Using Small Data to Support the Design Process of Learning Environments

Christian Ostlund, Lars Svensson, William Jobe
University West
School of Business, Economics and IT
461 86 Trollhättan, SWEDEN
{Christian.Ostlund, Lars.Svensson, William.Jobe}@hv.se

***Abstract**. Learning Analytics is seen as a promising way of improving online education, focusing on how students learn and how teachers' can help the students' to be effective in that process. Most approaches rely on the idea of "Big Data" to be the foundation upon which to design for assisting the students' learning process, the teachers' scaffolding efforts or enhancement of the interaction design of the software. In this paper, we depart from a case of e-training of clerks in a public organisation. It is argued that learning analytics and educational mining based on "Small Data" (i.e. use-pattern log data from a small set of learners) also have a promising potential for extracting design guidelines for an e-training environment and not only provide support after the system is developed and implemented, but also guide the design of future e-training systems.*

Keywords. learning analytics, online education, design science research, design, small data

1 Introduction

The interest in using web mining techniques in e-learning (i.e. learning analytics and/or educational mining) has increased over the last decade. Earlier, similar approaches were commonly used in e-commerce to encourage users to shop (e.g. Chou et al. 2010). Learning analytics can e.g. be used for analyzing social networks and help predicting the success or failure of the students based on previous behavior (Gašević et al.,2013). The focus of e-learning is shifting from the effectiveness of e-learning to focus more on teaching and learning practices (Hung 2012) and to that end educational mining can be effective. Romero and Ventura (2006) identifies three different actors having an interest in educational mining; students, teachers, and academic administrators. One example of student support could be a recommender system aiming at improving course material navigation (Ota & Kashihara 2010) as well as assist the overall online learning process (Zaiane 2002; Khrib et al. 2008; Svensson 2003). To support the student, Zaiane (2002) use web mining to build recommender system suggesting learning activities or shortcuts in an e-learning environment based on learners' previous activity in the e-learning environment. Gellerstedt et al. (2014) present strategies on how to address diversities in student learning styles using learning analytics for measuring and visualizing student activity.

To support the teacher Ueno and Nagaoka (2002) used web mining to create a system that identifies irregularities in the learning process, e.g. if the response time for a module is exceptionally long the teacher could further investigate if the module was too difficult. Zorrilla et al. (2005) reports from the early stages of a project aiming at supporting the teachers' and the students' in an e-learning environment, but they also identify the possibility for the site administrator to use the parameters from the tool to adapt the site to users' behaviour in order to improve the site efficiency e.g. by optimizing the server size. Another example of using data mining is to predict how plausible it is that students enrolled in a community college will return to school after they have finished (Luan 2002). This information can then be used by the faculty for direct or indirect intervention (Luan 2002).

2 Case and method

The case for this paper is set in a workplace environment, but as the educational landscape is changing with a higher degree of students combining work and education, the results is relevant for the educational sector as well. The case organization is the county administrations of Sweden. The county administration is a government authority that ensures decisions made by the government and parliament is carried out locally and can be seen as the link between the people and the municipal authorities on the one hand and the government, parliament and other central authorities on the other. Tasks that the county administration are obligated to carry out include:

- implementing national objectives
- coordinating the different interests of the county
- promoting the development of the county
- establishing regional objectives

• safeguarding the rule of law in every instance

There are in total 21 counties in Sweden, and the initiative for a joint “Swedish Academy of County Administrations”, sprung from the experiences from having to educate most of their clerks on the consequences on new EU-legislation in the area of environmental management. New guidelines for how to process claims and errands relating to respect for the environment triggered all 21 counties (in parallel) to quickly develop training modules for their staff that could support the implementation of new work-practices. In the aftermath of this experience, representatives from some of the leading counties realized that collaboration in the area of competence development and staff-training could save substantial resources and at the same time utilize ICT for delivery and mediation of the training.

As a consequence, an action- and design-oriented research project was created with the purpose of understanding how technology and instruction could be designed to support joint e-training of clerks in the county administrations of Sweden. As a first step, a survey was conducted to explore what type of competence development the counties perceived they would need (short and long term perspective), and the following 2-by-2-matrix show a synthesis of four archetypical techno-pedagogical genres corresponding to their overall demands (Svensson, 2004).

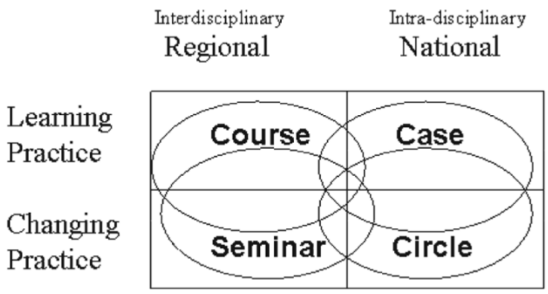

Figure 1. Design-space for work-integrated e-training (Svensson, 2004)

Secondly, the researchers decided to do a pilot-training module in the intersection between the genres ‘course’ and ‘seminar’ – An interdisciplinary topic that partly aimed at training new employees in existing practice, and partly aiming at improving how experienced clerks conducted their work. The topic chosen was “How to search digital resources, databases and the www” (Ostlund, 2012). The research was staged as a design science project with four iterative design cycles of (i) problem awareness, (ii) conceptual suggestion, (iii) system development, and (iv) evaluation (Vaishnavi & Kuchler, 2008). The data collected for problem awareness and evaluation of the different versions of the system drew on interviews (n=20) , questionnaires (n=25) and (for the purpose of this paper) also log-data on learners’ system use (small data).

Educational mining and learning analytics studies are mostly conducted in a school environment, but could in a similar way be helpful in an IT supported workplace learning environment (Svensson, 2004). However, when designing IT-support for workplace training, one must be sensible for the specific and situated nature of each societal sector, each organisation, and each individual.

Since the problem domain of IT supported workplace training is at the intersection of work, learning and IT, the approach of design science research is well suited. For IT to support workplace learning, the conditions of work needs to be taken into consideration. If any kind of designer chooses a scientific approach and process to designing, it will limit what and how things can be done and the eventual outcome itself (Nelson and Stolterman, 2012). Design science research puts a greater focus on the design research process, which makes design science research a less limiting approach and well suited for answering design-oriented research questions (Hevner et al. 2004).

One aspect that links the design science research theories and frameworks together are the stages of the research process such as problem formulation, suggestion for a solution to that problem, designing and developing the solution and evaluation of the artifact instantiation (e.g. Hevner et al., 2004; Vaishnavi and Kuechler, 2008). All these elements are represented in an article by Peffers et al., (2007) attempting to create a commonly accepted methodological framework for design science research. Based on a review of seven representative papers within design science, Peffers et al., (2007) present a design science research methodology that includes six steps: problem identification and motivation, definition of the objectives or a solution, design and development, demonstration, evaluation, and communication (Peffers et al., 2007).

Another aspect that design science research theories and frameworks have in common is that the artifact construction should be rooted in a kernel theory. The kernel theory for the pedagogical design of the IT artefact what that of authentic e-learning (Herrington et al., 2010). Authentic e-learning consists of nine pedagogical principles:

- Authentic Context
  Provide an authentic context that reflects the way the knowledge will be used in reality.
- Authentic Activities
  Provide an ill-defined authentic activity that encourages the students to find and solve problems.
- Expert Performance
  Provide access to expert performances by letting the student observe the task before they try themselves.
- Multiple Perspectives

Provide the learner with the opportunity to investigate multiple roles and perspectives.
- Collaboration
  Support the collaborative construction of knowledge.
- Reflection
  Promote reflection to enable abstractions to be formed.
- Articulation
  Promotes articulation to enable tacit knowledge to be made explicit.
- Coaching and Scaffolding
  Provides for scaffolding of support and coaching at critical times.
- Integrated Assessment
  Provide for integrated assessment of learning within the tasks.

(Herrington et al., 2010)

Given the context of e-learning, it is important that the outcome of the research activity not only contributes to the research domain of e-learning, but also has the potential of being relevant to practice. The result of a design science research project should be presented in a way that technicians understand how the artifact was created and evaluated and thereby knows how to use that in their own practice. In a similar way management should be able to evaluate if it is cost effective (Hevner et al., 2004). If research is intended to guide the design of future information systems, then we argue that design science research is a sound approach.

3 Findings

In a design science research project aiming at a design theory for IT supported training in the workplace, educational mining was used as a step to present a framework for workplace training. The learners' navigational pattern was used to extract behavioural patterns (Ostlund, 2012). The web logs using data pre-processing, pattern discovery and pattern analysis (Cooley et al. 1997). After grouping the session from the log data as the final step in the data pre-processing phase i.e. transaction identification (Marquardt et al. 2004) seven groups or interaction patterns emerged:

Table 1. Interaction patterns

Peeking (P)	Just visiting the home page of the WLA or just the start page for the video module and nothing more
One go (OG)	Going through all the modules in the course in one go
Partial order (PO)	Going through 2 or more video modules of the WLA, but in the intended order
Partial unordered (PU)	Going through 2 or more video modules of the WLA, but not in the intended order, ep.g. module 4 then module 7 and then module 3.
Single module (SM)	Just visiting one of the video modules
Mixed modules (MM)	Mixing e.g. streamed video modules with other modules like e.g. read more.
Non video modules (NV)	Browsing the non video modules e.g. read more and discussion.

The learner's activity peaks were also derived from the web logs. The activity statistics show that the most active day of the week was Thursday, although the activity was quite evenly spread throughout the workdays of the week with no activity during the weekends (Figure 2). On average over a day the activity was higher in the mid-morning than in the afternoon and peaked before and after lunch (Figure 3).

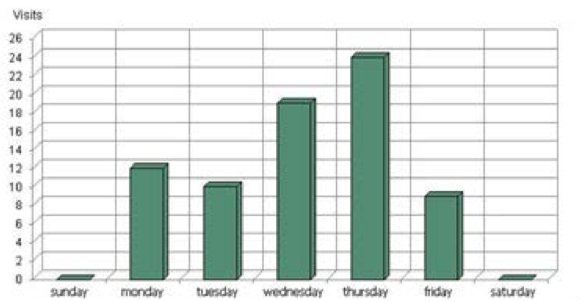

Figure 2. Visits by day of week

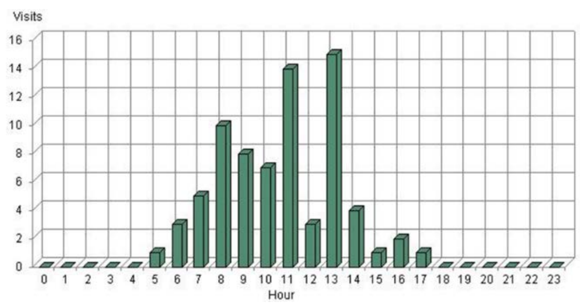

Figure 3. Visits by hour of day

The education mining data was complemented with interviews and a questionnaire. From the questionnaire it was e.g. noted that two thirds of the learners reported that they took breaks or was interrupted while taking the course and one third did not. This was further validated through the use of educational mining, where peeking and partial ordered or unordered use (Table 1) are evidence of interruptive use and these three categories made up more than 50% of the interaction patterns in table 1.

The four design research cycles has also through stepwise refinement of concepts and instantiations of the kernel theory of authentic e-learning (Herrington et al., 2010) resulted in a design theory of e-training through web lectures that adds to the prescriptive knowledge on formal e-training. The design theory consists of eight principles for workplace training Three of the design principles emerged from the design cycles, and the remaining five design principles originated from the kernel framework of authentic learning (Herrington et al., 2010), and were either confirmed as transferable to the context of workplace e-training as they are presented in the original framework or adapted to comply with the conditions of e-training in a workplace. The design theory is summarised in table 2 below:

Table 2. A design theory for work-integrated e-learning

Principle #I (Emergent)	The limitations and affordances of the organizational and infrastructural context must be carefully considered as a frame for the design of e-training.
Principle #II (Emergent)	Design to support for interruptive use, so that learners are supported in prompt repetition of previous use-sessions.
Principle #III (Adapted)	Design for explorative use so learners can create flexible and individual use-patterns.
Principle #IV (Adapted)	Design to promote learners' reflections with peers in their local workplace context.
Principle #V (Adapted)	Design for variation in participants' levels of expertise, so that scaffolding is integrated into the design.
Principle #VI (Confirmed)	Design for a learning environment that provides an authentic context, authentic activities and authentic tasks.
Principle #VII (Adapted)	Use experts that reflect the ethics and the culture of the organization as content providers.
Principle #VIII (Emerged)	Design for attention with mild reminders.

4 Conclusions

This paper adds to the ongoing discussion on small data vs. big data (e.g. Huang and Huang, 2015). In a blog post from 2013, Rufus Pollock, the director of the Open Knowledge Foundation, stated that the size of the data does not matter, what is important is having the data of whatever size it may be, as long as it helps solve a problem or address a question. This paper sets out to add to this discussion by describing how big data methods like learning analytics and web mining were used on small data to support the design process of a learning environment.

In the current body of research, learning analytics and web mining in educational contexts are to a large extent used to support the learners or the teachers (Zaiane 2002; Ueno and Nagaoka 2002; Khrib et. al 2008), but seldom to guide the overall design of the learning environment. Educational mining and learning analytics have great potential in supporting the learner (Zaiane 2002; Khrib et. al 2008) and/or the teacher (Ueno and Nagaoka 2002), but it is argued in this paper that there is also great potential to support the design process of the learning environment as well, not only how to make the learning environment more efficient (Zorilla et. al. 2005), but utilizing learning analytics on small data.

Ostlund (2012) used educational mining to extract design guidelines that will support the learners in their learning process. Learning analytics reveal actual use of an e-learning system in a way that is very hard if not impossible to do through other ways of gathering empirical data. Observation as a method for collecting data would be very time consuming due the anytime/anyplace character of e-learning. Interviews and surveys will tell you how the learners think they used the system, but learning analytics show how they actually use the system. It is therefore argued that it is important to acknowledge the value of applying learning analytics on small data.

When analysing big data the results can sometimes be too generic and to some degree lose its value. When designing for IT supported learning, context is crucial, and small data is less likely to lose contextual values. It is easier to interpret and bring the context into the analysis when the data set is scaled down. We do not argue that big data is not useful, but rather that small data (e.g. use-pattern log data from a small set of learners) also have a promising potential for extracting design guidelines for e-training and e-learning.

No matter if analysis is done on a small or big data set, the analyst plays a crucial role when deciding which data should be collected, picked, integrated, etc. (Huang and Huang, 2015). Examples of institutions that commonly have access to big data are industry, government, and research universities, whereas the small data rich institutions may be all around us on a more local level (Abreu and Acker, 2013). Although more research is needed, it could be argued that the analyst with access to small data may be more familiar with the data set and thereby perform a more relevant analysis than would be the case with large and more anonymous data sets.

References

Abreu, A., & Acker, A. (2013). Context and collection: A research agenda for small data. iConference 2013 Proceedings (pp. 549-554).

Chou, P. H., Li, P. H., Chen, K. K., & Wu, M. J. (2010). Integrating web mining and neural network for personalized e-commerce automatic service. Expert Systems with Applications, 37(4), 2898-2910.

Cooley, R., B. Mobasher, and J. Srivastava. Web mining: Information and pattern discovery on the world wide web. 1997.

Gašević, D., Zouaq, A., & Janzen, R. (2013). "Choose your classmates, your GPA is at stake!": The association of cross-class social ties and academic performance. *American Behavioral Scientist*, 0002764213479362.

Gellerstedt, M., Svensson, L., & Östlund, C. (2014). Navigating in a New Pedagogical Landscape with an Introductory Course in Applied Statistics. In *Topics from Australian Conferences on Teaching Statistics* (pp. 119-148). Springer New York.

Herrington, J., Reeves, T. C., and Oliver, R. (2010). A guide to authentic E-learning. New York: Routledge.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75-105.

Huang, P. C., & Huang, P. S. (2015). WHEN BIG DATA GETS SMALL. International Journal of Organizational Innovation (Online), 8(2), 100.

Hung, J. L. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British Journal of Educational Technology*, 43(1), 5-16.

Khribi, M. K., Jemni, M., & Nasraoui, O. (2009). Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval. *Educational Technology & Society*, 12 (4), 30–42.

Luan, J. (2002). Data Mining and Knowledge Management in Higher Education-Potential Applications.

Marquardt, C. G., Becker, K., & Ruiz, D. D. (2004). A pre-processing tool for Web usage mining in the distance education domain. Paper presented at the *Database Engineering and Applications Symposium*, *2004. IDEAS '04*. Proceedings. International.

Nelson, H. G., and Stolterman, E. (2012). The design way : intentional change in an unpredictable world. Cambridge, MA: MIT Press.

Ostlund, C. (2012). E-learning use patterns in the workplace–Web logs *from interaction with a web based lecture. International Journal of Advanced Corporate Learning (iJAC),* 5(4), 4-8.

Ota, K. & Kashihara, A. (2010). Reconstructing Web Resource with Learning History Mining. In J. Sanchez & K. Zhang (Eds.),Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2010 (pp. 2662-2670). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45-77.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert systems with applications, 33(1), 135-146.

Svensson, L. (2003). Less is more in distance education: the contradictory relationship between rapid adoption and radical innovation. *Educational Technology & Society*, 6(1).

Svensson, L. (2004). Challenges for work-integrated e-learning: The case of the Swedish Academy of County Administrations. *Journal of Workplace Learning*,*16*(8), 492-502.

Ueno, M., & Nagaoka, K. (2002). Learning Log Database and Data Mining system for e-Learning–On-Line Statistical Outlier Detection of irregular learning processes. In Proceedings of the International Conference on Advanced Learning Technologies (pp. 436-438).

Vaishnavi, V. K., & Kuechler , W. (2008). Design Science Research Methods and Patterns: Innovating Information and Communication Technology: Auerbach Pub.

Zaíane, O. R. (2002). Building a recommender agent for e-learning systems. In Computers in Education, 2002. Proceedings. International Conference on (pp. 55-59). IEEE.

Zorrilla, M., Millan, S., & Menasalvas, E. (2005, July). Data web house to support web intelligence in e-learning environments. In 2005 IEEE International Conference on Granular Computing (Vol. 2, pp. 722-727). IEEE.