenel and Bryan (2015) argue that it is high time that researchers, teachers and managers start working together to combine efforts in learning analytics research to understand how context, learner characteristics such as motivation and behaviour, and learning design to impact the learners.

2 APPLICATIONS OF LEARNING ANALYTICS: PREDICTION OR REFLECTION

Greller and Drachsler (2012) state that learning analytics can focus on two kinds of applications: prediction and reflection. Prediction is intended to identify at risk students at the earliest instance as possible before or during a course. Within every course or subject other variables may contribute to whether or not a student is at risk. If a student is identified, a system could either be programmed in such a way that the student will be offered additional, adaptive, feedback or material or the decision to intervene and how to intervene could be placed with the teacher of the course. Most learning analytics research, especially in MOOCs research, is focused on predicting whether or not students will finish the course. The variables that prove to be of importance for predicting completion are variables well known from student success research (see e.g. Van den Bogaard, 2012): gender, prior achievement, motivation and student behaviour in the course account for most of the variance. However, most of these variables are not actionable and in that sense the research into MOOCs does not show a lot of innovation in itself. Predictive analytics are a first step to adaptive learning environments where a student is offered a pathway through the learning environment based on characteristics, behaviour and performance in a virtual learning environment, however, little progress is being made with developing diverse pathways through the MOOCs or with experimenting with new arrangements for learning (Gašević & Dawson, 2015; Skrypnyk, Hennis, & De Vries, 2015).

With reflective analytics, the aim is to provide feedback to the teacher and/or the learner for critical self-reflection to obtain self-knowledge. On an individual level this pertains to the learning processes by offering information on progress. On the institutional level, it could enhance monitoring process and use that to suggest interventions. This could for instance be feedback on how well the student is performing in comparison to other students in the course, or a summary of how the student has been participating in the course, e.g. has the student only been online just prior to deadlines or has the student been studying consistently over time. Reflective analytics can also provide the teacher with information that is useful for 'digital classroom management'. In the regular classroom, it is relatively easy for teachers to keep track of which students contribute to the class and who comes prepared, in a VLE this is much harder to keep track of. Reflective analytics can provide teachers with real time information on how the students are participating, what kinds of contributions they make to the class in the VLE, etc. With this kind of analytics it is essential to consider what data needs to be logged to facilitate effective reflection. Again, it is up to the teacher and/or student how to respond to the information. For effective use of reflective analytics, it is important that dashboards are provided to the users. Defining effective dashboards is a field in its own right and we will not touch on that topic in this paper (SURF, 2015).

Predictive analytics and reflective analytics are both yet not fully matured: the technology is still under development, tools are being developed and institutions are working hard to create infrastructures to collect, combine and present data in a more

efficient manner. The developments in this field are rapid and involve things such as Artificial Intelligence (AI) and machine learning, however it will take time before these applications are sufficiently accessible for teachers who would like to use analytics in their virtual classrooms. The use of student data will definitely evoke privacy and ethics issues that need to be considered by the institutions and students, such as the question of algorithms taking over parts of our teaching, which pertains to technology assessment and requires a dialogue between all the stakeholders in education and analytics. Until than, we have a lot of data already available and we need to further develop out thinking about what and how to use this resource for the well-being of teaching and learning. The available data is often relatively easy to understand. In the next section of this paper we discuss a few applications of analytics in engineering education that can serve as examples of analytics and that can aid in building an understanding of the use of data that may make it easier to catch on as soon as new applications to run analytics become available.

2.1 Example 1: Analytics of VLE data in X University of Technology

X University of Technology uses Blackboard as a VLE, in which all user data are logged and stored. We ran a small experiment with analytics. It took place in the first semester of the academic year of 2013/2014. It concerned logging data directly from the servers of the VLE for two courses, one in civil engineering on Fluid Mechanics (in BA2) and one in mechanical engineering on Thermodynamics (in BA1), of which it was known that the teachers used the VLE extensively. In this paper, we report on the thermodynamics course. For the duration of the course all the log data, click data and ping data, which indicate how long a certain window was open, were downloaded from the servers. At the end of the course this data was combined with the final grades obtained for the course. This data was analysed using educational datamining techniques for subgroup analysis where we looked at which variables in the data gave the highest predictive score out of one, two and three variables for three different outputs: predicting a fail, predicting a pass and predicting the final grade (LIACS, 2017). The outcomes of the prediction of a fail are presented in Table 1.

Output variables	N predictive variables	Thermodynamics course (Mech. Eng, BA1)
Predicting a fail	1 variable	64 % predicted correctly
		n distinct sessions <= 13
	2 variables	69 % predicted correctly
		n times access to content X <=48 n distinct sessions <=14
	3 variables	73 % predicted correctly
		n times access to content X <=48 average duration of the intervals between access in hours <=424 n distinct sessions <=14

Table 1: Predictive variables failing the Thermodynamics course.

Overall it is safe to state that engagement in the VLE is a good predictor for passing the course with a good grade (Van den Bogaard, 2012). The actual question is whether or not it is possible to discern between the students at risk and the students who are doing well based on these analyses? The outcomes in the table show how complicated it is to answer that question. There are different variables for each depth of analysis of the outcome variable. Even if the same variable shows up, the cut off scores tend to vary. When we study the predictor variables, we note that not all of these variables are straight forward to interpret. For instance, the predictor variable 'average duration of the intervals between access in hours <=424' in the

Thermodynamics course, it is hard to understand how students who wait less than approximately 17 days between sessions would be more likely to fail than students who do not.

Overall some clear patterns emerge, but it would be very difficult to translate that into reliable analytics that will accurately predict the success of each individual student. It is important to recognise that the variables that came up in the post hoc analysis, may not necessarily be the variables that have predictive powers during the course. Outcomes of analytics for these course prior to the course, during and after the course may be very different.

2.2 Example 2: Reflective analytics in the ABLE project

Within the Erasmus+ project on Achieving Benefits through LEarning Analytics (ABLE), the Nottingham Trent University (UK), KU Leuven (Belgium) and Leiden University (The Netherlands) were collaborating to develop dashboards for the students to reflect on their participation in education and on their progress (www.able-project.org, 2017). In Leuven and Leiden there are progress requirements in place for first year students, but student counsellors find that students at risk have a tendency to make uninformed decisions when it comes to resit exams. Students tend to attribute failure to a lack of effort and often believe that they can make up for lost exams by working very hard. Students often try to take too many resits and fail as a result. Student counsellors warn students about these decisions, but students lack the experience of the counsellors when it comes to making decisions regarding planning, resits, and barriers in the curriculum that require the students' attention. At Leuven a dashboard has been developed to aid the conversations between student counsellors and students on their options for resits. This dashboard is based on performance data from previous student cohorts and will show the odds of a student passing resit exams based on the performance of students in similar situations in previous years. This closes the information gap between a student counsellor and a student when discussing the options a student has for resits. The counsellor can inform students about the progress requirements in the programme and show options to students including the odds of passing.

The dashboard is currently being piloted by KU Leuven and Leiden University and it is found that most students and counsellors find this dashboard very useful to have more meaningful conversations on progress and resit options. In that sense the dashboard proves to be useful as a tool for reflection. It is based on data already available in any university at this point in time and therefor is an example of analytics that most universities or programmes could implement in a very short timeframe.

3 CONCLUSIONS AND REFLECTIONS

From this discussion and examples of learning analytics, we draw a number of conclusions, that we present here as four topics of discussion.

3.1 The need of technology assessment of learning analytics

The implications of learning analytics will be far reaching if this technology enters into a state of maturity that allows institutions and teachers to apply the results in their daily practice. It could mean that decisions over learning and, ultimately, on the lives of our students will be made based on or even by machine learning algorithms that are beyond the direct influence of humans. This brings about ethical considerations and it requires an informed dialogue between all stakeholders in education and learning analytics on the ownership and value of data, the analytics and the decisions made. The easiest way out is not to use any of the data that is available, but at the same time it might not be ethical not to use any of the data if it could provide better opportunities to students. It is important to start this discussion in our institutions, but also on a wider scale (Rathenau Institute, 2017; Institute for Technology Assessment, 2017)

3.2 A focus on technology versus a focus on learning

Most research into learning analytics has been done within the context of MOOCs with the focus on models and algorithms that can predict retention in online courses. This research has yielded many interesting outcomes, but few of them will have implications for education practices. This is partly due to the fact that participants in MOOCs tend to be professional learners with different attributes than students enrolled in a university (Hennis, Topolovec, Poquet, & Vries, 2016). Little research within the MOOC community has been done on experiments with learning design and learning arrangements and therefore pointers are lacking for how to offer online learning activities in more effective ways or offer more adaptive learning pathways. Most LA research so far did not focus on learning as such. As Bos (2016) found in her research on the effectiveness of blended learning applications in relation to study behaviour in a VLE, the effect diminished as soon as measures of student motivation were included in the models. As long as such variables and actionable variables are not included in the research, it will be very difficult to apply such research in the learning design as suggested by McKenny and Mor (2015). In the VLE example in this paper it is clear that the VLE will not automatically yield the right information that can be used to make meaningful changes to the learning environment either. If LA is to make a contribution to education, we have a lot of ground to cover in terms of applying LA in the right way and in a situation where the student and teacher are capable of using the data to improve the teaching and learning process.

3.3 Effectiveness and acceptance of interventions

Learning analytics can only be effective if the outcome can generate insights in the pedagogical didactic consequences for the teaching and learning practices. The VLE example in this paper showed that it is very difficult to interpret the outcomes of the VLE analytics. We do not really know why some of these variables show up in the analyses, nor what kind of interventions could be designed based on these analytics and what the analytics should present to clarify whether or not an intervention has been successful. Interventions should be designed based on a theoretical model that helps understand and explain effects, so it can be reproduced. Additionally, to design an effective intervention, we need to learn to understand the needs of our students. The ABLE project is a great example of this: Talking with the students using the dashboard made clear what the students needed. Interventions that are not designed with the learning in mind, are likely to have no effect or to have a detrimental effect. Gašević and Dawson (2015) state that the evaluations of interventions are often botched and are not as effective as researchers initially report. Without an understanding of theory and of what our students are willing to embrace, interventions are unlikely to have any sustained effects.

3.4 Learning analytics is about learning, let's not forget the teachers!

We argued in this paper that there is a large body of knowledge on teaching and learning that should be used to deal with LA effectively. The main outcome of the VLE experiment is that prolonged engagement with the learning materials, in this case the VLE, is a strong predictor for success. This finding is consonant with Skrypnyk, Hennis and De Vries (2015). Feedback is essential for learning and we could use the VLE outcomes as a reminder that we need to look for ways of giving

students meaningful feedback, in the VLEs and in our classrooms. How to do that effectively and efficiently, we do not know yet and the research on LA does not give us a lot of hints. The options of real time LA are promising to this purpose, because it will help teachers to consider what kind of feedback is useful for which kind of student. Grades should never come as a surprise to students. As LA is not yet mature enough to provide adaptive feedback to students, it is up to the teachers to make links between the meso- and micro-levels. It is known from research into student success that the interaction between teachers and students has most impact on student learning (Hattie, 2009; Van den Bogaard, 2015). If we leave the learning design and teacher inquiry to the teachers only, very little will happen. If we leave it to the managers and policy makers, there will only be developments on the macro- and meso-levels in organisations. We should create a situation where teachers are empowered and facilitated to engage with the data they already have. They can complement this with data that can be collected in simple ways and that is easy to interpret and actionable. The challenge is to come up with meaningful and effective interventions on the level of teaching and learning and the core rationale of the institution. Teachers do not have to wait for the technology and applications to mature. They could start today with the data that is already available to them. If they do, they will be ready to start using technology meaningfully and responsibly once it becomes available in the (near) future.

4 ACKNOWLEDGMENTS

The authors would like to thank the E-merge Consortium and the team at the ABLE project.

REFERENCES

- Bos, N. R. (2016). Effectiveness of Blended Learning: Factors Facilitating Effective Behavior in a Blended Learning Environment. Open Universiteit.
- Colvin, C., Wade, A., Dawson, S., Gašević, D., Buckingham Shum, S., Nelson, K., ... Fisher, J. (2016). *Student retention and learning analytics : A snapshot of Australian practices and a framework for advancement. Draft final report.* Canberra: ACT: Australian Geoverment Office for Learning and Teaching.
- Forsman, J., Linder, C., Moll, R., Fraser, D. M., & Andersson, S. (2014). A new approach to modelling student retention through an application of complexity thinking. *Studies in Higher Education*, 39(1), 68–86. http://doi.org/10.1080/03075079.2011.643298
- Gašević, D., & Dawson, S. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1). http://doi.org/10.1007/s11528-014-0822-x
- Gašević, D., Kovanović, V., & Joksimović, S. (2017). Piecing the Learning Analytics Puzzle: A Consolidated Model of a Field of Research and Practice. http://doi.org/10.1080/23735082.2017.1286142
- Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology and Society*, *15*(3), 42–57. http://doi.org/http://hdl.handle.net/1820/4506
- Hattie, J. A. (2009). Visible learning. A synthesis of over 800 meta-analyses relating to achievement. New York City NY: Routledge.
- Hennis, T. A., Topolovec, S., Poquet, O., & Vries, P. de. (2016). Who is the Learner: Profiling the Engineering MOOC Student. *SEFI 44th Annual Conference*, (August).
- Institute for Technology Assessment. (2017).www.ifta.org
- LIACS. (2017). Cortana Subgroup Discovery. Retrieved April 30, 2017, from http://datamining.liacs.nl/cortana.html
- McKenney, S., & Mor, Y. (2015). Supporting teachers in data-informed educational design. *British Journal of Educational Technology*, *4*6(2), 265–279. http://doi.org/10.1111/bjet.12262
- Nottingham Trent University, KU Leuven, & Leiden University (2017). ABLE. Retrieved April 20, 2017, from www.able-project.org
- Rathenau Institute (2017). Research and dialogue, science, technology and innovation. www.rathenau.nl/en
- Rienties, B., Toetenel, L., & Bryan, A. (2015). "Scaling up " learning design : impact of learning design

activities on LMS behavior and performance. In LAK '15 March 16-20. Poughkeepsie NY.

Sclater, N. (2014). Learning analytics The current state of play in UK higher and further education.

- Siemens, G., & Gašević, D. (2012). Guest editorial learning and knowledge analytics. *Educational Technology and Society*, *15*(3), 1–2. http://doi.org/10.1207/s15327752jpa8502
- Skrypnyk, O., Hennis, T. A., & Vries, P. De. (2015). Reconsidering Retention in MOOCs: the Relevance of Formal Assessment and Pedagogy. *Proceedings of the Third European MOOCs Stakeholder Summit*, 1(23), 166–172. http://doi.org/10.13140/RG.2.1.1881.3286
- SURF. (2015). Grand Challenges Learning Analytics & Open En Online Onderwijs. Een verkenning. Utrecht the Netherlands,: SURF Foundation.
- Van den Bogaard, M. E. D. (2012). Explaining student success in engineering education at Delft University of Technology: a literature synthesis. *European Journal of Engineering Education*, 37(1), 59–85. http://doi.org/10.1080/03043797.2012.658507
- Van den Bogaard, M. E. D. (2015). Towards an action-oriented model for first year engineering student success. A mixed methods approach. Delft University of Technology: dissertation.