Internet and Higher Education 27 (2015) 54–63

Contents lists available at ScienceDirect

Internet and Higher Education

Predicting course outcomes with digital textbook usage data

Reynol Junco [a,b,*], Candrianna Clem [c]

[a] School of Education, Human Computer Interaction, Iowa State University, 2628 Lagomarcino Hall, Ames, IA 50011, United States
[b] Berkman Center for Internet and Society, Harvard University, 23 Everett Street, 2nd Floor, Cambridge, MA 02138, United States
[c] Department of Sociology, The University of Texas at Austin, Austin, TX 78712, United States

ARTICLE INFO

Article history:
Accepted 5 June 2015
Available online 11 June 2015

ABSTRACT

Digital textbook analytics are a new method of collecting student-generated data in order to build predictive models of student success. Previous research using self-report or laboratory measures of reading show that engagement with the textbook was related to student learning outcomes. We hypothesized that an engagement index based on digital textbook usage data would predict student course grades. Linear regression analyses were conducted using data from 233 students to determine whether digital textbook usage metrics predicted final course grades. A calculated linear index of textbook usage metrics was significantly predictive of final course grades and was a stronger predictor of course outcomes than previous academic achievement. However, time spent reading, one of the variables that make up the index was more strongly predictive of course outcomes. Additionally, students who were in the top 10th percentile in number of highlights had significantly higher course grades than those in the lower 90th percentile. These findings suggest that digital textbook analytics are an effective early warning system to identify students at risk of academic failure. These data can be collected unobtrusively and automatically and provide stronger prediction of outcomes than prior academic achievement (which to this point has been the single strongest predictor of student success).

1. Introduction

Learning analytics is the “measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (Siemens, 2010). Learning analytics projects collect and assess student-produced data in order to predict educational outcomes with the goal of tailoring education. In other words, learning analytics is an application of big data and predictive analytics in educational settings (Ferguson, 2012; Manyika et al., 2011). These methods do not involve direct input from students but instead use data collected unobtrusively from their use of educational technologies such as a university's learning and course management systems (LCMS) (Mattingly, Rice & Berge, 2012; Verbert, Manouselis, Drachsler & Duval, 2012). Data are collected and analyzed in real-time giving educators the ability to identify students at risk of academic failure (Picciano, 2012). Such predictive modeling is the ultimate form of student formative assessment—educators can have information about how a student might fare in their courses even before the student submits gradable work.

Campbell and Oblinger (2007) identified five steps of the learning analytics process:

1. *Capturing*: Data are captured from real-time learning analytics systems and combined with student information, such as demographic or background information (Ferguson, 2012). Specifically, data are collected in real-time from virtual learning environments (VLEs), learning management systems (LMSs), virtual machines, personal learning environment (PLEs), web portals, intelligent tutoring systems (ITS), forums, chat rooms and email (Mattingly, Rice & Berge, 2012).
2. *Reporting*: Data collected in the first step is reported to key individuals, identifying trends and patterns in the data to generate accurate models for measuring student progress and success. In order for the data to be reported in a coherent manner, it must be translated from raw data to comprehensive information through theoretical constructs, algorithms, and weightings (Greller & Drachsler, 2012). Oftentimes, the transformed data are visualized using learning analytics dashboards so that they can be more easily understood.
3. *Predicting*: A key affordance of learning analytics is using data to identify predictors of student success and to create models to predict student outcomes and identify at-risk students. In addition, predictive modeling is used to examine real-time learning in the classroom, to predict and make decisions about course and resource allocation, and to inform institutional decision-making (Ferguson, 2012).
4. *Acting*: Instructors and students act on the information discovered through learning analytics, for instance by intervening when a student is not doing well.
5. *Refining*: Ideally, as information is gathered, processed, and evaluated, learning analytics are refined through a cyclical process and continuous improvement model (Ferguson, 2012).

* Corresponding author at: 2628 Lagomarcino Hall, Ames, IA 50011, United States.
E-mail address: rey.junco@gmail.com (R. Junco).

http://dx.doi.org/10.1016/j.iheduc.2015.06.001
1096-7516/

While learning analytics show great promise for education, presently these tools and projects collect limited data on student academic work. For instance, most learning analytics tools collect information solely from the LCMSs and restrict the data collected to number of times students log on to the system, number of posts in online discussions, and assignment grades (Dawson, McWilliam & Tan, 2008; MacFayden & Dawson, 2012). The Purdue Signals learning analytics project mines data from the student information system and LCMS to flag at-risk students by posting a traffic signal indicator (i.e., red for high risk, yellow for moderate risk, and green for low risk) on the LCMS homepage (Arnold & Pistilli, 2012). Based on the results of the system's predictive student success algorithm, an intervention schedule is created consisting of automated email message or reminders, text messages, referral to an academic advisor, or a referral for face to face meetings with the instructor. Such a system helps instructors become aware of students who might be at risk of failure and helps focus their limited time on the students who need help.

Predictive models based on LCMS data show that activity on these sites is related to course outcomes. For instance, Arnold and Pistilli (2012) found that students in courses that used the Purdue Signals system earned better course grades and were retained at higher rates. Macfayden and Dawson (2012) found that number of discussion messages posted, read, and replied to positively predicted final course grades. Student use of LCMS-based course content and visits to the online grade book were also positively correlated with final course grade (Dawson, McWilliam & Tan, 2008). Smith, Lange, and Huston (2012) found that LCMS log-in frequency, site engagement (activities such as viewing the course syllabus, viewing an assessment, and completing an assessment), and points submitted were all correlated with successful course outcomes. Also, the more a student performs a certain course activity the better they will score in that area as course outcomes are predicted with increased accuracy as more activity and grade information accumulate over the duration of a course (Smith, Lange, & Huston, 2012).

Learning analytics applications to this point have focused on collecting data from LCMSs, although newer methods allow for collection of broader types of student-generated data (Junco, 2013a; Junco, 2013b). For instance, using monitoring software to collect data on student use of social media might predict how well they will perform in a course (Junco, 2013a; Junco 2013b). Furthermore, the predictive ability of models based solely on LCMS data is overestimated because they relate graded activities (like number of discussion board posts) to course grades. The relationship between discussion board activity and course grade should be significant because students are being graded on such activity.

1.1. Digital textbooks

Digital books continue to be a fast-growing sector of the publishing market. The Association of American Publisher's (2014) data on trade books found that there was an increase of 43% in ebook sales between 2011 and 2013. Youth digital book usage has grown substantially since 2011. Nielsen (2014) reports that while only 13% of children less than 12 years old read ebooks on a tablet in 2011, 34% did so in 2014. These numbers have increased since 2011 with Pew Internet Project data showing that more adults and youth own ereading-capable devices (such as tablets and dedicated ereaders) and read ebooks than ever before (Rainie, 2012; Zickuhr & Rainie, 2014).

While adoption of digital textbooks in higher education is growing, there are mixed findings when looking at student preferences. Longhurst (2003) found that 64% of students preferred printed materials to online texts, and that almost all students printed out materials when they had the option. Among both graduate and undergraduate students, 92% of students printed out materials when working with someone else, over 80% of students printed out materials if they were long or complicated or they wanted to study from them and 75% of students printed out materials if they wanted to take notes (Sandberg, 2011; Spencer, 2006). The reason students give for choosing print material over online text is because of difficulty reading from a screen (Rockinson-Szapkiw, Courduff, Carter & Bennett, 2013), that it is easier to concentrate when using paper than on a screen, and that it is easier to remember and understand information with paper than with online text (Le Bigot & Rouet, 2007). Dennis (2011), on the other hand, found that 36% to 84% of students preferred a digital textbook depending on the course; however, the strongest factor predicting preference was whether the instructor made substantial use of the digital text (assigning readings, making annotations, and referring to the book in class), instead of solely using it as a reference. Previous experience using a digital textbook for class also influenced student preference. In a related study, Weisberg (2011) provided students with tablets, e-readers, and print textbooks. If given a choice, 87% of students said they would use a digital textbook. Students reported that lower cost, convenience, and the ability to easily search book content were positive factors for adoption; while reporting that it was easier to concentrate on and comprehend paper textbooks as negative factors for adoption (Weisberg, 2011).

Capitalizing on the growth of ereading-capable devices and the ubiquitousness of learning and course management systems, textbook publishers have invested a great deal of resources into developing and promoting digital textbooks (Young, 2013). Newer versions of digital textbooks are intended to serve not just as texts for the course, but also as primary sources of course content. Indeed, some have argued that the publishers may soon control not just the textbook material but the course content as well (Young, 2013). In their push to increase the interactivity and usefulness of digital textbooks, publishers have included interactive content such as dynamic quizzes that feed results back into LCMS grade books (Young, 2013). Such integration will become more commonplace as textbook companies have moved to acquire educational technology companies, like those that develop LCMSs. In addition to acquiring companies that develop platforms like LCMSs, textbook companies are acquiring adaptive learning and learning analytics startups. For example, in 2013 *Pearson* acquired *Learning Catalytics*, a company that uses predictive analytics to help provide feedback to faculty and improve student engagement (New, 2013).

The merger of textbook companies with LCMS, adaptive learning, and learning analytics products hints at the future of digital textbooks. Indeed, companies have developed textbooks that are not only intended to help students become more engaged, but that can track that engagement much in the same way that LCMSs use learning analytics to track and predict student success. The advent of digital textbooks, then, affords educators the opportunity to unobtrusively collect learning analytics data from student use of reading materials. While digital textbooks have had the ability to collect usage data, only recently have technology companies started to develop methods to use these data to predict student outcomes (Bossaller & Kammer, 2014).

Textbook analytics is an emerging sub-category of learning analytics and such applications can not only provide additional data to existing learning analytics models but can serve as stand-alone predictors of student success, for engagement with digital textbooks should be predictive of course outcomes. For instance, computer science students who completed more exercises in a digital text scored better on written exams in the course (Fouh, Breakiron, Hamouda, Farghally, & Shaffer, 2014). Fouh et al.'s (2014) data show that while students did not read the static text, they did engage with the digital text's interactive elements, even when such engagement was ungraded.

1.2. Reading and student performance

Unsurprisingly, reading a textbook is directly related to course outcomes (Daniel & Woody, 2013). For instance, Landrum, Gurung, and Spann (2012) found that the self-reported percentage of reading completed in a textbook was positively correlated with quiz scores, total course points, and final course grades. On average, students spend less