Computers in Human Behavior 47 (2015) 157–167

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

In search for the most informative data for feedback generation: Learning analytics in a data-rich context

Dirk T. Tempelaar [a,*], Bart Rienties [b], Bas Giesbers [c]

[a] Maastricht University, School of Business and Economics, PO Box 616, 6200 MD Maastricht, Netherlands
[b] Open University UK, Institute of Educational Technology, UK
[c] Rotterdam School of Management, Erasmus University, Netherlands

ARTICLE INFO

Article history:
Available online 28 June 2014

Keywords:
Blended learning
Dispositional learning analytics
e-Tutorials
Formative assessment
Learning dispositions

ABSTRACT

Learning analytics seek to enhance the learning processes through systematic measurements of learning related data and to provide informative feedback to learners and teachers. Track data from learning management systems (LMS) constitute a main data source for learning analytics. This empirical contribution provides an application of Buckingham Shum and Deakin Crick's theoretical framework of dispositional learning analytics: an infrastructure that combines learning dispositions data with data extracted from computer-assisted, formative assessments and LMSs. In a large introductory quantitative methods module, 922 students were enrolled in a module based on the principles of blended learning, combining face-to-face problem-based learning sessions with e-tutorials. We investigated the predictive power of learning dispositions, outcomes of continuous formative assessments and other system generated data in modelling student performance of and their potential to generate informative feedback. Using a dynamic, longitudinal perspective, computer-assisted formative assessments seem to be the best predictor for detecting underperforming students and academic performance, while basic LMS data did not substantially predict learning. If timely feedback is crucial, both use-intensity related track data from e-tutorial systems, and learning dispositions, are valuable sources for feedback generation.

1. Introduction

Learning analytics provide institutions with opportunities to support student progression and to enable personalised, rich learning (Bienkowski, Feng, & Means, 2012; Oblinger, 2012; Siemens, Dawson, & Lynch, 2013; Tobarra, Robles-Gómez, Ros, Hernández, & Caminero, 2014). With the increased availability of large datasets, powerful analytics engines (Tobarra et al., 2014), and skillfully designed visualisations of analytics results (González-Torres, García-Peñalvo, & Therón, 2013), institutions may be able to use the experience of the past to create supportive, insightful models of primary (and perhaps real-time) learning processes (Rienties, Slade, Clow, Cooper, & Ferguson, submitted for publication; Baker, 2010; Stiles, 2012). According to Bienkowski et al. (2012, p. 5), "education is getting very close to a time when personalisation will become commonplace in learning", although several researchers (García-Peñalvo, Conde, Alier, & Casany, 2011; Greller & Drachsler, 2012; Stiles, 2012) indicate that most institutions may not be ready to exploit the variety of available datasets for learning and teaching.

Many learning analytics applications use data generated from learner activities, such as the number of clicks (Siemens, 2013; Wolff, Zdrahal, Nikolov, & Pantucek, 2013), learner participation in discussion forums (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014; Macfadyen & Dawson, 2010), or (continuous) computer-assisted formative assessments (Tempelaar, Heck, Cuypers, van der Kooij, & van de Vrie, 2013; Tempelaar, Kuperus et al., 2012; Wolff et al., 2013). User behaviour data are frequently supplemented with background data retrieved from learning management systems (LMS) (Macfadyen & Dawson, 2010) and other student admission systems, such as accounts of prior education (Arbaugh, 2014; Richardson, 2012; Tempelaar, Niculescu, Rienties, Giesbers, & Gijselaars, 2012). For example, in one of the first learning analytics studies focused on 118 biology students, Macfadyen and Dawson (2010) found that some (# of discussion messages posted, # assessments finished, # mail messages sent) LMS variables but not all (e.g., time spent in the LMS) were useful predictors of student retention and academic performance.

* Corresponding author. Address: Department of Quantitative Economics, Maastricht University, School of Business and Economics, PO Box 616, 6200 MD Maastricht, Netherlands. Tel.: +31 43 38 83858; fax: +31 43 38 84874.

E-mail addresses: D.Tempelaar@MaastrichtUniversity.nl (D.T. Tempelaar), Bart.Rienties@open.ac.uk (B. Rienties), bgiesbers@rsm.nl (B. Giesbers).

http://dx.doi.org/10.1016/j.chb.2014.05.038
0747-5632/

Buckingham Shum and Deakin Crick (2012) propose a dispositional learning analytics infrastructure that combines learning activity generated data with learning dispositions, values and attitudes measured through self-report surveys, which are fed back to students and teachers through visual analytics. For example, longitudinal studies in motivation research (Järvelä, Hurme, & Järvenoja, 2011; Rienties, Tempelaar, Giesbers, Segers, & Gijselaers, 2012) and students' learning approaches (Nijhuis, Segers, & Gijselaers, 2008) indicate strong variability in how students learn over time in face-to-face settings (e.g., becoming more focussed on deep learning rather than surface learning), depending on the learning design, teacher support, tasks, and learning dispositions of students. Indeed, in a study amongst 730 students Tempelaar, Niculescu, et al. (2012) found that positive learning emotions contributed positively to becoming an intensive online learner, while negative learning emotions, like boredom, contributed negatively to learning behaviour. Similarly, in an online community of practice of 133 instructors supporting EdD students, Nistor et al. (2014) found that self-efficacy (and expertise) of instructors predicted online contributions.

However, a combination of LMS data with intentionally collected data, such as self-report data stemming from student responses to surveys, is an exception rather than the rule in learning analytics (Buckingham Shum & Ferguson, 2012; Greller & Drachsler, 2012; Macfadyen & Dawson, 2010; Tempelaar et al., 2013). In our empirical contribution focusing on a large scale module in introductory mathematics and statistics, we aim to provide a practical application of such an infrastructure based on combining longitudinal learning and learner data. In collecting *learner data*, we opted to use three validated self-report surveys firmly rooted in current educational research, including learning styles (Vermunt, 1996), learning motivation and engagement (Martin, 2007), and learning emotions (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). This operationalisation of learning dispositions closely resembles the specification of cognitive, metacognitive and motivational learning factors relevant for the internal loop of informative tutoring feedback (e.g., Narciss, 2008; Narciss & Huth, 2006). For *learning data*, data sources are used from more common learning analytics applications, and constitute both data extracted from an institutional LMS (González-Torres et al., 2013; Macfadyen & Dawson, 2010) and system track data extracted from the e-tutorials used for practicing and formative assessments (e.g., Tempelaar et al., 2013; Tempelaar, Kuperus, et al., 2012; Wolff et al., 2013). The prime aim of the analysis is predictive modelling (Baker, 2010; Sao Pedro, Baker, Gobert, Montalvo, & Nakama, 2013), with a focus on the roles of (each of) 100+ predictor variables from the several data sources can play in generating timely, informative feedback for students.

2. Literature review

2.1. Learning analytics

A broad goal of learning analytics is to apply the outcomes of analysing data gathered by monitoring and measuring the learning process (Buckingham Shum & Ferguson, 2012; Siemens, 2013). A vast body of research on student retention (Credé & Niehorster, 2012; Marks, Sibley, & Arbaugh, 2005; Richardson, 2012) indicates that academic performance can be reasonably well predicted by a range of demographic, academic integration, social integration, psycho-emotional and social factors, although most predictive models can explain only up to 30% of variance. Recent studies in learning analytics (Agudo-Peregrina et al., 2014; Macfadyen & Dawson, 2010; Tempelaar et al., 2013; Wolff et al., 2013) seem to indicate that adding LMS user behaviour to these models can substantially improve the explained variance of academic performance. However, according to Agudo-Peregrina et al. (2014) there is no consensus in the learning analytics community on which user behaviour and interactions data are appropriate to measure, understand and model learning processes and academic performance.

Clow (2013, p. 692) argues that "as a field, learning analytics is data-driven and is often atheoretical, or more precisely, is not explicit about its theoretical basis". Although several researchers have worked to link learning analytics to pedagogical theory (Clow, 2013; Dawson, 2008; Macfadyen & Dawson, 2010; Suthers, Vatrapu, Medina, Joseph, & Dwyer, 2008), this is still the exception, rather than the rule. However, Macfadyen and Dawson (2010, p. 597) note that "knowledge of actual course design and instructor intentions is critical in determining which variables can meaningfully represent student effort or activity, and which should be excluded". For example, Tempelaar et al. (2013) found empirical evidence for the role of a broad range of learning dispositions in learning analytics applications in a study amongst 1832 students. Demographic characteristics, cultural differences, learning styles, learning motivation and engagement, and learning emotions, all proved to be facets of learning dispositions having a substantial impact on learning mathematics and statistics. This study extends the analysis of predictive modelling for generating learning feedback by looking at the role of any data source in a multivariate context, so in the presence of several alternative data sources.

In Verbert, Manouselis, Drachsler, and Duval (2012), six objectives are distinguished in using learning analytics: predicting learner performance and modelling learners, suggesting relevant learning resources, increasing reflection and awareness, enhancing social learning environments, detecting undesirable learner behaviours, and detecting affects of learners. Although the combination of self-report learner data with learning data extracted from e-tutorial systems (see below) allows us to contribute to at least five of these objectives of applying learning analytics (as described in Narciss & Huth, 2006) (as described in Narciss & Huth, 2006), we will focus in this contribution on the first objective: predictive modelling of performance and learning behaviour (Baker, 2010; Sao Pedro et al., 2013). The ultimate goal of this predictive modelling endeavour is to find out which components from a rich set of data sources best serve the role of generating timely, informative feedback and signalling risk of underperformance.

2.2. Formative testing and feedback

A classic function of testing is that of taking an aptitude test. After completion of the learning process, we expect students to demonstrate mastery of the subject. According to test tradition, feedback resulting from such "classical" tests are typically limited to a grade (Boud & Falchikov, 2006; Whitelock, Richardson, Field, Van Labeke, & Pulman, 2014). Another limitation of classical summative testing is that feedback becomes available only after finishing all learning activities (Segers, Dochy, & Cascallar, 2003). An alternative form of assessment, formative assessment, has an entirely different function: that of informing student and teacher (Segers et al., 2003). This information should help to better shape teaching and learning and is especially useful when it becomes available prior to or during the learning process. Feedback plays a crucial part to assist regulating learning processes (Boud & Falchikov, 2006; Hattie, 2009; Lehmann, Hähnlein, & Ifenthaler, 2014; Whitelock et al., 2014). Several alternative operationalisations to support feedback are possible. For example, using two experimental studies with different degrees of generic and directed prompts, Lehmann et al. (2014) found that directed prereflected prompts encourage positive activities in online environments. In a meta-study of 800+ meta-studies, Hattie (2009) found that the way students receive feedback was one of the most powerful factors in enhancing learning experiences. Diagnostic