Experiences from Digital Learning Analytics in Finland and Sweden: A Collaborative Approach

Mikko Apiola*, Thashmee Karunaratne†, Erkki Kaila*, and Mikko-Jussi Laakso*
*University of Turku, Finland, Department of Future Technologies
Email: firstname.lastname@utu.fi
†Stockholm University, Sweden, Department of Computer and Systems Sciences
Email: thasmee@dsv.su.se

Abstract—Digital learning management systems (LMS) are revolutionizing learning in many areas, including computer science education (CSE). They are capable of tracking learners’ characteristics, such as prior knowledge, and other learning habits, and may offer more personalized learning or guidance on useful learning practices. LMSs collect large amounts of data. Proper processing of such collected data can offer valuable insights about the learning process, support for higher quality education, insights on why some students drop out of courses, and so on. In this paper, we briefly review and discuss the global trends in digital learning and learning analytics (LA), specifically from the viewpoint of two LMS systems and related LA research, one in Finland and one in Sweden. In this paper, we address the context-, and course-specific nature of LA by developing the idea of cross-country and cross-systems learning analytics. Second, we consider our research especially from an educational perspective to identify the most beneficial practices for teachers and students. Third, we discuss, based on findings from our projects, future avenues for research.

Index Terms—learning analytics, digital learning

I. INTRODUCTION

Digital learning management systems (LMS) are revolutionizing learning in many areas. In computing, a vast array of learning management systems (LMS) are being used in teaching many topics, especially programming, in schools and universities, globally (eg. [1]). Modern LMS systems are often web-based, and they are labeled as adaptive, semi-intelligent, offering a range of learning content and supportive tools and services for learning content creation and management. LMS systems can track learners’ characteristics, such as prior knowledge [2], and other learning habits, and may offer more personalised learning [3] or guidance on learning practices that are beneficial for personalization. These modern approaches are already having major implications to education.

Learning management systems collect a lot of data, which is being analyzed by research under the titles of learning analytics (LA) or educational data mining (EDM) [1, 4]. Proper processing of such collected data can offer valuable insights about the learning process, support for increasing learners’ success, insights on learner’s dropping out, and so on. In addition to many opportunities of LA, there are challenges, too. Among the challenges is that a large amount of LA and EDM research in CSE has focused on simple metrics analysis and has been conducted in the context of a single institution or a single course (eg. [1]). Other challenges are how to use the collected data in the most beneficial ways, as well as ethical issues. A review, in 2014, concluded that: “To date, the complementary disciplines of learning analytics and educational data mining have focused predominantly on analyzing data systematically gathered in educational settings, which at the tertiary level includes factors of prior academic performance, demographic data, such as age and gender, and data gathered by logs recording student behaviour in online learning environments.” [4]. Moreover, it has been argued that important variables, such as ability, personality, motivation, self-regulation—while being well-proven predictors of academic achievement—are not much included in learning analytics [4]. In CSE, much of the analytics research is still done in the contexts of single courses and institutions (eg. [1]), lacking a holistic and generalisable approach to research.

In this paper we discuss the global trends in digital learning and LA from the viewpoint of two research groups on learning analytics, one in Sweden and one in Finland. With this paper, we have the following aims. This article aims to find answers to the following questions.

1) What are the most beneficial practices for teachers and learners in digital learning resulting from our research?
2) What are the most promising ongoing research approaches in our research efforts in learning analytics?

The rest of the paper is organized as follows. In Section [1] the background of digital learning in computer science (CS) is presented and our two educational systems are briefly described. Section [II] discusses the learning analytics research in our systems. Discussion about further possibilities for joint work and cooperation and influence of emergent development that e-learning offers to different educational stakeholders is given in Section [IV]. Section [IV-D] concludes the paper.

II. BACKGROUND: DIGITAL LEARNING IN CS

A. Digital Learning Overview (Stockholm)

The case study from Stockholm focuses on the department of Computer and Systems Sciences (DSV) at Stockholm university and its learning management system iLearn, which is an instantiation of the commercial tool Moodle, and the thesis management system SciPro [5].

iLearn is used at DSV to manage the course contents and student-teacher interactions in the courses offered to students.
As the unit that is responsible to support the university teachers to use digital tools in education, DSV manages all the online and on-campus courses in iLearn. DSV, at the moment, offers at least three Masters programs completely in online fashion. There are a significant number of students enrolled in 29 distance courses. In addition, there are 18 other bachelors and masters programs using iLearn to manage contents, submit assignments, peer collaboration, group activities as well as digital examination. Courses in iLearn are structured to follow a carefully chosen pedagogy, which is well-grounded in learning theories, and well tested evaluations (eg. [6], [7]).

SciPro is a system that is developed by DSV for managing students at bachelors and masters levels with a main purpose of finding relevant supervisors for students’ thesis projects based on interests of the supervisors and the needs of the students [3]. SciPro supports communication during many levels in the thesis process. SciPro answers to the needs to increase quality of thesis projects, and the rights of students to receive proper feedback in reasonable time [9]. SciPro facilitates peer communication during peer review process [10]–[12], student-supervisor communication [9], and efficient thesis writing [13]–[15]. SciPro brings flexibility to meetings between students, supervisors, and examiners to plan, discuss and exchange information efficiently [6], [10]. The quality control of theses is automatically done in SciPro with integrated plagiarism control [16], [17]. The use of SciPro has significantly increased performance in thesis writing [6].

B. Digital Learning Overview (ViLLE)

The development of ViLLE started in 2004 at the Department of Information Technology, University of Turku, Finland. The first edition was mostly a program visualisation tool. Over the course of time, through massive development efforts, ViLLE has grown into a full scale learning tool, with extensive digital tools for learning, tools for learning material creation, and tools for learning analytics (see [18] pp. 17–34).

Through the development lifecycle of ViLLE, research has increasingly become one pillar for development. This means that the functionalities and pedagogical approaches are evaluated through research. Through these efforts, a number of features that are beneficial for learning have been identified, which include automatic assessment, immediate feedback, support for a range of exercise types, and teaching modes such as pair programming [19], [20]. Positive results for learning and teaching have been demonstrated in numerous introductory computer science and basic mathematics courses both in university and high school courses [18], [20].

ViLLE has gained a lot of popularity in Finland [20]. Finnish school teachers, all of whom hold a higher university degree (master’s level) in education, are autonomous in selecting learning tools of their preference, which means that teachers must perceive digital learning as significantly beneficial in order for them to use it. In schools, ViLLE is currently being used in teaching of a wide range of topics. ViLLE is actively used in a large number of courses at schools, universities, and polytechnics institutes in Finland. In universities, ViLLE is being used in various computer science courses [20], [21]. Using ViLLE as compared to traditional instruction has improved learning outcomes in programming courses in high school [18], [39–41] and in various CS courses at university [18], [41–42], [18], [22], [23].

III. Thematic Overview of Research

A. Learning analytics research (Stockholm)

The learning analytics group at DSV with its extensive networks with research organisations, industry, and society, conduct research both with its digital platforms (iLearn and SciPro), but also extensively with respect to other cases and systems nationally and internationally. The research is grounded in learning theories and aims in using learning analytics and artificial intelligence in enhancing education. Together with the researchers and other stakeholders who share the interest in tracing learning, understanding learning, and improving learning with the support of LA and AI, DSV’s LA research broadly covers advances in learning analytics and artificial intelligence for education, including but not limited to the topics of tools and methodologies for LA for education, applications of LA for education in real-world settings, and, theoretical perspectives on LA for education. More specifically, our research focuses on the following themes, covering education from primary school to higher education.

1) Learners’ activity, interactivity, engagement and motivation: Learning management systems (LMS) harbor a lot of data, but until recently that data has been a largely unused source of information for pedagogical decision making. Relying on the log data in LMS for decision making on the other hand is challenging due to the variations of the information richness of this data. Understanding learning engagement as part of learning activities in online courses is one of our research tracks [5], [24]–[26]. For example, in one masters course for research methods in computer science [7] a clear difference between the click logs of high and low performing students is shown (see Figure I).

**Figure 1.** Average clicks of excellent, good, fair and failed students.

**Figure 2.** shows that there are significant differences of behavioural patterns of high achieving and low achieving students. For instance, students who fail to complete the course have the minimal login in the LMS in the beginning of the course. Such indications perhaps could be an alarm for the teacher for early interventions on the course for reducing the
failure rate. Our further research on the matter [26–28] have shown particular applications of log data to measure learning scenarios, such as lifelong learning, self-regulated learning and collaborative learning.

2) Evaluate, inform and develop current teaching practices and learning designs: Another important line of research has been support in thesis process with increasing communication and peer learning [6], [24]–[26]. This line of research has investigated how to design the thesis process to support increased communication and peer learning [6], how students’ temporal behaviors in digital learning platforms can predict performance and learning and form the basis for formative feedback [24], how learning analytics can be used to better understand and through feedback to support collaborative learning [25], and, how the engagement in the LMS throughout a span of a professional development course explains the throughput of the course [26].

3) Technical applications: machine learning, data mining, predictive analytics, social network analysis and network modelling: All the LA studies conducted at DSV follow at least one of these techniques. The use of complex machine learning techniques for building predictive models that explain the performance of prospective students based on data of previous courses has been demonstrated in studies [27], [29], [30]. In [29] flipped classroom contexts to predict low and high-performing students in a part to provide automatic formative feedback using machine-learning methods is discussed. With a pre-hypothesis that social network analysis may be of significant value in studying online collaborative learning, in one research [27] student collaboration in a digital learning platform was investigated. The results show that social network analysis can enhance the understanding of the collaborative process, predict the under-achievers and uncover the dynamics of students’ interactions [27]. The results further show that some students took teacher or instructor roles in the collaboration process, and that social network analysis was powerful in classifying students according to their achievements [27].

4) Improve feedback practices and support personalized learning: Analysis of behavioral patterns of students can inform pedagogical and institutional decisions [30]. On the level of classroom, patterns in learners’ behavior reflected in the data can inform the feedback teachers provide to learners, this feedback, in turn, can inform a conscious effort in learners to improve their studying habits and learning strategies. Our studies in this line have shown that systematic employment of formative feedback can support self-regulated learning [24]. Also, we are planning a research study to build on our previous work by extending self-regulation through formative feedback mechanisms. The plan is to explore possibilities of gaining insights through adopting a systemic and holistic approach, i.e. through collecting and analyzing data from multiple sources that reflect learning that takes place within as well as outside digital learning platforms.

5) Modelling learners interactivity in learning and teaching in blended environments: Almost all the third cycle courses at DSV are offered in blended fashion where, the content and some learning activates are taking place at a digital environment complemented by face to face teaching sessions. LMS record a large amount of data, i.e. traces of learning processes, which allows to investigate the learning processes in-depth. However, such approach has a major deficiency that it does not count on the behaviours outside the learning environment and focus exclusively on one data source (that of LMS), thus, contextual factors are not considered [31]. The focus on a single data source is especially problematic when trying to understand learning processes that take place in blended learning environments, i.e. in courses that combine offline and online learning, where the processes in both environments are intertwined and affect one another, including on the aspect of regulation of learning. The assumption that learning occurs within the learning platforms only has been pointed out as the main reason for LA research findings to lack in generalizability and transferability [31]. Thus, we argue that an analytical and methodological shift is needed, from the digital platform as the unit of analysis, which is the current practice, to the blended learning environment as the primary unit of analysis.

B. Learning analytics research (Turku)

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>THREE CLASSES OF LEARNING ANALYTICS IN ViLLE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(1) Experimental (pre-post) for comparing features and pedagogies</strong></td>
<td>Research that uses experimental pre-post research designs, which looks at the most efficient features or pedagogies in developing ViLLE. The mostly used dependent variables are learning performance measured by course grades and assignment scores.</td>
</tr>
<tr>
<td><strong>(2) Learning metrics (live learning data and achievements)</strong></td>
<td>Research that uses variables such as assignment submission times, prior skills, numbers of assignment submissions, log data, attendance data in learning sessions, attendance data in groupwork sessions, and patterns of code writing behavior in programming courses. Possible associations such as causalities are investigated by using statistical methods.</td>
</tr>
<tr>
<td><strong>(3) Psychometrics</strong></td>
<td>This class includes research that uses psychometric measurement instruments to classify learners’ behavior, motivation, problem solving abilities, attitudes, mindsets, or orientations to learning, and finds statistical associations with the measurements and other variables such as learning performance.</td>
</tr>
</tbody>
</table>

This section gives an overview about the learning analytics research conducted around the ViLLE. That research can
be roughly divided into three classes: 1) experimental pre-post setups, 2) analysing technical and log data, and 3) psychometrics. These three classes are summarised in Table I and explained in detail in the related forthcoming three subsections.

1) **Experimental Research for Comparing Learning Performance:** ViLLE is the result of 15 years of development. One of the cornerstones of success is that the functionalities and pedagogies are evaluated through research methodology [18]–[20]. Thus, this class of research forms one of the founding pillars of the development work, and includes research that evaluates courses and features. Examples of evaluation research include looking at the impact of visualisation exercises in high school programming [18] pp 35–39], looking at programming courses at university level [18] 41–42], and at university level for computer science minors [18] 43–44], and computer science majors [18] 45–47], as well as object oriented programming courses [18] 48–49], and introduction to computer science courses [22]. This class of research has helped to distill the most beneficial features in ViLLE, such as active learning and continuous assessment, heterogeneous exercise types, electronic examining [32], tutorial-based learning, and continuous feedback [18]. Perhaps most importantly, it has given evidence about the positive impact of digital learning for learning outcomes on a range of courses and topics. The research methods in this class are mostly pretest-posttest experimental setups, which use the pedagogical approach (such as tutorial-based learning versus traditional instruction) as the independent variable.

2) **Learning Metrics:** This class of research contains research that investigates statistical associations in the domain of a number of variables, which can be divided to live learning data and achievement data. The data includes variables, such as submission exercise times, number of exercise submissions, log data, and attendance data. Examples of research designs include looking at associations between students’ time usage and their course grades [21], 33], issues in regards of collaboration patterns seen in log data [34]–[36], finding associations in students’ previous knowledge and skills—as measured by course grades in previous courses—and course performance [2], and misconceptions in programming tasks by looking at the exercise completion patterns [2]. The data used in this class can be considered mostly as “technical”. The research methods include a range of measures of statistical association, depending on case.

3) **Psychometrics:** This class of research includes the use of well-tested and validated psychometrics instruments in order to gain deeper understanding about the various processes of learning. For example, while research in Class 1 (comparing pedagogies) has proven that a number of ViLLE-based pedagogical approaches outperform other, “traditional” approaches to instruction, it is currently not well understood why this is. The more deeper psychometric understanding about learning can provide important insights about a number of issues, such as learning deficits, special needs, talent, and groupwork. An overview of the research in ViLLE-project that falls under this class is summarised in Table II which divides the research into completed, in progress, and future research.

Table II shows the current psychometrics research in ViLLE. In regards of Problem Solving Inventory (PSI) [37], it has been investigated in the contexts of University of Turku courses on introductory programming and basic course in algorithms. The current results show that PSI scores can be used to predict students’ final exam performance, to a certain extent [38]. Also, it has been shown that by using a PSI-based machine learning model, it was possible to predict the learning performance in a test set by the accuracy of 80.01%. PSI was also tested in identifying students at risk of dropping out. Other research has used the The Nature of Attitudes Towards Learning Mathematics Questionnaire (NALMQ) [39], and the MAKEKO [40], which is a well established Finnish test for identifying students’ threshold concepts in mathematics as part of learning analytics within ViLLE [41]. In regards of Mindset [43], research is underway [44].

As can be seen from Table II the psychometrics research in ViLLE team is currently still at an early stage. Psychometrics measures and constructs with proven potential are listed in the future research-section of Table II. This list provides measures that are well-tested and many of which have proven power to predict factors such as academic achievement. Research based on the listed measures is currently being planned.

4) **Conclusions:** This section provided an overview of learning analytics research in ViLLE. During the extensive development process of ViLLE, spanning over 15 years, learning analytics research has started with comparing features and pedagogies in ViLLE by experimental settings (Class 1) towards analysis of a variety of technical and log data (Class 2), and is now expanding also to psychometrics research (Class 3). The current research in regards of psychometrics is relatively small in scale, considering the possibilities provided by extensive research in psychometrics, and the extensive pos-
sibilities that digital learning and the ViLLE system provides for data collection.

IV. DISCUSSION AND CONCLUSIONS

This section discusses the most important findings from our research. We align our discussions with the questions that we set in the introduction of this article.

A. Concrete benefits of digital learning

First, significant increases in learning outcomes in various courses have resulted from introduction of digital learning to teaching. Second, proven benefits to recognise students in risk of dropping out have resulted in new models of student counselling, with proven positive effects to students’ learning outcomes. Third, the introduction of the LMS systems have resulted in significant benefits to teachers, and development of new pedagogical models such as tutorial-based learning, and peer-based instruction. Fourth, identification of students’ learning styles and preferences in learning has helped to individualise the learning paths of students.

B. Most promising research approaches

LA research require a major shift from one dimensional single unit case study into a holistic approach that allow capturing and triangulation of data of teaching and learning derived by several sources. The necessary first step would be to cover both the physical and digital environments where learning occur. What data to be curated, and how, are the exact next steps to be performed. Understanding the indicators for learning in regards of benchmarking and recognizing LA are of crucial importance. Important avenues for future research also include school computing education [46], computational thinking, and bringing innovation-friendly teaching [47] into digital learning.

C. Challenges in research on digital learning

In our digital world, learning is not happening in an isolated unit. Inter-dependencies of many factors influence the end results of learning. The big question is now, which data is best in explaining learning, and how highly contextualised LA models can be generalised to allow wider applicability.

Also, there is a very thin line between use of learner data for supporting learning and use of the same data for surveillance purposes. Building the trust between the data owners and data users is challenging. Privacy has become more important than ever before since the new GDPR strengthens the ownership of data. If LA is to become a useful endeavour, richness of the data is the key. The debate about how to conquer this challenge is ongoing, but yet to reach into its success.

The new rules for data protection imposed by the European Union (GDPR) have given people rights to control who they expose their personal data to. The essential consequences and implications of such implementations of data regulations are that the stakeholders in education can, at any point be opted out from allowing LA on their data. We are interested in exploring the perceptions of students, teachers, schools and universities, parents and the society on use of data about learning and the environment where learning occurs.

D. Conclusions

In this article we have reviewed the central results from development and research within two digital learning systems that are extensively used in digital education. The positive impact of digital learning, automatic assessment, continuous feedback, visualisation, and other features have been shown, and the introduction of digital learning systems to education has been clearly shown to be a good idea. By looking at the research of two widely used systems in two different countries, we are building the possibility for inter-system collaboration and research where features and approaches are simultaneously tested within more than one system. By doing this, we are building grounds for a European network of learning analytics in CSE research, where approaches can be tested and data can be collected from several contexts, significantly strengthening the research results. In the long term, we believe that this approach will lead to increased opportunities for comparing and testing various new pedagogies and approaches, collecting and comparing data obtained from different institutions, making it possible to replicate and conduct experiments in two or more contexts, increasing the validity of the results.

REFERENCES
