Abstract: Learning analytics tools help online educators visually extract meaningful performance and behavioral patterns from learners’ trace data. While many learning analytics solutions have addressed how educators monitor and provide summative feedback to learners, most are pedagogically neutral, and do not feature or support formative feedback. We explore the research principles underlying the design and implementation of a learning analytics tool rooted in theories of self-regulation, formative feedback and design-based research that address challenges unique to online educators and learners. The tool, itself a source of formative feedback, is intended to improve educator efficacy and the provision of timely feedback, leading to greater learner retention and overall satisfaction.

Introduction

The number of learners participating in online education has increased dramatically since 1999 (Radford, 2011), so much so that the number of students in a single massive open online course (MOOC) today may easily number in the thousands (Pappano, 2012). As this explosion of interest increases the number and variety of learners in online classrooms, it also increases the need for educational technologies that easily scale up to provide a personalized learning experience, supporting the varying degrees of learners’ academic preparedness and experience. The methods used to assess the effectiveness of learning and instruction online are clear departures from face to face learning interaction techniques. Learning analytics tools (LATs) help educators visualize the dynamics of online learning, to identify struggling learners and to present personalized feedback. Assessments of learner’s understanding, motivation, and engagement previously analyzed through speech, intonation and body language are now communicated through the analysis of trace data collected and housed within online learning management systems (LMS). Educators need the ability to extract meaningful behavioral and performance patterns from this data to guide learners in the achievement of their educational goals (Sadler, 1989), particularly at-risk and nontraditional learners who may have underdeveloped self-regulatory skills.

Fully 62% of online learners are nontraditional students (Radford, 2011). Nontraditional learners are described as individuals who are full-time employees, older than 25 years of age, enrolled part-time, caretakers for children, parents or other family members, or disabled students who cannot attend brick and mortar schools or who choose not to (Radford, 2011). A November 2011 report by the Babson Survey Research Group found that more than 6.1 million students took at least one online class during the fall of 2010, a 10% increase over the previous year and nearly four times the number of students taking online courses a decade ago (Webley, 2012). Of the traditional undergraduate population, 25% will not achieve their degree within 4 years (National Center for Educational Statistics, 2002). Nontraditional students carry and even higher risk of attrition (Horn, 1996), so successful interventions designed for this population have the potential to positively impact a large number of learners.

Students who self-regulate their learning monitor, evaluate, and adjust their behavior, cognition, and motivation as necessary to successfully complete their academic tasks (Zimmerman, 1989). Learning analytics tools
Latent (LATs) provide a visual representation of the learning process, giving educators and learners an interactive aggregation of individual and group goals, tasks, resource usage, connections and achievements in real-time. The proposed LAT enables self-evaluation through intrinsic reinforcement, which according to Bandura’s social cognitive theory of self-regulation (Bandura, 1991), is potentially more influential than external reinforcement.

This study is motivated by the research linking self-regulatory skill to academic success (Boekaerts & Corno, 2005; Pintrich & de Groot, 1990; Zimmerman, 2000), along with the unique needs of online learning communities. We endeavor to create a learning analytics tool to examine how educators use data visualizations, self-regulatory strategies and formative feedback to regulate their instruction, benefiting educators and learners alike. The tool is designed to assist educators by (1) detecting pedagogically important patterns in learners’ assessments and classroom behaviors, (2) alerting educators to individual learners’ needs quickly so instructional methods may be adjusted appropriately, and (3) supporting their unique working styles through the provision of interactive visual feedback.

Background

Formative Feedback in Education

Michael Scriven first proposed the terms formative and summative assessment in 1967 to differentiate between the roles of evaluation in curriculum. Though the difference between summative and formative feedback may be minute at times in the learning process, feedback is formative only when used to motivate necessary improvements in academic strategy or behavior (Shute, 2008; Tyler, Gagné, & Scriven, 1967). This integral component of the learning process takes many forms, and may be delivered at any time. Formative feedback can range from the complexity of a hint to a fully fleshed out tutorial, complete with response accuracy verification and step-by-step problem-solving explanations.

Formative feedback is powerful; Black and William established that student gains in learning triggered by formative assessment are “amongst the largest ever reported for educational interventions” (Black & William, 2009). Black and William’s studies brought together assessments done by educators and learners with learner’s motivation and achievement practices (Black & William, 1998a; 1998b), surmising that effective formative assessment must involve: (1) learner engagement through self-assessment, (2) educator feedback accompanied by the information learners need to improve, (3) and continuous adjustment to instruction according to these assessments. The motivational aspect of formative feedback may stem from its impact on learners’ goal orientation, as it highlights the effectiveness of their academic strategies.

Like Black and William, Sadler’s research on the role of formative assessment in the development of expertise (Bandura, 1991; Sadler, 1989) sheds light on the benefits of formative assessment practice. Sadler’s theory of formative feedback – namely that it must motivate learners to close gaps between their actual status and desired goals (Sadler, 1989) – is in line with LATs that visually aid users in identifying and closing these gaps. Learners’ self-monitoring is regarded as a distinct process from that of educators’ use of these assessments for instructional adjustment and teaching learners to effectively utilize their own feedback. Additionally, Sadler’s research calls attention to the value of previous evaluative experience educators bring to each formative interaction with learners (Sadler, 1998). Educator helps learners conceptualize what is required to achieve academic standards, and to develop the self-regulatory skill necessary to identify their achievement gaps. Inexperienced educators do not have the same wealth of knowledge to draw from to react appropriately to maladaptive learner behaviors; LATs can help educators identify at-risk learners to meet their needs in a timely fashion, when learners can still recover from their mistakes. Without the development of self-regulatory skills, learners remain completely reliant on educators as their primary source of learning assessment.

Self-Regulated Instruction and Learning

Social cognitive models of learning include the motivational, social and environmental factors that contribute to learner success (Bandura, 1991). These models give credence to a measure of academic performance that values resources such as time management, information and help seeking behaviors, goal setting, self-motivation and emotional control. These practices have been linked with academic achievement through numerous studies (Boekaerts & Corno, 2005; Pintrich & de Groot, 1990; Zimmerman, 2000); specifically, self-regulation is
Learning Analytics

Different types of web-based pedagogical tools support different self-regulated learning processes (Dabbagh & Kitsantas, 2004). In online environments, learning analytics support the activity of learning through the collection, analysis, and interpretation of educational trace data (Ferguson, 2012). This data is a rich source of interactive, visual formative feedback. Like business intelligence, the field of learning analytics uses massive data sets and predictive modeling to yield insight-generating visual information. These insights lead to improved learning strategies and behaviors at the classroom level, and drive decision-making from classrooms to the halls of government (Siemens & Long, 2011). Learning analytics draw on research in human computer interaction, social network analysis, latent semantic analysis, and educational data mining to provide sophisticated guidance to educators and policymakers alike.

As distance education pedagogy moves from cognitive-behaviorism to constructivism and now connectivism, educational technologies must reflect and support the evolving practices emerging from online learning environments. Connectivism is based in the premise that learning builds on a network of information; learners must be able to identify, interpret and apply this information when needed (Siemens & Long, 2011). Connectivist models involve exposing users to networks and “providing opportunities for them to gain a sense of self-efficacy in networked-based cognitive skills” (Anderson & Dron, 2011). Learning analytics tools are part of the information network of learners and educators, improving their technical literacy and confidence with repeated use. While all learning analytics systems leverage trace data, few provide formative feedback or employ self-regulated learning pedagogies since historically learning analytics tools have been pedagogically neutral.

Turnitin (“The Scientific Basis of Turnitin,” 2010) is arguably the most well-known commercial educational technology tool for formative feedback and assessment. The web-based platform detects learners’ written plagiarism, returning visual feedback on the percentage of semantic similarity between the submitted work and the millions of papers, journals, books and websites in the Turnitin database. The tool saves educators time – it is highly unlikely a single educator could review this amount of data in a timely fashion – and the tools’ visual feedback guides educators not versed in plagiarism detection. Jocoy and DiBiase’s research shows that Turnitin’s semantic analysis detected five times more instances of plagiarism than manual methods (Jocoy & DiBiase, 2006). Additionally, Davis and Carroll report that using Turnitin as a formative assessment tool helps learners to avoid plagiarism, decreasing their tendency to rely too heavily on sources and improving both citation practice and paraphrasing skill (Davis & Carroll, 2009).

Piloted in 2007 on the Blackboard Learning Management System (LMS), the Signals LAT uses trace data collected from the LMS to provide real-time, frequent intervention feedback directly to learners (Arnold, 2010). Though its use of predictive models to incite at-risk learners to take action is an excellent example of nudge analytics, it doesn’t support a specific approach to learning (Carmean & Mizzi, 2010). Evaluated by more than 1,500 users across five semesters, feedback on the Signals tool was resoundingly positive. Of those surveyed, 89% said that the tool provided a positive experience and 58% of the respondents would like to use it in every course (Pistilli, Arnold, & Bethune, 2012). The success of Signals surely influenced the design of the Blackboard Learn suite of commercial analytics applications, developed specifically the Blackboard LMS in 2012 (Blackboard Inc., 2012).
The analytics tool suite provides course-specific and institution-wide data; it is being deployed in field trials with select colleges and universities this year.

The CourseVis LAT was developed in 2004 to help online educators visualize learner performance data gathered from the WebCT LMS (Mazza & Dimitrova, 2007). It used trace data to visualize learners’ social, cognitive and behavioral data en masse to educators (Mazza & Dimitrova, 2004). To design the tool, the researchers surveyed educators to identify the information most used to aid at-risk learners and visualize it for educators’ use effectively. GISMO, the subsequent open source Moodle LMS tool built by the same research team, iterates on the results of the qualitative information collected (Mazza & Botturi, 2007). As with the GISMO tool two iterations of the LOCO-Analyst tool were developed to qualitatively evaluate users’ perceptions, using questionnaires and focus groups (Jovanović, Gašević, Brooks, Devedžić, & Hatala, 2007). The first studies gathered feedback on the tools’ functionality and features, while follow-up studies allowed users to use the tool for some time before providing additional feedback.

The IQ Learn tool suite applies Pintrich’s Motivational Components of Forethought, Cognitive Strategies and Learning Skills (Niemi, Nevgi, & Virtanen, 2003). It utilizes automated tutoring sets, interactive questionnaires prompting metacognitive reflection, and a learning diary that serves as a self-reflection tool for learners and a source of qualitative usage information for researchers. The tutorials ask learners to reflect on their academic habits, and then the intelligent system suggests skill improvements based on learners’ unique areas of need. Students receive their test results online as visual profiles, along with the mean values and standard deviations of their virtual study group (Niemi et al., 2003). Though not a learning analytics tool, the tool suite is one of few that supports learners’ self-regulation in online educational settings.

**Study**

The basis of this study involves the derivation and testing of innovative learning analytics visualizations to optimize educators’ and learners’ academic strategies and behaviors in online environments. The learning analytics tool’s (LAT) design is rooted in established frameworks of formative feedback (Black & Wiliam, 1998b), academic self-regulation practice (Winne & Hadwin, 1998; Zimmerman, 1989) and design-based research (Anderson & Shattuck, 2012). The first phase of this research is a visualization pre-design study (Isenberg, Tang, & Carpendale, 2008) exploring the kinds of data that can be collected from learning management systems to aid educators. Public data from the PSLC DataShop is used to model the first iteration of the visualizations. Collected from a 2005-2006 Algebra 1 course utilizing the Cognitive Tutor (Ritter et al., 2007) intelligent tutor software, this dataset is composed of solely of formative data. This is exemplary of the datasets housed in the PSLC DataShop repository and fits the purpose of the present study, given its emphasis on the provision of formative feedback. This particular dataset has been used in numerous learning analytics and educational data mining studies (Baker et al., 2008a; Baker et al., 2008b, Baker & Carvalho, 2008; Baker et al., 2009; Matsuda, 2007a; Matsuda, 2007b). The use of public data from repositories like this one makes it easier to externally validate an analysis (Baker & Yacef, 2009).

The proposed LAT would aggregate and visualize data from all sections of a course, past and present. We posit that the ability to view all of this data concurrently will give educators the formative feedback necessary to adjust their instruction to the needs of their students. The visualizations will help educators visually deduce the learners’ progress, as well as the behaviors that tend to lead to success in their courses. Further, the educator will be able to interactively search at the individual, class section, and course level.

Modeled using Tableau data visualization software (“Visual analytics for everyone,” 2003), the figure below represents the educator dashboard for the Algebra I course. The view represents the evaluation of a single step necessary to solve problem #47, to determine if the subject matter should be revisited. Represented by an anonymous student identifier, the top pane shows the number of learner attempts to solve the problem, including hints used. The bottom pane indicates the amount of time the learners spent on task, against the time they spent in the learning content. The trend line represents the average amount of time spent on the step. The time on task provides contextual information for the step attempts. For example, the step seen in the dashboard below has a single outlier. Highlighted in the main and alert panes, student XqBaV46VbC accessed the secondary hints 56 times, likely indicating an overreliance on hints to solve this step. The educator may surmise that this student needs to review this knowledge step, along with the learners who more time on task than average.
This mixed methods emergent study design will rely heavily on the qualitative data collected from educators. Like CourseVis, GISMO and LOCO: Analyst, this data will build the foundation of the learning analytics tool’s visualizations and feature set, guiding further design and development of the tool. After the open source tool is deployed in the Moodle LMS, additional quantitative trace data will be collected to provide usage and experience data to support the qualitative information.

A longitudinal within-subjects mixed methods study is appropriate for this research given the iterative nature of design-based research (Anderson & Shattuck, 2012), the context of a dynamic online learning environment, and the varied needs of the multiple stakeholders invested in its use. The framework introduced in this study will use behavioral theory, visualization and human computer interaction techniques to provide a mixed methods approach to constructing understanding around the attitudes and activities of educators and learners in online educational environments, providing a realistic view of the needs and practices of these users. Subsequent iterations of the tool will be studied with both learners and educators to validate its usefulness, address the phenomenology of the learning environment and contribute to the establishment of benchmark tasks for future study in learning analytics. Emergent themes from semi-structured interviews with both stakeholder groups will guide the coding scheme used in later study.

An important component of the value of a visualization lies in its long-term repeated use (Thomas & Cook, 2006), so the tool will aggregate usage log data and participant input over the course of a semester. Users will be prompted to give weekly qualitative feedback on their overall experience to supplement the trace data, illustrating a holistic view of what transpires in the classroom in real time. This data from users in their natural environment doing real tasks demonstrates the overall feasibility and in-context usefulness of the tool.
Conclusion

Self-regulation is instrumental to the development of personal empowerment – in academia and otherwise (Fetterman, 2000). Despite its efficacy in promoting academic success, limited research exists that explores the design of learning analytics tools employing self-regulated learning strategies. The research principles of this longitudinal mixed-method within-subjects study explore how learning analytics employing self-regulatory pedagogies may be used to improve educator efficacy, leading to greater learner retention and satisfaction overall. This study contributes to empirical evidence of the relationships between self-regulatory strategies and achievement, advancing the body of knowledge on how these strategies benefit both educators and learners. It also contributes to theory and practice in the areas of learning analytics, educational data mining, human computer interaction, and education.

References


