Literature-evidence base: Learning Analytics

Authors: Aneesha Bakharia, Hassan Khosvari

Prelude

The importance of learning analytics within the higher education context has already been acknowledged by The University of Queensland with the ongoing funding of a Learning Analytics team within ITaLI. This literature review has therefore not focused on highlighting the importance and relevance of learning analytics and instead sought to justify the selection of projects that have been funded by the Student Strategy initiative and provide supportive evidence for the design decisions make within these projects.

Course Insights Project

Learning Analytics emerged as a field in 2011 to apply analytics to educational data in order to better understand and support teaching and learning (Siemens et al., 2011).

Increasing the effectiveness and efficiency of instructors has been an important driver for learning analytics research (Buckingham Shum, Gašević, & Ferguson, 2012). In seeking to maximise the reach and impact of learning analytics across UQ, course instructors were seen as key strategic group to target with a course focused analytics dashboard (i.e., The Course Insights Project). Instructors have always been tasked with making sense of data about their students’ learning in order to guide their teaching (Borko & Shavelson, 1990), which is increasingly becoming a challenge with large student numbers, the adoption of blended learning, and the use of multiple learning tools. Providing instructors with visualisations and metrics delivered via a dashboard in many research studies, has failed to produce actionable insight (Spillane, 2012; Verbert et al., 2013), regular usage (Dazo, et al., 2017) and measurable student behavioral change. The Course Insights Project therefore sought to incorporate recent research findings that advocated the alignment of analytics to teacher inquiry (Sergis & Sampson, 2017), learning design (Lockyer, Heathcote & Dawson, 2013; Bakharia, et al., 2016), targeted student feedback (Pardo, et al., 2018; Pardo, et al., 2019) models and the co-design of analytics solutions (Dollinger et al., 2019).

There is a rich history of developing and evaluating dashboards that pre-dates the emergence of learning analytics as a field (France et al., 2006; Zinn & Scheuer, 2007). Most instructor dashboards have been designed to provide an overview of course activity and identify at risk or isolated students (Verbert et al 2013). Examples of recent teacher focused dashboard applications include GLASS (Leony et al., 2012) which included visualisations of aggregated daily activity and learning events; LeMo (Fortenbacher, 2013) which focused on pathway analysis; and Loop (Corrin et al., 2015) which provided clickstream and tool usage analysis for analysis of course learning design. Loop also provided visualisations to determine whether students had engaged with online resources before face-to-face lectures and visualised the impact of face-to-face delivery sessions (i.e. lectures and tutorials) and assessment due dates on student resource access and collaboration.

The design of Course Insights has been directed by the following four design principles:

**Involving instructors in the co-design of dashboard**

Co-design allows stakeholders to take ownership of a product and suggest features that are useful within their daily tasks (Dollinger et al., 2019). The agile development methodology adopted by the Course Insights team has allowed for continuous feedback to be addressed and new features to be added at regular release intervals. The prototype user interface was first validated at stakeholder engagement workshops.

**Providing filterable and comparative student sub-cohort visualisations**
The conceptual framework linking learning design with learning analytics (Bakharia et al., 2018), places the instructor as a decision maker analysing various types of student data. An essential element in instructor decision making is comparative analysis where the instructor is able to compare different student sub-cohorts and even different course offerings. Course Insights includes unique filtering functionality and automatically includes comparative (i.e. multiple series) visualisations to allow the whole cohort to be compared with the filtered cohort (e.g., learners from a specific program and/or demographics).

**Provide analytics across the course life-cycle**

Providing analytics that was relevant to time periods across the across the course life-cycle (i.e., orientation week, first week of semester, post census date, post assessment, end of course, etc) was seen as fundamental to the success of the Course Insights dashboard. Sergis and Sampson (2017) in a systematic review found learning analytics helpful for teacher inquiry, with data on assessment and learner engagement essential for analysing elements of course design. Course Insights provides histogram visualisation for all assessment items included in the Blackboard Grade Centre and time series visualisations of activity across all platforms used at UQ (i.e., Blackboard and edX Edge). Data from the student information system is also included (i.e. Enrolment Profile) to allow instructors to understand the cohort that has enrolled in the course early in the course life-cycle (i.e., orientation week). Data sources and visualisations included in Course Insights have been designed to facilitate the iterative process of creating and evaluating learning activities (Mor, Ferguson, & Wasson, 2015) and whole-class scaffolding (e.g., an instructor devoting face-to-face time to provide additional explanation of a concept) (Xhakaj et al., 2017).

**Facilitate the provision of personalised feedback**

The need to identify sub-cohorts of students and provide personalised feedback has driven the development and evaluation of tools such as OnTask (Pardo, et al., 2018; Pardo, et al., 2019). OnTask provides instructors with a user interface to build queries to identify groups of students and allows the created rules to be inserted in email templates where instructors can author appropriate feedback. Course Insights has been designed to bridge the gap between dashboards that provide aggregate data and feedback tools which allow sub-groups of student to be identified for targeted feedback. Inspired by OnTask, Course Insights includes an intermediate query builder that displays learners making the query criteria in a tabular format with email functionality. The query builder provides by Course Insights facilitates targeted instructor scaffolding (i.e., the instructor can identify students finding a concept difficult or not engaging with course resources and provide customised support either via email or an appointment) (Herodotou et al., 2017).

**Data Lake Extensions Project**

In 2018, the University of Queensland followed the industry trend (Gartner Inc, 2018) by piloting cloud-based data lake technology as an alternative to on-campus enterprise data warehouses. The UQ pilot used data lake infrastructure provided by leading cloud vendor Amazon Web Services (AWS). Data lakes offer many advantages over traditional data warehouses with significant differences between the two technologies (Datafloq, 2018). Data lakes store data in its default format, not requiring processing to transform the data to a fixed schema for analysis. Data lakes provide flexibility and cost efficient scalability. The AWS solution separates data storage and computation resources, resulting in minimal costs for storage and additional costs for compute only being incurred when data is being processed or analysed. Table 1, includes a comprehensive comparison between enterprise data warehouses and data lakes.

The UQ Data Lake has been an instrumental technology in delivering the Course Insights dashboard and enabling educational data science related activities. In particular, the data lake has allowed teaching and learning datasets from multiple systems to be integrated in a central location. A key design principles for Course Insights included the ability to provide visualisations that would be relevant across different course life-cycle events, linking analytics to course learning design.
Table Y-1: Data Warehouse and Data Lake Comparison (reproduced from AWS, 2019)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Data Warehouse</th>
<th>Data Lake</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Relational from transactional systems, operational databases, and line of business applications</td>
<td>Non-relational and relational from IoT devices, web sites, mobile apps, social media, and corporate applications</td>
</tr>
<tr>
<td>Schema</td>
<td>Designed prior to the DW implementation (schema-on-write)</td>
<td>Written at the time of analysis (schema-on-read)</td>
</tr>
<tr>
<td>Price/Performance</td>
<td>Fastest query results using higher cost storage</td>
<td>Query results getting faster using low-cost storage</td>
</tr>
<tr>
<td>Data Quality</td>
<td>Highly curated data that serves as the central version of the truth</td>
<td>Any data that may or may not be curated (ie. raw data)</td>
</tr>
<tr>
<td>Users</td>
<td>Business analysts</td>
<td>Data scientists, Data developers, and Business analysts (using curated data)</td>
</tr>
<tr>
<td>Analytics</td>
<td>Batch reporting, BI and visualizations</td>
<td>Machine Learning, Predictive analytics, data discovery and profiling</td>
</tr>
</tbody>
</table>

A second key design principle for Course Insights was to provide instructors with a way to dynamically filter and compare learner sub-cohorts in real-time. As the AWS data lake solution was unable to meet the real-time requirement, a NoSQL dynamic data store (TechTarget, 2018) was added specifically to meet the interactive query requirement. The ElasticSearch NoSQL database (ElasticSearch, 2019) currently provides search results for both the simple and advanced sub-cohort filters within the Course Insights dashboard. ElasticSearch (ElasticSearch Reference, 2019) also includes advanced time series aggregation functionality which processes data for the time series visualisations for learner engagement.

RiPPLE

Adaptive Learning Systems ALSs (Park & Lee, 2003) dynamically adjust the level or type of instruction based on individual student abilities or preferences to provide an efficient, effective, and customised learning experience for students. At a high level of generality, there are two main classes of ALS. The first class, commonly referred to as Intelligent Tutoring Systems (ITSs) (Anderson, Boyle, & Reiser, 1985) use AI-based algorithms to replicate the support that is often provided by a tutor by providing personalised step by step guidance in solving a problem. Carnegie Learning’s MATHiaU (Ritter, Carlson, Sandbothe, & Fancsali, 2015) is an established example of this class of ALS. The second class of ALSs focus on adaptively recommending learning activities from a large repository of learning resources to a student to match their current learning ability. Pearson’s MyLabs (using Knewton (Jose, 2016) for its adaptive functionality) and McGraw-Hill’s LearnSmart and ALEKS (Falmagne, Cosyn, Doignon, & Thiéry, 2006) are established examples of this class of ALSs. RiPPLE is also representative of this second class.

A consistent and growing body of knowledge provides evidence about the effectiveness of both classes of ALSs (Anderson, Corbett, Koedinger, & Pelletier, 1995; VanLehn, 2011; Ma, Adesope, Nesbit, & Liu, 2014). For example, a comprehensive meta-review by VanLehn (2011) reported that on average using ITSs have a learning gain effect-size (Cohen, 1992) of $d = 0.76$ relative to classroom teaching without tutors. For ALSs
that focus on recommending learning resources, Yilmaz (2017) and Mojarad, Essa, Mojarad, and Baker (2018) have reported improvement on student performance while using ALEKS or the popular ASSISTments Ecosystem (Heffernan & Heffernan, 2014). At a high level of generality, many ALSs rely on the following interacting components (Essa, 2016):

- **Domain model**: A knowledge space modelling what the students need to know. The domain model is commonly presented as a set of knowledge units that are “elementary fragments of knowledge for the given domain” (Brusilovsky, 2012).

- **Learner model**: An abstract representation of students often “overlaying” their knowledge state on the knowledge space defined in the domain model (Brusilovsky & Mill’an, 2007). The learner model may estimate a student’s ability level on different knowledge units based on their performance and interactions with the system. Importantly, using open learner models (Bull & Kay, 2010), which are learner models that are externalised and made accessible to students or other stakeholders, can be particularly effective in helping students to learn (Bodily et al., 2018).

- **Content repository**: A repository of learning resources which may include assessment-based and learning-based items designed to help the learner acquire the knowledge represented by the domain model. Each learning resource is tagged with knowledge units defined in the domain model.

- **Recommender engine**: A recommender engine that utilises information from the learner model and the content repository to select learning activities for each student that will be mostly likely to advance their learning of the domain knowledge.

Commonly, ALSs are developed using the publisher model (Oxman, Wong, & Innovations, 2014). In this model, the platform is designed with pre-existing learning activities, often based on textbooks from a publisher. Pearson’s MyLabs (using Knewton (Jose, 2016) for its adaptive functionality), McGraw-Hill’s LearnSmart and ALEKS (Falmagne et al., 2006) are established examples of this model. The publisher model has been successful in K-12, where course content follows a more simplistic structure, and often has to comply with national standards. However, higher education has been slow to embrace these systems, with adoption mostly restricted to research projects (Essa, 2016). The focus on specific restricted domains, limited flexibility for educators to tailor the learning activities to their context, and the high costs associated with the use of these platforms, have all contributed to their low adoption in higher education.

Responding to these limitations, an alternative has been established, referred to as the platform model (Oxman et al., 2014). The platform model provides a content-agnostic system infrastructure that enables educators to develop and author the content of their course. Smart Sparrow (Sparrow, 2016) and many learning management systems such as Desire2Learn, Loudcloud and edX that incorporate adaptive functionality into their course building tools follow this model. The platform model is relatively new and mostly suffers from an operational limitation rather than a technological one; implementing adaptivity in a course requires a large number of new learning activities and object tagging, which introduces significant overheads for teaching staff in both time and training. To overcome limitations of both of these models RIPPLE leverages ideas from crowdsourcing in education (Denny, Hamer, Luxton-Reilly, & Purchase, 2008), which has acted as a source of inspiration for RIPPLE. The use of crowdsourcing in ALSs is also beginning to receive attention. For example: Heffernan et al. (2016) proposed employing crowdsourcing within the popular ASSISTments platform. Williams et al. (2016) presented an Adaptive eXplanation Improvement System (AXIS) that uses crowdsourcing to generate, revise and evaluate explanations as learners solve problems; and Karataev and Zadorozhny (2017) proposed a framework that combines concepts of crowdsourcing, online social networks, and adaptive systems to

Responding to these limitations, an alternative has been established, referred to as the platform model (Oxman et al., 2014). The platform model provides a content-agnostic system infrastructure that enables educators to develop and author the content of their course. Smart Sparrow (Sparrow, 2016) and many learning management systems such as Desire2Learn, Loudcloud and edX that incorporate adaptive functionality into their course building tools follow this model. The platform model is relatively new and mostly suffers from an operational limitation rather than a technological one; implementing adaptivity in a course requires a large number of new learning activities and object tagging, which introduces significant overheads for teaching staff in both time and training. To overcome limitations of both of these models RIPPLE leverages ideas from crowdsourcing in education (Solemon, Ariffin, Din, Anwar, et al., 2013), by having students themselves generate and evaluate content that can then be adaptively served.

The use of crowdsourcing in education alongside insights from the students as partners approach (Matthews, 2017) makes way for respectful, mutually beneficial learning partnerships where students and staff work together. Successful examples of such partnerships have led to co-creation of curricula (Bovill, 2013), marking criteria (Meer & Chapman, 2014) and assessment items via the popular PeerWise platform (Denny, Hamer, Luxton-Reilly, & Purchase, 2008), which has acted as a source of inspiration for RIPPLE. The use of crowdsourcing in ALSs is also beginning to receive attention. For example: Heffernan et al. (2016) proposed employing crowdsourcing within the popular ASSISTments platform. Williams et al. (2016) presented an Adaptive eXplanation Improvement System (AXIS) that uses crowdsourcing to generate, revise and evaluate explanations as learners solve problems; and Karataev and Zadorozhny (2017) proposed a framework that combines concepts of crowdsourcing, online social networks, and adaptive systems to
provide personalised learning pathways for students. However, this preliminary work is yet to realise the full potential offered by crowdsourcing in ALSs or more broadly in education, which is the focus of RiPPLE.

An important assumption made by RiPPLE is that students, as non-experts, have the ability to create high-quality resources. While we were not able to find longitudinal or meta reviews that support this claim, there seems to be adequate evidence suggesting that students have the ability to create high-quality learning resources that meet rigorous judgemental and statistical criteria (Walsh, Harris, Denny, & Smith, 2018; Tackett et al., 2018; Denny, Hamer, Luxton-Reilly, et al., 2009; Galloway & Burns, 2015; Bates, Galloway, Riise, & Homer, 2014). In fact, students as authors of learning resources, may have an advantage over instructors; they can utilise knowledge of their own previous misconceptions towards the creation of resources that may have a lower chance of suffering from an expert blind spot. However, it is very likely that some of the learning resources developed by students may be ineffective, inappropriate or incorrect (Bates et al., 2014). As such, to effectively utilise resources developed by students, there is a need for a selection and moderation process to identify the quality of each resource. RiPPLE provides multiple options for moderation of the created resources. One of these options is “staff moderation”, where the created resources are moderated by the instructors before they are publicly released and added to the repository of the learning activities for the offering. However, this may not be feasible in large classes. Alternative options of moderation implemented in RiPPLE rely on the collective wisdom of the crowd and use methods that are generally used for reviewing of academic articles (e.g., “Competent student moderation” option) or content moderation on social networks (e.g., “flagging inappropriate content”). This raises an interesting research question; can non-experts accurately identify the quality of a learning resource? A recent paper by Whitehill, Aguerrebere, and Hylak (2019) provides evidence that non-expert subjective opinions may be utilised to accurately determine the quality of a learning resource, and that machine learning algorithms may be used to infer the reliability on an individual's opinion, which further increases the accuracy of the results. Further investigations about this topic are under way by the authors. Our initial results are aligned with the findings of (Whitehill et al., 2019).

References


Bovill, C. (2013). Students and staff co-creating curricula: A new trend or an old idea we never got around to implementing? In C. Rust (Ed.) Improving Student Learning through research and scholarship: 20 years of ISL. (pp. 96–108). Oxford: The Oxford Centre for Staff and Educational Development.


