Proceedings of the 1st Learning Analytics for Curriculum and Program Quality Improvement Workshop

April 25, 2016
Edinburgh, UK

Hosted by the Learning Analytics and Knowledge (LAK) 2016

Jim Greer, Marco Molinaro, Xavier Ochoa & Timothy McKay (Eds.)
Workshop Chairs
- Jim Greer, University of Saskatchewan, Canada
- Marco Molinaro, University of California – Davis, USA
- Xavier Ochoa, Escuela Superior Politécnica del Litoral, Ecuador
- Timothy McKay, University of Michigan, USA

Program Committee
- Christopher Brooks, University of Michigan, USA
- Catherine Uvarov, University of California - Davis, USA
- Christopher Pagliarulo, University of California - Davis, USA
- Craig Thompson, University of Saskatchewan, Canada
- Stephanie Frost, University of Saskatchewan, Canada
- Katherine Chiluiza, Escuela Superior Politécnica del Litoral, Ecuador
- Cristian Cechinel, Universidade Federal de Pelotas, Brazil
- Leah Macfayden, University of British Columbia, Canada
- Emily Miller, Association of American Universities, USA
- Vive Kumar, Athabasca University, Canada
- Phil Long, University of Texas at Austin, USA
# Table of Contents

**Workshop Description** .................................................................................. 1  
*Jim Greer, Marco Molinaro, Xavier Ochoa & Timothy McKay*

**Empowering instructors through customizable collection and analyses of actionable information** ................................................................. 3  
*Danny Y.T. Liu, Charlotte E. Taylor, Adam J. Bridgeman, Kathryn Bartimote-Aufflick and Abelardo Pardo*

**Using a Risk Management Approach in Analytics for Curriculum and Program Quality Improvement** ............................................................. 10  
*Wai Yee Wong and Marcel Lavrencic*

**LMS Course Design As Learning Analytics Variable** ........................................ 15  
*John Fritz*

**Simple Metrics for Curriculum Analytics** ..................................................... 20  
*Xavier Ochoa*

**Assessment Analytics for Peer-Assessment: A Model and Implementation** ... 25  
*Blaženka Divjak, Darko Grabar and Marcel Maretić*

**Data-Driven Programmatic Change at Universities: What works and how** ... 30  
*Jim Greer, Craig Thompson, Ryan Banow and Stephanie Frost*

**Promoting Instructor and Department Action via Simple, Actionable Tools and Analyses** ................................................................................. 34  
*Marco Molinaro, Matthew Steinwachs, Qiwei Li and Alberto Guzman-Alvarez*
Workshop Description

Workshop Chairs

Much of the research in LAK to date has been “student facing”, that is, using data to better understand learners and their need or to create interventions that directly support or influence learners. This workshop takes the perspective on how Learning Analytics can drive improvements in teaching practices, instructional and curricular design, and academic program delivery. While this does influence student outcomes in the long term, the data gathered and evidence generated is more instructor and administrator facing. We have seen examples of how LAK can help build the case for instructional, curricular, or programmatic change and further how LAK can be used to foster acceptance of change processes by teachers, administrators, and other stakeholders in the educational enterprise. When successful, these kinds of changes are often associated with educational reform or culture shifts in educational practice.

This workshop offered those in the LAK community an opportunity to share and explore how educational data, its analysis and visualization, and the evidence derived can change/improve the context of learning. The main research and practice questions addressed in this workshop were:

- How to provide relevant and actionable information to faculty, teaching assistants, departmental and college administrators to encourage a greater emphasis on student learning and use of evidence-based practices, thus encouraging a continuous improvement approach to teaching and learning.
- How to create visualization and data collection tools and approaches that encourage a community of instructors and administrators to engage in making evidence-based decisions to improve student learning whether at the activity, lesson, course, series, department, college or university-wide levels.
- How to extract information from the multiple modalities used in instructional environments and help capture, represent and evaluate faculty instructional approaches, student-faculty engagement, student-student interactions and student-technology interactions.
- How to represent, summarize and mobilize data from human interactions and student-technology interactions to motivate change and quality improvement. Is there a way to make the data and representations more useful for promoting sustainable change?
- How to change the evaluation of instructional activities to be more formative, actionable and multi-dimensional in nature, emphasizing individual and group improvement rather than a one-size fits all student survey by which instructors or courses are judged and compared. Ideally such evaluation systems would go well beyond student satisfaction as the sole measure of good teaching and should include learning outcomes, utility of outcomes, applicability to future learning, match or fit between learner and instructor, and more.
- How can analytics help the individual instructor to examine the success of a course that they teach? By including the analysis of individual courses we will keep the interest of all instructors, and not just those concerned with larger questions of course sequences and curricula. Curricula are built of individual classes of course, and campus
authorities at all levels need tools for examining the success and impact of individual courses.

Seven papers were accepted for presentation at this Workshop. These papers are reproduced here.
Empowering instructors through customizable collection and analyses of actionable information

Danny Y.T. Liu
Faculty of Science
The University of Sydney
danny.liu@sydney.edu.au

Charlotte E. Taylor
Faculty of Science
The University of Sydney
charlotte.taylor@sydney.edu.au

Adam J. Bridgeman
Educational Innovation
The University of Sydney
adam.bridgeman@sydney.edu.au

Kathryn Bartimote-Aufflick
Quality and Analytics Group
The University of Sydney
kathryn.aufflick@sydney.edu.au

Abelardo Pardo
Faculty of Engineering and IT
The University of Sydney
abelardo.pardo@sydney.edu.au

ABSTRACT
The use of analytics to support learning has been increasing over the past few years. However, there is still a significant disconnect between what algorithms and technology offer and what everyday instructors need to gather actionable insights from these tools into their learning environments. In this paper we present the evolution of the Student Relationship Engagement System, a platform to support instructors to select, collect, and analyze student data. The approach provides instructors the ultimate control over the decision process to deploy various actions. The approach has two objectives: to increase instructor data literacies and competencies, and to provide a low adoption barrier to promote a data-driven pedagogical improvement culture in educational institutions. The system is currently being used in 58 courses and 14 disciplines, and reaches over 20,000 students.

CCS Concepts
- Information systems--Decision support systems
- Human-centered computing--Visual analytics
- Computing methodologies--Machine learning approaches
- Applied computing--Education
- Software and its engineering--Software creation and management

Keywords
Learning analytics adoption; scaling up; instructors; curriculum design and delivery; teaching approaches; machine learning.

1. INTRODUCTION
Since the early days of learning analytics (LA), the promise has been that the collection and analysis of large educational datasets could yield “actionable intelligence” [8, p41] to improve the overall student learning experience. At some of the institutions that have adopted LA, this intelligence typically takes the form of algorithms that predict student outcomes and aim to reduce attrition and failure rates [10; 16; 44; 53]. The higher education sector has been one of the first to explore the adoption of these techniques [22]. Despite these initiatives, recent reviews highlight the lack of widespread adoption of LA in the higher education sector [10; 44]. Various explanations have been suggested for this. At a high level, these include policy and ethical challenges [41; 54], institutional leaders’ misconceptions of LA [10], and the sector’s general culture of resistance to change [19; 40]. At an operational level, other authors have reported the inflexibility of vendor solutions, and difficulties in accessing data [38], as well as the accuracy of such data [6]. To add complexity to this situation, evidence is mounting that the one-size-fits-all approach, typical in LA, may be inadequate in explaining student outcomes [21; 34; 55] and addressing the needs of students in different disciplines [43].

Notwithstanding, there is increasing interest in the instructor-facing benefits of LA. These include detecting patterns and trends, using data to support decision making, testing assumptions, and understanding the effect of learning designs [25]. Tools that display and analyze student data can help instructors reflect on their designs and better understand the relationships between variables [15; 51]. Moreover, new tools are being developed that address a long-held appeal to connect LA with the learning sciences [18], by helping instructors understand how learner behaviors correspond with their pedagogical intent [11]. Recent results in the area of artificial intelligence in education suggest a shift in focus away from fully self-contained decision systems to a paradigm based on human intelligence amplification [5]. However, low data literacies and competencies pose a significant barrier to address this shift and achieve wider LA acceptance and adoption [6; 24]. Taken together, these suggest that greater impact of LA (e.g. insight into curricular design and delivery versus prediction of retention), may be catalyzed by addressing, and indeed leveraging, identified adoption barriers. In this paper, we take the position that, to be effective, LA must empower instructors with tangible solutions to address pressing needs [15; 37]. For some, this may mean addressing immediate retention issues [10], that is, “to satisfy a tangible, small-scale problem” [38, p236], while pushing instructors along the adoption pipeline [35] to more involved insights. This builds on findings from early adoption of computers in teaching, where “use of computers for one purpose may encourage enthusiasm for further computer use” [26, p7]. We present a case study of a bespoke web-based LA solution at the University of Sydney, outline its capabilities and impact, to date, and highlight the flow-on impacts for shifting teaching practices, curricular design and delivery, and growing a culture of LA use. We use Greller and Drachsler’s [24] generic LA design framework to situate our work in terms of stakeholders, objectives, data, instruments, and limitations.

2. OVERVIEW OF OUR APPROACH
We opted for a bottom-up approach where a basic but high-utility system was developed and improved collaboratively with instructors. From an early stage, this meant that our system addressed pressing objectives of key stakeholders [14]. Our design philosophy shared common themes with other LA developments, including usability, usefulness, data interoperability, real-time operation, flexibility, and generalizability [8; 15; 23]. However, in contrast to other approaches, our system’s growth was instructor-centered and ‘organic’, initially addressing a small-scale need...
(originally, tracking class attendance) and iteratively building features into the system (e.g. personalized interventions, data mining to uncover hidden relationships in course design) as instructors’ data literacies and competencies grew. A recent review of LA implementations at Australian institutions suggests that such early small-scale applications can have large impacts on capacity building [10].

2.1 Data collection
The importance of having the right data in the right place is a central issue for LA [28]. Most practical LA implementations involve collecting data into a central database available to the instrument [e.g. 3; 15; 38] or building analytics directly into the data source [e.g. 33]. Recognizing that both LMS and student information system (SIS) data have shortcomings [21; 31], and in keeping with our instructor-empowering philosophy, we opted for a hybrid approach where instructors could decide which data were most important for their contexts. For example, our discussions with instructors identified that class engagement and attendance data were important, in keeping with evidence-based practice for student outcomes [42; 47].

Unsurprisingly, interim grade and other performance data were also relevant [9]. Therefore, we started by developing a web-based, smartphone-friendly, system that was easy and efficient to use and met these contextual needs (Figure 1). Since technology acceptance and adoption are closely linked with usefulness and usability [12], this was a first step in empowering instructors’ data usage.

Due to technical limitations of our institution’s information technology infrastructure and capabilities, our system could not programmatically access LMS or SIS data. Other authors have solved this issue by capitulating to vendor-locked solutions, which offer a level of automatization but at the cost of flexibility, customizability, and possibly even scalability [38]. We addressed the issue by building in an additional facility to import any student-matched data required through semi-automated data file uploads. This is a similar design philosophy to Graf et al. [23] in allowing free choice of data, and addresses realistic instructional situations where course-specific nuances can confound less flexible systems [38]. Serendipitously, this had the unintended advantage of forcing instructors to consider the data they were entering, in terms of its relevance to their context and pedagogical design. In fact, the criticality of these contextual factors is becoming much clearer [e.g. 15; 21], lending strong support to our approach. In terms of Greller and Drachsler’s [24] framework, our approach addressed the direct objectives of stakeholders in providing a stable, easy to use instrument that collected immediately relevant data.

2.2 Data extraction and affordances for action
Once the right data are in the right place, the typical progression in LA usually involves visualization via dashboards [45]. However, there is a danger that these visually appealing interfaces may distract users (such as instructors, students, and management) from a deeper understanding of the underlying data. Greller and Drachsler astutely describe that “enticing visualisations...[and] the simplicity and attractive display of data information may delude the data clients, e.g. teachers, away from the full pedagogic reality” [24, p52]. With this in mind, we decided to minimize visualizations and instead provide instructors with the ability to run large-scale customized queries on their students’ data. This meant that instructors of even very large courses could select, collect, and extract the data they wanted, and also run basic analyses that are of interest to their contexts [23]. Importantly, we aimed to avoid algorithmic black boxes [35], which are present in other solutions [e.g. 2], instead giving instructors full control of the process.

This level of functionality was built to respond to pressing institutional needs to address issues of student engagement, taking advantage of the data that were already being collected. Using the customizable analysis engine, instructors could specify conditions and efficiently identify particular groups of students (Figure 2). Once identified, instructors could then deliver personalized feedback to students via email or the cellular network. We observed that instructors “relied on their intuition and hunches to know when students are struggling, or to know when to suggest relevant learning resources” [13, p20].

In addition to this approach to extracting information at scale, we also focused on a seldom-raised concern, namely “the focus of LA appears fixed to an institutional scale rather than a human scale” [31, p4]. We therefore wished to promote the power of LA in augmenting human interaction. To this end, our system design allowed instructors to customize the information that could be immediately extracted and displayed to other staff (such as tutors and support staff) as they worked directly with students in face-to-face contexts (e.g. Figure 1). In a similar application, Lonn et al. [37] empowered academic advisors with pertinent student data. While use of our system in this way has been predominantly operational (e.g. redirecting students in class if they have not completed assigned pre-work), we envisage that, as more relevant data are available, this ‘mini human dashboard’ approach will spark deep human conversations supported by the relevant data.

In terms of Greller and Drachsler’s [24] framework, our approach allowed both staff (faculty as well as student support staff) and student stakeholders to take advantage of data through the instrument. In this process, information was prepared and presented to stakeholders, and the transparent analysis engine also forced instructors to develop data interpretation and decision-making competencies [24]. Moreover, we saw our approach as reflecting the human judgment and instructor empowerment roots of LA [52].
2.3 Guided semi-automated discovery

The closely related field of educational data mining has a greater focus on automated methods of discovering meaning in educational data than LA [4], which address one of the key opportunities for LA, namely "to unveil and contextualize so far hidden information out of the educational data" [24, p47]. Data mining techniques in LA [4] have primarily focused on outcome prediction through regression and classification [e.g. 21], semantic analyses [29], and social network analysis [e.g. 36]. However, data mining techniques typically require substantial technical understanding and are beyond the capabilities of most instructors [56]. Additionally, input variables are differentially predictive for each instructional context [21], necessitating a more nuanced and contextualized approach to information discovery.

To this end, we are in the initial stages of testing an approach that helps instructors uncover hidden relationships in data about their students. We are combining the data they have already collected in our system with the machine learning application programming interfaces (APIs) provided by BigML (https://bigml.com). Our approach involves instructors selecting data to analyze, based on their pedagogical context and intent, using a drag-and-drop graphical user interface where they can also transform and/or combine data (Figure 3) and select a target (dependent) variable (e.g. an interim grade). The system then runs a series of machine learning algorithms (see section 3.2) against these data and returns analysis results for instructors to interpret in their context. This approach is more user-friendly than a similar system designed by Pedraza-Perez et al. [46], and can also include data beyond the LMS. This process may provide novel insights into curriculum design and delivery, such as visual and statistical identification of factors that impact student outcomes, and identifying patterns in performance across multiple courses with different course designs.

Other possible insights are outlined in section 3.2.

Figure 3. Attribute selection interface allowing instructors to select, transform, and combine data they wish to analyze.

In terms of Greller and Drachsler’s [24] framework, this nascent approach adds algorithmic capability to the instrument to provide certain stakeholders with possibly hidden information, beyond that of prediction. However, it requires higher data literacies and competencies, such as critical evaluation skills (internal limitations [24]). By working through the other steps of the process already outlined (namely data selection, collection, extraction, and basic analyses), our presumption is that instructors will have gained some of these competencies. Together, we see this as a combination of LA and educational data mining, where instructor judgment is empowered through leveraging machine learning [52].

2.4 Preliminary outcomes

The first version of our system was trialed with four courses in 2012. Since then, it has been adopted in 14 disciplines and 58 courses, covering over 20,000 students. This approach has allowed our system to evolve functionality through collaboration with the instructors who are using it. Although lacking empirical data, anecdotal feedback indicates that uptake is, in part, due to the customizability and afforded actions (i.e. usefulness [12]) and ease-of-use of the system. This contrasts with the issues highlighted by Lonn et al. [38] around their scaled-up LA system with a vendor-locked approach not being "nimble enough to be responsive to idiosyncratic cases" [38, p238]. The interventions for students, using our system, have contributed to sustained improvements in retention as well as overall performance (Figure 4). Now that instructors have more experience working with their data, we are collaborating with them to expand opportunities afforded by our system to further understand, optimize, and transform their teaching.

Figure 4. Outcomes from a representative Science course. Percentage of students (y-axis) in each outcome category (HD, high distinction; DI, distinction; CR, credit; PS, pass; FA, fail; attrited, i.e. left the course) is presented against calendar years where the course was offered.

3. UNDERSTANDING, OPTIMIZING, AND TRANSFORMING TEACHING

3.1 Teaching practices

Too often the student experience at university is one of isolation from instructors, which is especially poignant for students transitioning to higher education where instructors can appear disconnected [30]. While LA may exacerbate this situation by defocusing the human aspects of learning [31], our approach encourages instructors to break this pattern: hence the name of our system, the Student Relationship Engagement System (SRES). The strength of the SRES lies in the ability for instructors to customize analyses to the needs of their course and students. One of the primary goals of the SRES is to personalize communication with students and engage them in conversations about their learning. This is particularly important when operating at scale with large cohorts, as data-driven personalizations are a key factor in promoting student engagement [7]. We see this as a blending of Greller and Drachsler’s [24] objectives of reflection and prediction, where timely data are extracted to aid co-reflection by instructors and students. We find that this approach can also encourage more meaningful student-faculty contact, thus addressing a constant warning in the field that students’ internal conditions must be taken into account [20].

3.2 Instructional and curricular design and delivery

Currently, we are trialing several newer developments in the SRES in our own courses to explore further ways to support decision making [24] about instructional and curricular design and delivery. Here, we present three proof-of-concept examples that attempt to derive meaning in our contexts by analyzing real course data (Table 1) using machine learning tools. Instructors can select (Figure 3) data that are most relevant in their contexts (for example, mid-term test grade, session length in the LMS, attendance count early in the semester, average grade of online quizzes early in the semester,
activity in online forums, etc), and apply these tools to uncover hidden patterns. For example, what relationship is there between class attendance, different aspects of online engagement, and test grades?

Table 1. Description of sample variables.

<table>
<thead>
<tr>
<th>Data/variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piazza_questions</td>
<td>number of questions asked on online forum</td>
</tr>
<tr>
<td>C_COURSE_ACTIVITYIN</td>
<td>Total session length in LMS</td>
</tr>
<tr>
<td>HOURS</td>
<td></td>
</tr>
<tr>
<td>online_worksheets</td>
<td>Total score in formative online quizzes</td>
</tr>
<tr>
<td>final_grade</td>
<td>Final course grade</td>
</tr>
<tr>
<td>early_attendance</td>
<td>Attendance pattern at first four practical classes of semester</td>
</tr>
<tr>
<td>Test_1</td>
<td>Mark in first mid-term exam/test</td>
</tr>
<tr>
<td>early_prework_</td>
<td>Average of first four pre-class online quizzes</td>
</tr>
<tr>
<td>quizzes</td>
<td></td>
</tr>
<tr>
<td>Piazza_answers</td>
<td>Number of replies posted to online forum</td>
</tr>
</tbody>
</table>

3.2.1 Decision trees

Decision tree algorithms generate hierarchical conditions-based predictive models that attempt to explain conditions or patterns in data that lead to a particular outcome [49]. In our context, the decision tree discovered through machine learning suggested that early quiz performance (which was only worth a low proportion of the final grade) was an important factor in student success (Figure 5). While instructor intuition about their students may predict this, there is value in having data demonstrating various ‘paths to success’. Additionally, when one considers that each of these quizzes are worth only 0.65% of a students’ final grade (again emphasizing the importance of context and design), this data-enabled discovery becomes the grounds for supporting the evidence-based practices of emphasizing time on task and continuous assessment. These analyses are now driving pedagogical changes (e.g. decisions on provision of feedback in these quizzes versus no feedback) to improve student performance. For instructors, this approach not only helps identify struggling students, but also supports decisions about learning activities and assessing course effectiveness [50; 51].

In many cases in LA and educational data mining, decision tree algorithms are used purely as opaque models for prediction of student outcomes [e.g. 27; 32]. However, this does not take full advantage of the fact that decision trees are one of the few machine learning algorithms that can produce easily human-interpretable and -understandable predictive models, in the form of choices and rules [49]. As in our example, analysis of LMS interaction and completion data with decision trees can reveal behavioral and early-performance characteristics of high- and low-performing students, and allows instructors to adapt their courses and interventions based on this information [17; 50].

3.2.2 Association rule mining

Association rule mining reveals typically hidden patterns in data that commonly occur together [4; 51]. These patterns are expressed as rules or relationships of varying strength from antecedent to consequent conditions. Our application leverages a BigML visualization to graphically represent these rules. In our context, association rule mining provided evidence that lower in-class attendance was associated with lower online activity, and that lower online activity was a central node between other disengagement measures (Figure 6, main network). On the other hand, common relationships were also found between strong mid-term test marks, high online quiz marks, and strong pre-class quiz performance (Figure 6, bottom-left network), although interestingly high online activity was not included. While again this might seem obvious, this data-driven finding could trigger curriculum or instructional design changes to better engage students [48]. The associations discovered could also inform intervention strategies by identifying linked problem areas [50].

3.2.3 Clustering

Clustering algorithms group members of a dataset (in this case, students) together based on similarity between their data [4]. In our context, the clustering algorithm identified a group of mid-performing students who had high engagement with an online forum (Piazza_questions, Figure 7, cluster 4), compared to relatively low engagement from higher-performing students. Interestingly, this cluster was differentiated from another cluster of mid-performing students, who had, overall, much lower online engagement (Figure 7, cluster 0). This finding counters the
common understanding that higher discussion forum engagement is correlated with higher performance [e.g. 39], and again emphasizes the importance of considering contextual and pedagogical factors [21]. In our context, the online forum functioned in a question and answer format, which may help to explain why a cluster of poorer-performing students had higher engagement, i.e. posting of questions. Together, these analyses and their data-driven findings can be powerful for instructors because they help to support or refute a priori assumptions about their students, pedagogical strategies, and curricular design. Clustering may also provide insight into behaviors common to groups of differentially-performing students [1]. Some have even suggested that clustering students based on observed behaviors may assist formation of congruous student groups [50].

3.3 Cultural shifts
Our approach leveraged existing instructor needs to introduce them to a data-driven LA system, the SRES. A consequence of doing so has been to force them to think about their contexts and the relevant data. We are currently analyzing these instructor capability outcomes, as others have suggested that “implementing early and to small scale, even if inadequately, will build capacity” [10, p38]. Our approach certainly started small-scale, and was perhaps somewhat inadequate in not providing automatic access to the plethora of data available in LMS logs and the SIS. However, our hope is that by starting small and introducing instructors to data-driven ways of operating, we can introduce them to deeper LA ‘by stealth’ and gradually expand their capabilities, in parallel with expansion of the system’s capabilities.

4. CONCLUSION
The field of learning analytics is under unprecedented pressure to effectively bridge the gap between technological capacity and tangible improvements of the student experience. The shift towards tools that enhance current instructional practice is occurring. In this paper we have presented the evolution of the Student Relationship Engagement System following an organic and instructor-centric approach. The platform provides a high level of control over data collection and processing as well as direct control over the actions derived from the analysis. The current uptake of the tool across disciplines suggests its suitability to promote data literacy skills and a culture of data-supported innovation. As further avenues to explore, we have identified the need to increase the understanding of how instructors are empowered through data-driven analysis of learning designs and delivery.

5. ACKNOWLEDGMENTS
We thank many instructors and student support staff for their input into the process, countless students for their enthusiasm, and the Australasian Society for Computers in Learning in Tertiary Education (ascilite) Learning Analytics Special Interest Group for their support.

6. REFERENCES


ABSTRACT
Learning analytics, with a risk management approach, provides relevant and actionable information to teaching and administrative staff to make evidence-based decisions in curriculum and program quality improvement. This paper outlines the development and pilot implementation of a risk management model with an online feedback system in a research-intensive Australian university. Providing teachers and executives with the opportunity, facilitated by the essential IT infrastructure, to contextualise data and to document their response to the identified risks is a proactive approach to empower staff to make enhancements to their teaching practices, and to influence academic management. In addition, the opportunity for individual teaching staff to examine the progress of their own courses is a fundamental step in curriculum and program quality improvement. Positive feedback has been received in terms of the ease of access and opportunity provided to contextualise the risk. Future development will incorporate dynamic data from different sources, such as student participation in the learning management system, to build a holistic risk management framework in teaching and learning.

CCS Concepts
• Social and professional topics—Professional topics—Management of computing and information systems—Project and people management—Systems analysis and design

Keywords
Risk management; analytics; teaching; curriculum; quality assurance.

1. INTRODUCTION
In the current highly competitive environment, new modes of governance that emphasise performance, quality and accountability of student learning and experience have become common practice in higher education institutions (HEIs) [1, 2, 3]. HEIs are under pressure to demonstrate their teaching quality with increasing degrees of accountability and quality assurance expectations [4]. In the Australian higher education system, the Australian Qualifications Framework (AQF) provides criteria for different types of qualifications, as well as the expected learning outcomes, skills and knowledge required for each qualification level [5]. Together with the Tertiary Education Quality and Standards Agency’s (TEQSA) risk assessment framework [6], these national frameworks evaluate and monitor the teaching, learning and assessment quality of HEIs [7]. Linking these national requirements to the field of learning analytics, the emergent question is how to best use the “measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and environments in which it occurs”, a definition of learning analytics by the Society for Learning Analytics Research [8], in the context of curriculum and program quality enhancement. Curriculum based analytics is defined as the actions of collecting, analysing and interpreting key stakeholder data, such as student admission, retention, satisfaction, course and program structure, and assessment, across multiple offerings to enhance the development, implementation and evaluation of curriculum and program quality [9]. Active engagement from university executives, academics and students in using evidence-based practices to evaluate curriculum design and make decisions about curriculum and program reforms is pivotal to the success and sustainability of efforts to curriculum and program quality improvement [9].

This paper outlines the development of a risk management framework in the revised Curriculum and Teaching Quality Appraisal (CTQA) process at a research-intensive Australian university, which will be fully implemented for the academic year 2016. The pilot phase of implementation concluded in January 2016. The paper also discusses how a risk management model, better facilitates data-driven decision making, and curriculum and program quality improvement, compared with the traditional performance management framework. Alongside with the risk management framework, a series of interactive reports and dashboards for University Executives, Program Convenors, Course Coordinators and teaching staff are also developed. This is an attempt to provide comprehensive, relevant, and actionable information to key stakeholders to encourage the use of evidence-based practices, as well as to assist individual teaching staff to examine the success of a course which is fundamental to curriculum and program quality improvement. Last, but not least, an online feedback system also acts as an effective means to close the loop of the risk management process. Staff are provided with the opportunity to document their response to the data provided via the online feedback system. Risk management with active participation from staff empowers the University community to make data-driven decisions in considering student learning and experience.

2. BACKGROUND
The CTQA is a key component of this University’s overall quality assurance process in teaching and learning. It is undertaken on an
annual basis, and involves an evidence-based consideration of the overall quality of its teaching programs. The previous CTQA process was established in 2008 and was based on a performance management model, which identified programs that did not meet the specified performance indicators. Since 2008, there have been changes in both the external and internal higher education environment. In order to align the University’s teaching and learning quality assurance process to the national agenda, and to maximise the internal benefits of this quality assurance process, a decision was made to revise the CTQA process.

3. THE REVISED CTQA PROCESS
The principle of the revised CTQA process is to collect relevant data, and undertake critical and diagnostic data analyses which focus on trends, issues, actions taken and outcomes to support ongoing curriculum and program quality improvement. The rationale of selecting a risk management framework, instead of using a performance management framework, is based on the concept that through identification and management of risk, it can impact performance. A performance management framework focuses on the measurement of the actual results and their deviation from the targets [10]. Academic staff reactively respond to the identified areas for improvements and implement strategies in an attempt to reach the university’s targets. A number of academic staff previously expressed their resentment to a performance management framework, as they felt that they should not be penalised for the poor performance of the indicators that they have limited control on, such as the student load. In contrast, a risk management framework emphasises the importance of proactive actions for risk mitigation [10]. The premise of this framework lies in the fact that when an indicator is identified as at risk, it may not necessarily signal poor performance of a specified course/program. Instead, the identification of risk provides an opportunity for the staff to mitigate and contextualise the risk, and make a conclusion of whether current actions are adequate to address the identified risk or further actions are required. Academic staff who participated in the pilot welcomed the change from a performance to a risk management framework, as it lessens the punitive perception of the process and encourages conversations between staff and senior executives to investigate the identified risks.

The first step in developing the revised process is key stakeholder consultation to ensure that relevant and actionable information is provided to teaching staff and University executives. A broad consultation was conducted with the Associate Deans (Academic) in each Faculty, Chairs of Teaching and Learning Committees of each School, Heads of Schools, Program Convenors and Course Coordinators. Through committee meetings, presentations and individual discussions, a community of teaching and administrative staff was encouraged to engage in making evidence-based decisions to improve student learning. Based on the outcomes of the consultation, in alignment to the TEQSA risk assessment framework [6] and the University’s strategic plan and policies, separate sets of risk indicators were defined for courses and programs. The future plan is to include dynamic data from other sources, such as the student learning management system, as the model evolves in time.

3.1 Risk Indicators for Programs
The set of risk indicators for programs and the rationale, based on the TEQSA risk assessment framework [6] and the University’s strategic plan and policies, are outlined as follows:

1. Year 12 Student First Preferences to a Program with an Overall Position (OP) 1-5 (OP ranges from 1 – the highest to 25 – the lowest): This indicator shows the ability of a program at this University to attract students with high academic achievements in comparison to its competitors. A significant decrease may signal a decline in the quality or value of the program offered. However, recruitment strategies and employment in a profession need to be considered when interpreting this indicator.

2. Student Load: An unplanned significant increase in student load could potentially impact on the quality of student experience. Conversely, an unplanned significant and continuing decrease may signal a decline in the quality of the programs offered as perceived by prospective students.

3. Domestic Retention: A low retention rate may suggest that there are potential quality issues in the process of student admission, teaching and learning, and the overall student experience. Prompt actions to address early attrition are critical to minimise the compound effect on attrition in the later years of the program.

4. International Retention: Rationale same as Indicator 3.

5. Full-Time Employment after Graduation: A very low employment rate could indicate that students may not be well-equipped with the necessary graduate attributes for successful transition to the next stage of their chosen profession. However, volatility in the labour market needs to be factored in when interpreting this indicator.

6. Overall Satisfaction: A core quality indicator in higher education and provides an overall guide as to whether the program met student expectations. Poor satisfaction is a risk to the institution’s future market demand.

7. Pass Rate: A core indicator of student success and quality of the academic environment. When the pass rate is at very high/low levels, it may suggest that there are potential quality issues in student teaching and learning, and/or the overall student experience.

8. Completion Times: This indicator represents one dimension of the effectiveness of the delivery of educational services. Number of students in different study mode (full-time or part-time) need to be factored in when interpreting the results. Prompt actions to identify at-risk students, at an early stage, who are not being able to complete a program and to provide them with appropriate support are essential to minimise the possibility of reaching the stage of non-completion.

3.2 Risk Indicators for Courses
The set of risk indicators for courses and the rationale, based on the TEQSA risk assessment framework [6] and the University’s strategic plan and policies, are outlined as follows:

1. Enrolments: An unplanned significant increase in student enrolments could potentially impact on the quality of student experience. Conversely, an unplanned significant and continuing decrease may signal a decline in quality in courses offered as perceived by prospective students.

2. Pass Rate: A core indicator of student success and quality of the academic environment. When the pass rate is at very high/low levels, it may suggest that there are
potential quality issues in student teaching and learning, and/or the overall student experience.

3. Student Evaluation of Course and Teacher (SECaT) Response Rate: This is one of the indicators to reflect student engagement with the course in providing feedback. However, strategies implemented and timing at which the SECaT was administered need to be considered when interpreting this indicator.

4. Average SECaT Score for Q1: I had a clear understanding of the aims and goals of the course.

5. Average SECaT Score for Q2: The course was intellectually stimulating.

6. Average SECaT Score for Q3: The course was well structured.

7. Average SECaT Score for Q4: The learning materials assisted me in this course.

8. Average SECaT Score for Q5: Assessment requirements were made clear to me.

9. Average SECaT Score for Q6: I received helpful feedback on how I was going in the course.

10. Average SECaT Score for Q7: I learned a lot in this course.

11. Average SECaT Score for Q8: Overall, how would you rate this course?

For indicators 4 to 11, these are core quality indicators to provide a guide as to whether a course met student expectations. Prompt actions to address low student satisfaction scores in specific areas will assist in identifying the issues and implementing appropriate strategies to minimise student attrition and increase overall student satisfaction over time.

Using separate sets of risk indicators for courses and programs enable individual Course Coordinators and teaching staff to examine the success of the courses that they have taught in a semester. This is an obvious progression from the former CTQA, as previously only faculty- and school-level data were available with limited individual course/program information. Nevertheless, individual courses are the building blocks of the curriculum and program. The provision of course-level data will further engage teaching staff in the curriculum and program quality improvement.

Most importantly, the key feature of this risk management model is the opportunity provided for teaching and administrative staff to contextualise and mitigate the identified risk, to make a decision on whether the identified risk should be closely managed, or the risk is expected and actions have been in place to minimise its impact. Staff can also document their feedback to the data provided via an online feedback system which will be further discussed in Section 5. This active engagement from teaching and administrative staff in the revised CTQA process encourages them to reflect on the relevant student learning data and adopt a continuous improvement approach to teaching and learning. Staff are able to review individual program data on an annual basis, and individual course data on a semester basis. By using trend data of each program and course, teaching and administrative staff are proactively managing risks rather than reactively managing performance. The revised process not only identifies the at-risk courses and programs, but also the minimal-, neutral-, increasing-risk courses and programs.

The opportunity to explore the risk indicators, which contribute to a heightened risk for increasing-risk courses and programs, as well as those result in a lesser risk for neutral- and minimal-risk courses and programs, allows staff to adopt a proactive approach in managing risks. For example, course staff are able to modify their teaching practices, such as the use of a flipped classroom model to allow more interactive sessions with students, in anticipation of an increasing trend of student enrolments. Unlike the reactive management approach, staff only formulate a solution after an increase in student enrolments is evident. The revised CTQA is an annual process that focuses on data-driven decision making through contextualising and mitigating risks, evidence-based action planning, and revisiting and evaluating proposed actions in subsequent annual reviews.

This section outlined the development of the revised CTQA process. The next section will focus on how to create visualisations that encourage a community of teaching and administrative staff to engage in making evidence-based decisions to improve student learning at both course- and program-levels.

4. DATA VISUALISATION

The ultimate goal of data visualisation is to provide clear and useful information to the targeted audience. However, it is an iterative process to find the best way to visually present data to meet the needs of the stakeholders [11]. Being able to easily access the required data is the key starting point to make data-driven decisions in teaching practices, curriculum design and academic program delivery. Therefore, the aim of the first iteration of data visualisation for the revised CTQA process is to provide University executives, academic and administrative staff with quick and easy access to both high-level overview and detailed-level information about the courses and programs offered, with the incorporation of simple visual cues, such as differential colour coding to provide greater ease in interpretation of risks. Three levels of data visualisation are created. The first level is the new executive dashboards and reports (see Figure 1), which provide University executives with an overview of the minimal-, neutral-, increasing- and at-risk courses and programs.

![Figure 1. A snapshot of a program executive dashboard.](image)

The second level is the new Faculty and School dashboards and reports (see Figure 2), which provide the Associate Dean (Academic) of each Faculty, Heads of Schools, Chairs of Teaching and Learning Committees, Program Convenors and Course Coordinators with an overview of the minimal-, neutral-, increasing- and at-risk courses and programs offered within their Faculty and School.
The third level is the detailed course/program report for an individual course/program (see Figure 3). Previously, Course Coordinators or individual teaching staff were required to collate and compile their own reports from the available and relevant teaching and learning data about a course/program. The new reports consolidate all the required data and provide the stakeholders with an integrated report for each course/program.

![Figure 2: A snapshot of a Faculty dashboard.](image)

### 2015 Program Report:

<table>
<thead>
<tr>
<th>Category</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onsite</td>
<td>1200</td>
<td>1300</td>
<td>1400</td>
</tr>
<tr>
<td>Online</td>
<td>800</td>
<td>900</td>
<td>1000</td>
</tr>
<tr>
<td>Total</td>
<td>2000</td>
<td>2200</td>
<td>2400</td>
</tr>
</tbody>
</table>

Staff, who have access to these modified detailed course/program level reports, are already actively using them to explore the strengths and limitations of their courses/programs. They have also provided positive feedback about the reports and process. This unified approach reduces a considerable amount of administrative time in collating data. As a result, they can use the time to engage in data-rich conversations focused on improving curriculum and pedagogical practices, reflection and decision-making as to how to improve student learning in their course/program.

In addition, these three levels of reports and dashboards are interrelated, which provide the opportunity for key stakeholders to either drill down to the details of the strengths and limitations of a course/program, or zoom out to look at the relationship of a particular course/program to the relevant group of courses/programs. These three levels of data visualisation aim to generate conversations, initially, between individual teaching staff, and gradually expand the conversations with the Course Coordinators and Program Convenors, and collaborate to make evidence-based decisions to improve teaching practices, curriculum and program quality.

Apart from the three levels of data visualisation, it is essential that reasonable requests of teaching and learning data from individual teaching staff are adequately addressed. Nevertheless, courses are the building blocks in a curriculum and program. Providing individual teaching staff with customised reports could, in fact, extend their engagement in the curriculum and program quality improvement process. The additional data that an individual teacher requests may also be beneficial to other courses/programs. Hence, consideration should be made to incorporate those in the new iteration of the reports and dashboards. An example is the request of analysing the distribution of assessment types (that is, examinations, presentations, essay writing) in the compulsory courses of a program. These relevant and actionable data about assessment allows teaching staff and Program Convenors to have a holistic view of student learning and assessment experience in a program. When data revealed that a large percentage of assessment was examinations, one would expect that investigation into the rationale of the existing assessment regime is conducted and changes will be made to provide students with the opportunity to demonstrate their knowledge and skills via different modes of assessment. This process is the start of a continuous improvement approach to teaching and learning, in which assessment is a core component, and should be encouraged in other Faculties/Schools.

The first iteration of data visualisation for the revised CTQA process only includes static and historical data about student learning. In the second iteration of data, the aim is to create interactive reports and dashboards with automatic drill-down functions to reveal dynamic data, such as student access patterns to online resources and assessment, and student and teacher engagement patterns with the Learning Management System (LMS). As part of the curriculum and program quality improvement, these additional data about student interactions with online resources and technologies would provide insight into the optimal structure of a course/program that will engage and motivate students to learn.

### 5. ONLINE FEEDBACK SYSTEM

The continuous process of reviewing, reflecting and proposing new solutions is a core part of the quality improvement process. One of the strategies to engage a community of teaching staff in curriculum and program quality improvement is to empower them to complete the revised CTQA process loop via an online feedback system (see Figure 4). The purpose of this online feedback system is to provide an opportunity for staff, firstly, to provide contextualised information around selected courses/programs, such as those identified as increasing- or at-risk. Secondly, to confirm or disconfirm the identified risk and determine the residual risk for relevant courses/programs as minimal-, neutral-, increasing- or at-risk. Finally, to document proposed actions that will be undertaken to address the confirmed risks.

![Figure 4: A snapshot of the online feedback system.](image)

The documentation of feedback is pivotal in the continual cycle of curriculum and program quality improvement, as the feedback
collected from academic staff, Course Coordinators/Program Convenors, and Faculty Executives establish the basis for the required actions to address the risks. All key stakeholders can review their feedback and document progress in comparison to the previous release of data. The program reports and dashboards are updated on an annual basis, whereas the course reports and dashboards are released after the conclusion of a semester. Once these reports are available, each Faculty and School will have the autonomy to decide which group of courses or programs to focus on in order to enhance their delivery, and the approach they use in response to the data provided. This autonomy provides opportunities to generate conversations among staff to develop a Faculty/School-wide response to the issues identified and raised during the review process and the ability to apply the learnings of best practice to other courses or programs requiring intervention and/or reward. In summary, this online feedback system is developed to enable collection and consolidation of feedback and proposed actions to address risk.

6. FEEDBACK FROM PILOT PROCESS

The purpose of this pilot was to ascertain the effectiveness of the new process and associated communication strategy. The information gathered provided an opportunity for the Learning Analytics and Evaluations teams to mitigate risks associated with a University-wide implementation, and facilitate resolutions to any identified issues prior to the formal rollout of the new process across the University.

Feedback from the participants was positive. They appreciated the integrated course/report program which provide all the relevant data for a particular course/program. This unified approach reduces a considerable amount of administrative time in collating the data from different sources. In addition, the Faculty/School reports provided an overview of the minimal-, neutral-, increasing-, and at-risk courses/programs in a Faculty/School, which assists in directing attention, resources or recognition to particular groups of courses/programs. The identified courses/programs risk dashboard appeared to have face validity based on the participants’ knowledge and experience. Participants also acknowledged that the revised process provides them with the opportunity to contextualise and mitigate the identified risk, to make a decision of whether the risk should be closely managed, or the risk was expected and actions have been in place to minimise its impact via the online feedback system.

7. CHALLENGES

This paper presents how learning analytics methodologies play a pivotal role in developing understanding, optimising and transforming courses/programs, using a risk management framework with an online feedback system. The two major challenges encountered in the development of the revised CTQA process are the institutional culture change from a performance management to a risk management framework, and collaboration with the business intelligence and IT departments. The lessons learnt in developing and implementing the pilot revised CTQA process revealed that effective communication, with the support from the University senior executives, is the best strategy in dealing with these challenges. Although a cultural shift in an institutional-wide system can take up to a few years, consistent communication and clear expectations from all key stakeholders involved are the important incremental steps in shifting the culture from a performance to a risk management model. In terms of collaboration with business intelligence and IT departments, the message needs to be focused on the value-adding role of learning analytics to the current business intelligence and IT functions, instead of being perceived as a threat to their operation.

The development of the risk management framework, and its associated reports and dashboards and online feedback system, is still evolving. Continual support to the teaching and administrative staff in terms of understanding the data, as well as possible pedagogical enhancement that they could implement in their courses/programs, is required to sustain their engagement with the data to make evidence-based decisions in the curriculum and program improvement process. Future development will incorporate dynamic data from additional sources, such as student participation in the LMS, to build a holistic risk management framework in teaching and learning in higher education.

8. REFERENCES

LMS Course Design As Learning Analytics Variable

John Fritz
Univ. of Maryland, Baltimore County
1000 Hilltop Circle
Baltimore, MD 21250
410.455.6596
fritz@umbc.edu

ABSTRACT
In this paper, I describe a plausible approach to operationalizing existing definitions of learning management system (LMS) course design from the research literature, to better understand instructor impact on student engagement and academic performance. I share statistical findings using such an approach in academic year 2013-14; discuss related issues and opportunities around faculty development; and describe next steps including identifying and reverse engineering effective course redesign practices, which may be one of the most scalable forms of analytics-based interventions an institution can pursue.

Categories and Subject Descriptors
K.3.1 [Computer Uses in Education]: Collaborative learning; Computer-assisted instruction (CAI); Computer-managed instruction (CMI); Distance learning.

General Terms
Algorithms, Design, Human Factors, Measurement, Standardization, Theory

Keywords
Course Design, Instructor Pedagogy, Learning Analytics Methodology

1. INTRODUCTION OF PROBLEM
Given wide spread use of the learning management system (LMS) in higher education, it is not surprising this form of instructional technology has frequently been the object of learning analytics studies [1, 2, 3, 4, 5]. While methods and results have been mixed in terms of predicting student success, let alone leading to actual, effective and scalable interventions, there is one potential LMS analytics variable that has received comparatively little attention: the role of course design.

Part of the problem is how to operationalize something as theoretical, subjective or varied as instructor pedagogy. Indeed, Macfadyen and Dawson [6] attributed variations in “pedagogical intention” as a reason why the LMS could never serve a “one size fits all” dashboard to predict student success across an institution. Similarly, Barber and Sharkey [7] eliminated theoretical student engagement factors such as self-discipline, motivation, locus of control and self-efficacy because they were “not available” (i.e., quantifiable) in the LMS data set, which was their primary object of analysis. Basically, how does one quantify course design that seems qualitatively different from usage log data like logins?

Despite these operational challenges, some of the most frequently cited LMS analytics studies referenced above actually provide a surprisingly uniform characterization of course design that can be roughly broken down into three broad, but distinct categories:

1. User & Content Management (e.g., enrollment, notes, syllabi, handouts, presentations) 1
2. Interactive tools (e.g., forums, chats, blogs, wikis, announcements)
3. Assessment (e.g., practice quizzes, exams, electronic assignments, grade center use)

If we are willing to accept LMS course design as an aspect of instructor pedagogy – and accept student LMS activity as a proxy for attention, if not engagement – then it may be possible to use one to inform the other. Specifically, patterns of student LMS behavior around tools or functions could retroactively shine light on implemented course design choices that align with the broad, research-based LMS course design types described above.

For example, if students in one course appear to use the online discussion board more than students in another course, could one reasonably assume that instructors of the two courses varied at least in their conceptual value and effective use of this interactive tool? Perhaps this is evident by how instructors differ in their weighting or reward for the discussion board’s use in the course’s grading scheme, or model and facilitate its use, or simply enable it as a tool in the LMS course’s configuration. Admittedly, the challenge is determining how much variance in student LMS course usage is statistically significant or attributable to and indicative of instructor course design. For assessment purposes, though, these three broad LMS course design types (content, interaction and assessment) provide at least a theoretical way to operationalize variability in faculty LMS course design and usage.

While there may be a default institutional LMS course configuration most instructors blindly accept, in trying to explain why one tool or function is used by students more in one course vs. another, it seems odd that we shouldn’t be able to consider the pedagogical design choices of the instructor as an environmental factor that may impact student awareness, activity and engagement. True, this may also reflect an instructor’s capability or capacity to effectively express his or her pedagogy in the LMS,

---

1 Dawson et al (2008) proposed a 4th type of LMS use called “administration” that roughly equates to course logistics of enrollment, file management, etc. For convenience, I’ve combined this into the “user & content management” category.
but to simply ignore the possible impact of course design on student engagement seems un-necessary and disingenuous if we want to use learning analytics to predict and hopefully intervene with struggling students. If students who perform well use the LMS more, do we not want to know what tools, functions and pedagogical practices may facilitate this dynamic?

2. SOLUTION & METHOD
Despite the striking similarity in how several LMS-based analytics studies have categorized LMS course design practices (if not pedagogical intent), what’s needed is a plausible, systematic approach to operationalize these common definitions.

2.1 Weighted Item Count by Design Type
Conveniently, Blackboard used these same research-based definitions of course design for its Analytics for Learn (A4L) product. Specifically, A4L’s “course design summary” is a statistical comparison of a Bb course’s relative, weighted item count compared to all courses in a department and the institution based on the three major item types found in the LMS analytics literature. Essentially, all items in any Bb course, such as documents or files, discussions or chats, and assignments or quizzes, are grouped into 1) content, 2) interactive tools or 3) assessments. Then, A4L’s course design summary uses a simple algorithm to categorize all courses into department and institutional statistical quartiles through the following process:

1. Sum all course items by primary Item Type (e.g., Content, Tools, Assessments).
2. Multiply the group total using a weighting factor (wf): Content (wf = 1), Interaction (wf = 2) and Assessments (wf = 2).²
3. Statistically compare each course to all other courses in the department and all other courses across the entire institution.
4. Tag each course with a quartile designation for both the department and institution dimension.

Again, the “course design summary” is already provided in A4L and is really just a way of categorizing how a course is constructed, compared to all courses in the department and across the institution, not necessarily if and how it is actually used by students. To understand and relate student activity to course design, we need to calculate a similar summary of student activity from existing A4L measures.

2.2 Student Activity Summary
Blackboard Analytics 4 Learn (A4L) contains several student activity measures that include the following:

- Course accesses after initially logging into the LMS;
- Interactions with any part of the course itself, equivalent to “hits” or “clicks”;
- Minutes using a particular course (duration tracking ends after 5 minutes of inactivity);
- Submission of assignments, if the instructor uses assignments;
- Discussion forum postings, if the instructor uses discussions.

However, for calculating the companion student activity summary to correlate with A4L’s course design summary, I have only used the first three measures (accesses, interactions and minutes) because ALL courses generate this kind of student activity, regardless of design type. Not all instructors use electronic assignments or discussion forums, but short of simply dropping a course, all students generate at least some activity that can be measured as logins, clicks or hits and duration.

To calculate the student summary, we must first convert each raw activity measure to a standardized Z-score, which shows how many standard deviations and in which direction a particular raw score is from the mean of that measure in a normal distribution of cases. Because the scale of each activity varies greatly during a semester (e.g., accesses or logins could be under one hundred, interactions or hits could be in hundreds and duration or minutes could be in the thousands), converting these variables to Z-scores allows us to compare and summarize them across measures more efficiently. It also allows us to identify and remove outliers, which for this purpose is defined as scores greater than three (3) standard deviations from the mean. The formula for converting Z-scores is as follows:

\[ Z = \frac{X - \mu}{\sigma} \]

The Z-score is equal to X (value of the independent variable) less \( \mu \) (the value of the class mean for X), divided by \( \sigma \) (the class standard deviation of X).

Accordingly, the steps to analyze and summarize student activity in all courses include the following:

1. Convert accesses, interactions and duration student Bb activity measures to Z-scores.
2. Average the combined student activity scores into a summary measure.
3. Assess the internal consistency of items using a Cronbach alpha test of reliability for each approach (e.g., comparing converted Z-scores).

In addition to student LMS activity and course design measures described above, I used a “threshold” approach to academic performance. Specifically, I used “C or better” final grade in a course and “2.0 or better” term grade point average (GPA) as dependent variables.

3. FINDINGS
3.1 Data
The participants for my study were all first-time, full-time, degree-seeking, undergraduate freshmen or transfer students starting their enrollment in Fall 2013. According to the UMBC Office of Institutional Research and Decision Support (IRADS), this included 2,696 distinct students (1,650 freshmen and 1,046
transfers) or 24.48% of all 11,012 degree-seeking undergraduates. The demographic distribution was as follows:

<table>
<thead>
<tr>
<th>Gender</th>
<th>Fresh.%</th>
<th>Trans.%</th>
<th>Total%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>57</td>
<td>48</td>
<td>54</td>
</tr>
<tr>
<td>Female</td>
<td>43</td>
<td>52</td>
<td>46</td>
</tr>
<tr>
<td>Subtotal</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th>Fresh.%</th>
<th>Trans.%</th>
<th>Total%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian</td>
<td>24</td>
<td>11</td>
<td>19</td>
</tr>
<tr>
<td>Black</td>
<td>11</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Hispanic</td>
<td>5</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>White</td>
<td>45</td>
<td>41</td>
<td>43</td>
</tr>
<tr>
<td>Other</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Unknown</td>
<td>6</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Subtotal</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 1: Study Sample, 2013-14 FT Freshmen & Transfers

3.2 Grades by Student LMS Activity
Generally, students who performed well academically in courses and a given term overall, showed a higher, statistically significant (p < .001) use of Bb compared to peers who did not perform as well. Specifically, using logistic regression to control for other factors such as gender, race, age, Pell eligibility, academic preparation and admit type, students were 1.5 to 2 times more likely to earn a C or better in Fall 2013 and Spring 2014, respectively. Similarly, students were 2.4 to 2.8 times more likely to earn a 2.0 term GPA in Fall 2013 and Spring 2014, respectively.3

3.3 Student LMS Activity by Course Design
Generally, students were much more active in Bb courses that used a wider array of Bb functionality. Specifically, after using linear regression, both the institutional course design quartile and instructor use of the grade center were statistically significant (p < .001) terms of freshmen and transfer LMS activity in both semesters. As indicated by the R² change, course design and grade center use contributed more than 20% to the overall models, whose adjusted R² of .265 and .239 explained 26.5% and 23.9% of the variance in student Bb usage for freshmen and transfers, respectively in Fall 2013. A similar pattern emerged in Spring 2014, with course design and grade center use contributing more than 22% to the overall models’ adjusted R² of .333 and .278, which explained 33.3% and 27.8% of freshmen and transfer student use of Bb, respectively.

3.4 Student Grades by Course Design
Generally, there was a statistically significant (p < .001) relationship for student academic outcomes based on the interaction of course design and student activity in the LMS. However, there was a marked difference in the Expected (B) or odds ratio for both groups of students across both terms, depending on whether I used institutional course design quartiles (ICDQ) or course grade center use as the covariate interaction effect with student Bb activity. For example, the ICDQ * Bb activity interaction effect never produced an odds ratio higher than 1.009, which translates into little more than 1 times the likelihood of earning a C or better final grade (essentially, a 50/50 chance).

By contrast, the odds ratio for the grade center use * Bb activity interaction effect was no less than a 1.571 (for transfers in Spring 2014) and reached a high of 2.455 (for freshmen in Spring 2014). This means that selected subsets of my sample of students had a 1.6 to 2.5 times chance of earning a C or better after controlling for other demographic and academic variables.

Using the same approach for 2.0 or better term GPA, the odds ratio for freshmen under the grade center * Bb activity interaction effect model was 2.610 and 3.504 for Fall 2013 and Spring 2014, respectively. This means freshmen were 2.6 to 3.5 times more likely to earn a 2.0 term GPA in their Bb courses that used the grade center. By contrast, the institutional course design quartile (ICDQ) * Bb interaction effect model remained essentially the same as the C or better findings described above.

4 The “Other” category is my combination of relatively small numbers for “International,” “Native American,” “Pacific Islander,” and “Two or More” UMBC Census Data categories.
4. DISCUSSION
While the correlation between LMS course design and student outcome is compelling, I cannot confirm or reject a hypothesis that it is a causal relationship. I’d want to study these relationships over a longer time, across the entire student population, and even replicate it at other schools. However, is it necessary to establish causality to leverage let alone prove a prediction? Desirable: yes. Necessary: I’m not so sure.

I tend to view LMS use – by faculty and students – as a real-time proxy for their respective attention, engagement and effort in the larger context of teaching and learning. As such, we’ve developed a simple “Check My Activity” (CMA) feedback tool for students allowing them to compare their own LMS activity with peers who earn the same, higher or lower grade for any assignment – provided the instructor uses the grade center. [3] After controlling for other factors (e.g., gender, race, academic prep, Pell eligibility, etc.) freshmen using the CMA were 1.7 times more likely to earn a C or higher final grade (p < .001), but transfers were barely 1 times more likely and the findings were not statistically significant. We also show students how active the LMS course is compared to other courses in the discipline, and recently extended this same view to faculty themselves. This way, everyone can decide how to gauge or interpret the importance of their own – or even an entire course’s – LMS activity in the context of that exhibited by others.

Additionally, Blackboard has developed a compelling predictive risk model based on this combination of student activity and course design to derive a student “engagement” indicator that is reflected in UMBC’s actual full-time freshmen and transfer retention status from Fall 2013 (see figures 3 and 4 below).[5]

![Figure 3: Freshmen Retention by Bb Learn Risk Profile, FA13](https://umbe_box.com/fritzpelashortpaperimages)

![Figure 4: Transfer Retention by Bb Learn Risk Profile, FA13](https://umbe_box.com/fritzpelashortpaperimages)

Notice how less successful but more engaged students (#3) are retained next year at higher rates than more successful but less engaged peers (#2), particularly transfers (figure #4). Moving forward, I can see the Bb integrated model becoming a valuable tool in studying the long-term impact of an LMS on student retention, persistence and graduation. If so, it might also reinforce the value of using the LMS as a real-time indicator of student engagement, not just the passive, one-way delivery of content for which it has typically been used.

4.1 Course Design as Scalable Intervention
If course design has a relationship with student academic performance, then faculty development could be a necessary first step toward a more scalable form of institutional intervention with at-risk students. In fact, in describing self-directed learning, Ambrose et al [8] suggest that “students must learn to assess the demands of a task, evaluate their own knowledge and skills, plan their approach, monitor their progress, and adjust their strategies as needed” (p. 191). However, instructors also need to be pedagogically ready and secure in their own roles as teachers to desire this kind of empowerment for their students, let alone seek it out by design.

For example, Robertson [9] proposed what is now considered a classic model for how faculty beliefs about teaching influence their evolving pedagogical practice, including the following stages:

- **Epocentrism** – focusing mainly on their role as teachers;
- **Aliocentrism** – focusing mainly on the role of learners; and
- **Systemocentrism** – focusing on the shared role of teachers and learners in a community.

If this evolution of thought and practice occurs at all among teachers, Robertson identifies telltale signs of the transformation. First, as faculty move from one stage to the next, they bring the benefits and biases of the previous stage. Second, they typically change their beliefs and practices only when confronted by the limitations of a current stage, which is brought about by teaching failures. Finally, the desire for certainty, stability and confidence either keeps faculty frozen in a current, status quo framework or drives their progression to the next one in an effort to avoid a potentially paralyzing neutral zone: “a familiar teaching routine that they have deemed inappropriate and with nothing to replace it” (p. 279).

Just as Robertson showed how faculty beliefs about teaching influenced their practice, Steel [10] showed how teaching beliefs influenced their perceptions about what they believe various instructional technologies will allow them to do. For example, using detailed case studies about faculty use of online discussions in an LMS, Steel illustrates the creative tensions between how faculty conceptualize teaching and how they perceive the affordances of web-based technologies like an LMS.

“The velocity of change in the affordances offered by learning technologies presents a significant challenge as does the minimal incentives available to university teachers to use technologies effectively in their teaching practices.” (p. 417)

Whether faculty like it or not, when they teach online or use online tools as supplements in their traditional classrooms, they...
also become webmasters. As such, they need to understand the potential affordances and limitations of web technologies as they attempt to express and implement their pedagogy in course designs. Steel argues that this “reconciliation process” between pedagogical beliefs and rapidly changing technology affordances “needs to be incorporated more fully into informal teacher development approaches as well as formal programs” (p. 417).

To me, faculty who are in Robertson’s “neutral zone” between “teaching failures” and “nothing to replace [them]” may be ripe for a course design intervention based on learning analytics, but only if they are aware of peers who believe to have a more effective approach. This is why and how learning analytics may be able to identify, support, promote and evaluate effective practices and practitioners, to serve as a standard by which faculty not only measure themselves, but also point to a way forward, by ideally helping students take responsibility for learning. Yes, technology may help, but per Robertson’s and Steel’s research, it may not do so unless faculty first believe that it can, enough so as to try or look for peers who have done so. Just as students taking responsibility for their learning is the only scalable form of learning, so too must faculty take responsibility for “teaching failures.” This includes being open to other pedagogical examples and working hard to master and implement them, which requires a willingness to explore, practice, refine and self-assess.

5. NEXT STEPS

In recent posts, e-Literate bloggers Michael Feldstein and Phil Hill lament the ubiquitous, but essentially boring LMS [11] and even equate it to the minivan of education technology that has long-lasting utility, but not much zip or cache [12]. But if we are willing to go beyond a conventional view of the LMS as more than a content repository or one-way (ego centric?) delivery of knowledge from instructor to student, we might just find that variations in student behavior can shine light on effective course design practices.

Toward this end, we are beginning to look at the LMS as a way to identify effective course design practices and practitioners. While a given semester is underway, we monitor positive outlier courses that appear to generate inordinately high student LMS usage. When the semester is over, we correlate final grades and follow up with instructors whose students may also be performing higher than peers within a department or the institution. To be sure, we conduct these qualitative interviews without necessarily relying on student LMS usage. But taken together, high student LMS usage and grade distribution analysis adds a real-time indicator of student engagement and academic performance that is no longer limited to the end of semester post-mortem.

Finally, as instructional technology support staff, it is not our job to shine light on instructors or course designs that could be better. We’ve learned instructors learn best from each other, but we can help by using the technology and methodology of learning analytics to identify and reverse engineer effective course design practices we wish all faculty knew about and would emulate. In this way, course redesign could be the most scalable form of analytics-based intervention any institution could pursue.

6. REFERENCES


Simple Metrics for Curricular Analytics

Xavier Ochoa
Escuela Superior Politécnica del Litoral
Vía Perimetral, Km. 30.5
Guayaquil, Ecuador
xavier@cti.espol.edu.ec

ABSTRACT
The analysis of a program curriculum is traditionally a very subjective task. Perceptions and anecdotes, faculty preferences and content or objectives check-lists are the main sources of information to undergo the revision of the structure of a program. This work proposes a list of simple metrics, that can be easily extracted from readily available academic data that contains the information about the actual interactions of students with the curriculum. These metrics, divided into time- and performance-related, are calculated at program level. The use of these metrics provides objective information in which to base discussions about the current state and efficiency of the curriculum. To exemplify the feasibility and usefulness of the metrics, this work presents some illustrative analysis that make use of the simple curriculum metrics.

CSC Concepts

• Applied computing → Education;

Keywords
Learning Analytics, Curriculum Analytics

1. INTRODUCTION
Learning Analytics has been traditionally applied to understand and optimize the learning process at course level. The learning process is analyzed through the captured interactions between students and instructor, content or tools. However, Learning Analytics are not restricted to act at this level. Adapted techniques, applied to different sources of information, could be used to understand and optimize learning at the program level, as exemplified by the works of Pechenizkiy et al. [8] and Gonzalo et al. [7]. Due to the interconnection between learning at the course and the program level. Program Analytics are an indispensable complement to traditional Learning Analytics in order to have an effective end-to-end learning process.

There are several sources of information that can be used to analyze a program curriculum. The first main categorization of this information responds to its level of objectivity. Surveys about needs, perceptions and sentiments are a common tool in curricula analysis. These surveys can be directed to students [6], faculty [7], alumni [3] or the labor market [5]. The result of these surveys provide subjective information. On the other hand, curriculum analysis could also employ factual data obtained from the curriculum and its usage. This data can be classified as objective information.

The objective information could be further classified in three main groups:

• Intrinsic: This is the information that is contained in the curriculum itself. For example, Sekiya et al., used the descriptions provided in the syllabi of several Computer Science curricula to compare their compliance to the Computer Science Curriculum recommendation from ACM [9].

• Extrinsic: This is the information external to the program that influence its content or structure. For example, Sugar et al., [10] found required multimedia production competencies for instructional designers, compiling information from instructional design job advertisements.

• Interaction: This is information that is generated when the students interact with the curriculum. The most common interactive information is the course selection and the grades obtained by students. This information is commonly refereed as student academic records. For example, Bendatu and Yahya [1], inspired by the curriculum mining idea of Pechenizkiy et al. [8], use student records to extract information about the course-taking behavior of students.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

PCLA ’16 Edinburgh, UK
© 2016 ACM. ISBN 123-4567-24-5678/08/06 . $15.00
DOI 10.475/123_4
perform illustrative curriculum analysis. The paper closes with conclusions and recommendations for further work.

2. CURRICULUM METRICS

Metrics are objective measurements or calculations of the characteristics of an object that simplify their understanding and analysis. While obtaining the value for a given metric is not the same as performing an analysis, metrics are the base of quantitative analytics. This work proposes curriculum interaction metrics than can be used to perform quantitative analysis of the status and efficiency of program curricula. These proposed metrics will be calculated exclusively from curriculum interaction data (academic records).

Academic records can be seen as the capture of the interactions between students and the program curriculum. Table 1 presents an example of the usual information present in the records of an academic institution. As a minimum, academic records contain information about two main interaction events: 1) the decision of the student to register in a given course during a given academic period, and 2) level of success of the student within the chosen courses. Due to these two different interaction aspects, the curriculum interaction metrics will be grouped into two sets described in the subsections below.

2.1 Temporal metrics

In academic programs where students have the flexibility to select courses at different temporal points during their studies, that selection could provide useful information for curriculum analyzers. This work proposes three metrics associated with the temporal information of the academic record.

2.1.1 Course Temporal Position (CTP)

This simple metric measure the average academic period (semester or year) in which a course is taken by the students of a program. This information can be used to establish the real position of a course in the program.

To calculate this metric, the raw academic period information needs to be converted into a relative value. For example, in a semester based program, if a student started their studies during the first semester of 2004 and he or she took the relevant course during the second semester of 2006, the relative period will be 6, because the course was taken on the sixth semester relative to the student’s first semester. To avoid to inflate the metric, only active periods, that is periods where the student has been actively pursuing the program, should be counted. Once relative period of a course is calculated for all the students, the average is calculated according to Equation 1, where \( RP_c \) is the relative period of the analyzed course \( c \) for a given student \( s \). Additionally, this metric can be configured to obtain the temporal position when a course was initially taken or when it was finally approved. Depending on the type of analysis, these two different metric versions could be useful.

\[
CTP_s = \frac{1}{N_{\text{student}}} \sum_{s=1}^{N} RP_{\text{grades}}_s \quad (1)
\]

2.1.2 Temporal Distance between Courses (TDI)

This metric establishes how many academic periods, in average, pass between a student taking two different courses. This information can be used to establish the actual sequence in which courses are taken.

While a simple way to calculate this metric will be to subtract the CTP of the second course from the first course, information about the actual time difference for each student is lost due to the average nature of the CTP. To calculate TDI (Equation 2, the relative periods of the relevant courses \( c_1 \) and \( c_2 \) are subtracted for each student. Then, the average is taken.

\[
TDI_{1,2} = \frac{1}{N} \sum_{s=1}^{N} (RP_{s,c_2} - RP_{s,c_1}) \quad (2)
\]

2.1.3 Course Duration (CDU)

This metric measures the average number of academic periods that students need to pass a given course. This metric provides information about the effect that a course has in the length of the program.

CDU is obtained by subtracting the relative period of the first time each student took the course \( (RP_{first}^s) \) from the relative period when the student finally passed \( (RP_{pass}^s) \) it and then averaging these values between all the students (Equation 3). A variation of this metric only consider the periods where the course was taken. In this case, the metric is identical to the average number of times that students need to repeat the course before passing.

\[
CDU_s = \frac{1}{N} \sum_{s=1}^{N} (RP_{pass}^s - RP_{first}^s) \quad (3)
\]

2.2 Difficulty metrics

Each time a student undertakes a course, performance information is captured and stored. The way in which this information is represented varies, but usually involved a grading scale. This scales could be categorical (letters, passing/not-passing etc.) or numerical (20 out of 100, 4 out of 5, etc.). The information stored in the student grades can be processed to produce useful information about the difficulty of different courses in the program. This work summarize some simple difficulty metrics proposed by previous works and propose two new profile-base metrics.

2.2.1 Simple Difficulty Metrics

The most basic metrics of the difficulty of a course are the passing rate (PR), the number of students that have approved the course divided by the number of students that have taking the course, and the average grade (AG), the sum of the grades of all students (converted to a numerical value) divided by the number of students. This metrics, however, are not comparable between courses because they depend of the group of students that the course. A course with relatively good students will have a better PR and AG than a course when only regular or bad students.

<table>
<thead>
<tr>
<th>Student Id</th>
<th>Course Id</th>
<th>Semester</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>200002608</td>
<td>ICM00604</td>
<td>2001-1S</td>
<td>6.75</td>
</tr>
<tr>
<td>200002608</td>
<td>ICM00604</td>
<td>2001-2S</td>
<td>8.32</td>
</tr>
<tr>
<td>200002608</td>
<td>ICF00687</td>
<td>2002-1S</td>
<td>4.23</td>
</tr>
<tr>
<td>200300341</td>
<td>ICM00604</td>
<td>2003-2S</td>
<td>9.01</td>
</tr>
</tbody>
</table>

Table 1: Example of Academic Records
Calulkins et al. [2] proposed more robust difficulty metrics. Two metrics, Grading Stringency, also called $\beta$ (Equation 4) and Multiplicative Magnitude, also called $\alpha$ (Equation 5) eliminate the bias introduced by the group of students taking the course by subtracting from the GPA of each student ($GPA_{c}$) the grade that he or she obtained in the course ($r_{sc}$) and averaging those values over all the students ($N$). However, the calculation of $\beta$ and $\alpha$ metrics assume a normal distribution of grades that is usually not the real case.

$$\beta_{c} = \frac{1}{N_{c}} \sum_{s=1}^{N_{c}} (GPA_{c} - r_{sc})$$  \hspace{1cm} (4)$$

$$\alpha_{c} = \frac{\sum_{s=1}^{N_{c}} GPA_{c}^{2}}{\sum_{s=1}^{N_{c}} (r_{sc} \times GPA_{c})}$$  \hspace{1cm} (5)$$

### 2.2.2 Profile-Based Metrics

Simple Difficulty metrics (PR, AG, $\beta$ and $\alpha$) reduce the difficulty of a course to a single number. However, as demonstrated by Mendez et al. [7], course difficulty is different for different types of students. To account for this difference, this work proposes a set of profile-based difficulty metrics.

The basic idea behind profile-based metrics is to divide the population of students in different groups according to their performance (usually their GPA). For example, in a program with grades between 0 and 10 and a passing grade of 6, students could be grouped into the following schema: students with [GPA higher than 8.5], [GPA of 7.5 to 8.5], [GPA of 6.5 to 7.5], [GPA of 5.5 to 6.5] and [GPA lower than 5.5]. Then the relevant metric for a course is calculated separately for each group using only information from the performance of its members.

The use of profile for the difficulty metrics reduce the bias of the PR and AG as it is calculated only for similar students in different courses. Also, the profile-based metrics preserve the basic grade distribution shape for $\beta$ and $\alpha$.

The proposed profile-based difficulty metrics are:

- **Course Approval Profile (CAP):** This is the profile-based version of the Passing Rate (PR) metric. For each student group, the number of students on that profile that have passed the course in a given period is divided by the number of students in the group that have taken the course in the same period.

- **Course Performance Profile (CPP):** This is the profile-based version of the Average Grade (AG) metric. For each group of students that have taken the course, the AG is calculated.

- **Course Difficulty Profile (CDP):** This is the profile-based versions of the metrics proposed by Calulkins et al. It can be Additive (CDP-$\beta$) or Multiplicative (CDP-$\alpha$), depending on the difficulty metric used for each group.

The result of the profile-based difficulty metrics is a vector. This representation enables the use of more sophisticated data mining techniques to compare and group courses according to their difficulty.

All the difficulty metrics could also be calculated for each Course-Instructor pair to provide a better difficulty estimation given that the characteristics and grade stringency of each instructor could bias the metric result if averaged over all instructors.

### 3. CURRICULUM ANALYSIS

The main purpose of calculating a set of well understood metrics over the different courses of a program curriculum is to easily find answers through more complex analysis based on a combination of the metrics’ results. This section provides five illustrative examples of these analyses using only the temporal and difficulty metrics presented before.

#### 3.1 Course Concurrency

One of the main tasks in curriculum analysis is to determine the workload that a student will receive over a given academic period. It is a usual practice that instructors from concurrent courses, that is, courses that are taken together in a period, interchange information about their course load (homework, projects, etc.) to avoid overloading the students over specific times during the period, for example near the exams. However, it is not always easy to determine which courses are actually concurrent, specially if the program is flexible.

This analysis can be performed mainly in two ways. Without previously calculated metrics, the recommended technique is to use a frequent itemset mining technique, such as FP-Growth [4]. This technique discovers courses commonly taken together more than a given percentage of times (support). However, it is not easy for instructors to determine the right value of the support and the crisp sets that this algorithm return hide information about less frequent but also occurring course concurrences.

In the second method, the determination of concurrency between courses can be easily obtained from the Course Temporal Position (CTP) metric. For example, in a semester-based program, all courses with at CTP between 1 and 1.5 could be considered to be part of the first semester, while all the courses with a CTP between 1.5 and 2.5 could be considered to be in the second semester. Moreover, overlapping sets could be used to assure that less frequent, but also relevant concurrences are taken into account in the period workload discussions.

#### 3.2 Neglected Courses

It is common to find curricula with small sequences of related courses. When those sequences are designed, it is expected that students follow the courses one after another in consecutive periods. This is specially important for difficult courses such as Calculus or Physics where concepts learned in a previous course are necessary to master the concepts of the next one. However, students, specially in flexible programs, could neglect taking some courses due to different factors (difficulty, personal preferences, reduced available time, etc.) If too much time pass between courses, some of the previously learned concepts could be forgotten by the time the next course requires them, generating lower than expected performance.

To find if there are courses that are consistently neglected by students, the Temporal Distance between Courses (TDI) can be used. TDI is applied to each pair of consecutive courses in the analyzed sequence. If a pair of expected consecutive courses have a TDI value higher than a threshold (for example 1.5) the second course could be considered neglected and actions should be taken to encourage the
students to take them as originally planned (for example, adding the second course as a prerequisite to a course with TDI between 2 and 2.5 from the first course).

3.3 Bottlenecks Identification

Due to economic constraints, the time that a student takes in completing the program has been of great interest for academic institutions. However, it is not always clear which courses are the bottlenecks that reduce the overall throughput of the program.

One way to identify the offending courses is to convert the curriculum into a graph. Each course will be a node in this graph. An edge will connect each pair of courses. The weight of each edge will be equal to the TDI between the courses it connects. All the edges with weights lower than 1 and higher than 2 are removed to leave only courses taken in sequence. Then the course with lower CTP is selected as the initial node and the critical path is found in the graph. The critical path determines the longest sequential path from the initial course. For each of the nodes in the critical path, the course duration (CDU) is calculated. Those courses in the critical path with the higher CDU could be flagged as bottlenecks because they are likely to increase the number of periods that a student has to stay in the program.

3.4 Section Planning

Physical or regulatory limitation often determine the maximum numbers of students in a given class. When there are more students than places in a class, it is common practice to create additional sections of the course taught either by the same or a different instructor. Planning the number of sections needed for the next period, before the end of the current period is sometimes a challenge and usually provide unreliable results. This lead to wasting of resources (for example, two half-full sections) or under-served students (for example, students that can not follow the course during the period due to full sections).

The average passing rate it the usual way in which the forecast about the number of students that will be available to take the next courses is calculated. However, given that each period the composition of students varies, the passing rate does not remain constant, leading to inaccurate results. The use of the profile-base approval metric (CAP) could provide a better way to forecast the actual number of students that will pass the course because it takes into account the different performance of the students taking the course. Those CAP could be refined by using a combination of Course-Instructor to also take into account the grading stringency of the instructor.

3.5 Course Similarity

One of the main curricular decisions that students make is the selection of the course load for each period. The number and difficulty of the courses has been found to have direct impact on the performance of the students [7]. This decision is so important that it is common for academic institutions to provide course-selection counseling for students that seems to be struggling with their workload. The counseling session, however, only transfer the burden of course selection to instructors or professors that do not necessarily have a current picture of the difficulty and load of all the courses in the program. The decision is taken with a better background knowledge, but still perceptions and beliefs are

the main sources of information.

The vector nature of the profile-based difficulty metrics could be exploited to apply straight-forward clustering techniques to group the courses according to their type of difficulty. These groupings could provide an easier way to characterize courses. For example, courses with the same passing rate AG, could be grouped separately according to their Difficulty profile (CDP). Difficult courses, with a linearly decreasing negative β for students with lower GPAs, will be clustered together. The same will happen to easy courses that have a constant β value among the groups. Courses with other distributions (for example, very easy for good performers, but hard for bad performers) will also be clustered with similar courses. Presenting this information for all courses in the program could help instructors to associate the difficulty of known courses to new or unknown courses. This potentially could lead to a better recommendation to the student.

4. CONCLUSIONS AND FURTHER WORK

Differently from data produced at course-level, program-level data tend to be more homogenous between programs and institutions. This similarity could lead to the development of a sub-field of Learning Analytics with a common set of metrics and methodologies for Program Curriculum analysis that could be called Curricular Analytics. This work is one of the first steps towards the creating this sub-field.

Even simple metrics, when well defined and transferable between programs, have the capacity to improve the way in which curricula is analyzed and improved. The list of metrics presented in this work is by no means comprehensive, but provide a starting point from which more advanced and informative metrics could be created.

The presented illustrative analysis served as an initial validation of the feasibility and usefulness of the metrics. However, a series of evaluation studies with real data from existing programs is needed before these metrics could be safely used by practitioners to draw conclusions from their programs. The operational complexity of these studies is very low given that only the raw data and simple computational tools (for example an spreadsheet) are needed to obtain the metrics. On the other hand, measuring the informational value of the metrics to solve real-world questions requires a more complex quantitative and qualitative analysis.

The relative homogeneity of the data could also lead to the creation of Curricular Analytics tools or plugins that could incorporate all the tested metrics and analysis developed inside this sub-field. The existence of this kind of easy-to-use tools could help transferring the research results into the practitioners field much faster than what has happened in Learning Analytics in general, where research results are much harder to make inroad in the day-to-day operation of academic institutions.

Finally, this work is a call to other Learning Analytics researchers to start focusing on the different levels of learning and education and the interrelation between those levels. While the focus on course-level analytics could help to improve the learning process in the classroom, only a holistic approach could ensure that these improvements are also reflected in the efficiency and effectiveness of learning programs and that society will receive the benefit of better prepared individuals.
5. REFERENCES


Assessment analytics for peer-assessment: a model and implementation

Blaženka Divjak
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2
42000 Varaždin, Croatia
bdivjak@foi.hr

Darko Grabar
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2
42000 Varaždin, Croatia
darko.grabar@foi.hr

Marcel Maretić
Faculty of Organization and Informatics
University of Zagreb
Pavlinska 2
42000 Varaždin, Croatia
mmaretic@foi.hr

ABSTRACT
Learning analytics should go beyond data analysis and include approaches and algorithms that are meaningful for learner performance and that can be interpreted by teacher and related to learning outcomes. Assessment analytics has been lagging behind other research in learning analytics. This holds true especially for peer-assessment analytics.

In this paper we present a mathematical model for peer-assessment based on the use of scoring rubrics for criteria-based assessment. We propose methods for the calculation of the final grade along with reliability measures of peer-assessment. Modeling is motivated and driven by the identified peer-assessment scenarios.

Use of peer-assessment based on a sound model provides benefits of the deeper learning while addressing the issues of validity and reliability.

Categories and Subject Descriptors
D.2.8 [Software Engineering]: Metrics—complexity measures, performance measures; H.4 [Information Systems Applications]: Miscellaneous

General Terms
Algorithms, Design, Measurement, Reliability

Keywords
peer-assessment, assessment, analytic tools for learners, assessment learning analytics

1. BACKGROUND ON ASSESSMENT LEARNING ANALYTICS
Learning analytics (LA) is all about usefulness of the data once they have been collected and analyzed [6]. Research in LA is interdisciplinary and it must be emphasized that LA includes the aspects of human judgments and it goes beyond data analysis (business analytics): it has to make sense of information, come to decisions and take action based on data [13]. This is the leitmotiv of the research presented in this paper.

LA has to be useful to a vast majority of students. The so-called average student has to be taken into account when setting the goals of LA, not only the under-performing or over-performing students. Teaching practice shows that a meaningful analysis of assessment results is of interest to all the students.

Assessment is both ubiquitous and very meaningful as far as students and teachers are concerned (Ellis in [6]). It is an essential part of the teaching and learning process, especially in the formal education because assessment guides learning for a vast majority of students. Ellis at the same time claims that assessment analytics are lagging behind other types of learning analytics. There are several reasons for this. Among these, we argue that insufficient granularity of assessment data presents a difficulty for an interpretation of results.

The so called networked learning (see [12], e.g. Massive Open Online Courses (MOOCs), social learning platforms, online learning and e-learning in general) presents a completely new playground for learning analytics. In networked learning the number of participants rapidly increases along with the interactions between learners in the form of discussions and mutual learning. We focus here on a special types of assessment: peer-assessment. Use of peer-assessment and self-assessment is appealing and very appropriate for a task leading to a certificate in a MOOC with enrollment measured in tens of thousands. This approach generates a huge amount of assessment data but also asks for sound metrics for the calculation of final grade and for estimates on the reliability of assessment data. Peer-assessment has additional benefits in the learning process, but also additional disadvantages (cf. [4]). Among the disadvantages there are issues of reliability and validity of assessment.

To address validity, we advise the use of the scoring rubrics as they contribute to the quality of assessments and by facilitating valid judgments of complex competencies [10]. Based on the analysis of 75 studies Jonsson and Svigby
in [10] conclude that the use of scoring rubrics enhances the reliability of assessments, especially if the rubrics are analytic, topic-specific, and complemented with examples and/or rater training. Otherwise, the scoring rubrics do not facilitate valid judgment of performance assessments. Besides this, rubrics have a potential to promote learning and/or improve instruction.

Aim of this paper is to model peer-assessment and to discuss issues of final grade calculation and reliability of raters’ judgments. Jonsson and Svingby note that variations in raters’ judgments can occur either across raters, known as inter-rater reliability, or in the consistency of one single rater, called intra-rater reliability. Referring to [1] Jonsson and Svingby state that a major threat to reliability is the lack of consistency of an individual grader. Reports rarely mention this measure. On the other hand, inter-rater reliability is in some form mentioned in more than half of the reports but many of these simply use percentage as a measure for agreement. This is in agreement with Sadler and Good’s critique in [14] of poor quality of quantitative research regarding self-assessment. Situation has improved since. Nevertheless, majority of current research still uses overly simple statistical measures in order to determine correlations that might indicate reliability.

In the following sections we describe two major peer-assessment scenarios we have recognized and for which we have developed a mathematical model. After that we present and analyze a model for these scenarios.

2. SCENARIOS FOR PEER-ASSESSMENT

Reliability of peer-assessment depends on many factors but consistency of individual evaluator was very early recognized as the most important (see [1]). On the other hand, having more assessments per assignment increases the reliability of peer-assessment with relatively inexperienced evaluators.

From experienced evaluators (experts) we presume a high expertise in the domain knowledge and prior experience in evaluation. Similarly, an inexperienced evaluator is an individual with a relatively high level of domain knowledge (high baseline), but lacking experience in evaluation (e.g. peer assessment by senior undergraduates).

We analyze scenarios with respect to the experience of evaluator as is shown in scenario grid (Fig. 1). We have placed a continuum of possible scenarios in a grid with four quadrants. Within four quadrants we recognize two interesting scenarios for peer-assessment and discard the other two as either unrealistic or inappropriate.

In the first scenario, let us call it Scenario A, participants are inexperienced evaluators (for example undergraduate students with introductory domain knowledge and no experience in peer-assessment) whereas in the scenario B evaluators have higher expertise in the evaluated domain (i.e. teachers, graduate students or senior undergraduates) and prior training in assessment. In scenario A, the lack of experience in evaluation must be compensated with a quantity of peer-assessments, i.e. having a larger group size in peer-assessment. On the other hand, setting a group size too large in scenario B is a needless waste of expert’s time.

![Figure 1: Scenario grid](image)

Detailed analysis is given in the Table 1.

3. OVERVIEW OF THE PEER-ASSESSMENT ACTIVITY

Peer-assessment activity starts after the work on the assignment task has completed. In a general case peer-assessment consists of two phases. We identify following activities in the whole process.

**Phase 1: Assessment of assignments**

i. Learners assess a (predefined) number of assigned assignments

ii. Analysis of peer-assessments (grouped by assignment)

iii. Calculation of the assignment grade

**Phase 2: Assessment of the assessments**

i. Analysis of peer-assessments (grouped by grader)

ii. Calculation of the assessment grade

First phase starts with learners assessing the assignment work of their peers. We assume that each participant grades several assignments (at least 2). At the end of the first phase a reliability check has to be performed and the final grade has to be calculated. Second phase is concerned with the quality of assessments relative to the evaluator. As on outcome of the second phase graders can receive a grade (points) for the quality of their assessments.
## 4. Mathematical Model for Peer-Assessment

We recognized three challenges: (1) calculation of the final grade based on different assessment scenarios, (2) measurement of the assessment’s reliability and (3) measurement of reliability of each grader (for grading of the graders).

### 4.1 Overview of the assignment grading

A grading $G$ from the scoring rubric with $n$ criteria is a tuple of numbers $G = (g_1, \ldots, g_n)$. We consider gradings as points in an $n$-dimensional space endowed with a metric $d$, i.e. a function that measures the distance between points (i.e. gradings) and satisfies the axioms of a metric space.

In [5] we proposed the use of the non-euclidean taxicab metric $d_1$, but for the purpose of this paper it is sufficient think of $d$ as any distance metric.

### 4.2 Calculation of the assignment’s final grade

An assignment graded through peer-assessment will receive several peer gradings. These will have to be analyzed. If estimated as reliable these gradings will be use as input for the calculation of the final grade.

A simplest approach is to calculate the final grade based on the mean value of received assessments.

Let $\mathcal{S} = \{S_1^1, \ldots, S_n^m\}$ denote a set of peer gradings for assignment $k$, then the mean grade is

$$M(\mathcal{S}) = (a_1^k, \ldots, a_n^k), \quad \text{where} \quad a_i^k = \frac{1}{m} \left( \sum_{j=1}^{m} c_{i,j}^k \right).$$

$M(\mathcal{S})$ is a center of mass of the set $\mathcal{S}$. This method for grade calculation is suitable for scenario A. We can say that $M(\mathcal{S})$ is sensitive to quantity, and less sensitive to outliers (i.e. a function that measures the distance between points (i.e. gradings) and satisfies the axioms of a metric space).

For scenario B, we propose an alternative grade calculation method (see [5]). In scenario B we assume that peers are experienced evaluators. Final grade is calculated as so-called optimal final grade $O(\mathcal{S})$ defined by

$$O(\mathcal{S}) = \left( a_1^o, \ldots, a_n^o \right), \quad \text{where} \quad a_i^o = \frac{1}{2} \left( W(\mathcal{S}) + B(\mathcal{S}) \right).$$

$W(\mathcal{S})$ and $B(\mathcal{S})$ represent amalgamations of worst and best received gradings respectively, defined by:

$$W(\mathcal{S}) = (w_1, \ldots, w_n), \quad w_i = \min_j c_{i,j}^k,$$

$$B(\mathcal{S}) = (b_1, \ldots, b_n), \quad b_i = \max_j c_{i,j}^k.$$

This approach is inspired by Hwang and Yoon’s TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method of multi-criteria decision making in [9]. When evaluators are trusted experts, we don’t expect “wild” gradings (outliers). Here, it is expected that after just a few initial evaluations any additional gradings will have no effect on the final grade $O(\mathcal{S})$. Please consult [5] for additional details.

A summary of our recommendations for two scenarios A and B is given in the Table 2.

## 4.3 Reliability of the peer-assessment

A prerequisite for the calculation of the assignment’s final grade is the determination whether a received set of peer-

### Table 1: Scenario table

<table>
<thead>
<tr>
<th>Scenario A</th>
<th>Scenario B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Playground – use cases</td>
<td>Multiple graduate/postgraduate assess complex student work [9]</td>
</tr>
<tr>
<td>Networked learning (MOOCs, online learning and e-learning in general, see [12])</td>
<td>Peer assessment of research papers</td>
</tr>
<tr>
<td>Voting for awards where general audience is involved</td>
<td>Evaluation of competitive research projects</td>
</tr>
</tbody>
</table>

### Table 2: Grading method recommendations

<table>
<thead>
<tr>
<th>Scenario A</th>
<th>Scenario B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suggested grading method</td>
<td>Mean value grading.</td>
</tr>
<tr>
<td>Reliability provided by quantity of evaluations.</td>
<td>Optimal value grading.</td>
</tr>
<tr>
<td>Reliability provided by the quality of evaluators.</td>
<td></td>
</tr>
</tbody>
</table>

With optimal value grading we have the opportunity to allow experts to skip grading for certain criteria. For example this would be reasonable if an expert is not an expert for all the criteria. To be able to calculate $O(\mathcal{S})$ it is sufficient to have every criteria covered by at least one expert.
assessments is (sufficiently) reliable, i.e. acceptable.

For reasoning about reliability it is necessary to have granular data. The importance of granular scoring data is illustrated in the example in Table 3. Gradings $S_1$ and $S_2$ agree on the summative level, but seem very distinct at the granular level. This is an example of an unreliable peer-grading set where this incoherence is not visible on the summative level.

Table 3: Highlighting the importance of granular data
\[
\begin{array}{cccc}
C_1 & C_2 & C_3 & C_4 & \sum \\
S_1 & 3 & 0 & 2 & 2 & 7 \\
S_2 & 0 & 1 & 3 & 3 & 7 \\
\end{array}
\]

A \textit{diameter} of a set of gradings $S = \{S_1, \ldots, S_n\}$ is defined as
\[
\text{diam} S = \max_{i,j} d(S_i, S_j).
\]

We consider a set $S$ of peer gradings as \textit{reliable} if $\text{diam} S$ (maximal pairwise distance between gradings) is less than $2e$ where $e$ is \textit{acceptable error} given in advance.

Note that the diameter of the set $S$ is also a diameter of an encompassing sphere. So, we can say that a reliable peer-grading set fits within an encompassing $e$-sphere.

If a set of peer-assessments is estimated as not acceptable (un-reliable) on the granulated level then the final grade cannot be calculated. A recommendation about acceptability of particular peer-assessment set can be given to teacher or course designer by LA. This can be implemented in the learning management system (LMS, for example Moodle). Practical related issues will be discussed in the section 5.

### 4.4 Grading process

Assessment set can turn out as unacceptable because of a single outlier grading. As an attempt to eliminate the outlier grading we propose to search for a maximal acceptable subset of the received peer-assessments. If such subset can be found, it is then used as input for the final grade calculation.

As a measure of final resort, an supervisor’s intervention is asked for. In a course with a large student enrollment (thousands for a MOOC) this will be avoided as much as possible. However, if present, instructor’s assessment becomes a final grade (no need for calculation). This is described in Algorithm 1.

### 4.5 Normalization

Metric $d$ can be linearly scaled to obtain a normalized metric $d_0$ with values within the interval $[0, 1]$. Distance of $d_0 = 1$ corresponds to the maximal distance between worst and best possible gradings.

This would facilitate having general recommendations for setting acceptable error $e$ on a normalized scale (setting $e_0 = 0.2$ for example). Additionally, this could facilitate comparison of data from different tasks (within a course, or from different courses).

### 4.6 Evaluation of peer-assessments (awarding the graders)

Goal of the second phase of the peer-assessment process is to reward the graders for their effort. Graders (peers) who have graded consistently and accurately (near the final grade) should be rewarded more than inconsistent and inaccurate graders.

Let us assume that a maximum of $A$ points is awarded for the peer-assessment task. Then grader $k$ can be awarded $A_i$ points for each of the $m$ gradings $G_i$ that he/she was assigned, where $A_i$ is calculated by the following formula

\[
A_i(d_i) := \begin{cases} 
\frac{A}{me} (e - d_i), & d_i < e \\
0, & d_i \geq e 
\end{cases}
\]

where $d_i = d(G_i, F)$.

This has the effect that 0 points are awarded for a grading outside of the $e$-sphere around final grade $F$. For a grading within this $e$-sphere $A_i$ is proportional to $(e - d_i)$ where $d_i = d(G_i, F)$.

Finally, grader $k$ is awarded a total of $A(k)$ points for his effort with gradings $G_1, \ldots, G_m$ where $A(k)$ is calculated as a sum of $A_i(d_i)$.

![Figure 2: Points awarded to grader for grading $G_i$](image-url)
5. IMPLEMENTATION

A support for peer-assessment LA is lacking in assessment analytics in general. We analyze the current implementation in the Moodle LMS where peer-assessment activity is implemented with the Workshop plugin.

In a Workshop activity, students receive a grade for their work and another grade for the quality of their assessment of other student’s assignments.

Each participant in Workshop gets a grade for his submission and a grade for her assessments. These grades are visible as separate grade items in student’s gradebook.

Current implementation of Workshop calculates the assignment grade as a weighted mean of received assessment grades. Received gradings are not analyzed for reliability. If the teacher wishes to override or influence the calculated assignment grade, he can (a) additional provide his own assessment and set its corresponding weight to a higher value or (b) even completely override the final grade. As we have argued here and in [5] we find this method as inadequate. Therefore, we proposed alternative methods for the calculation of the final grade.

Assessment grade calculation is more complex. The goal is to estimate the quality of each assessment. One assessment is singled out as the best one – it is the assessment closest to the mean value of all assessments. This selected assessment is assigned with highest grade. Other assessments receive grades based on the distance from the selected assessment. Teacher can influence in this process by setting the parameter which determines how quickly a grade should decreased relative to the distance.

We are currently developing a new Moodle plugin for peer-assessment. This plugin will address the identified problems of the current implementation according to our model.

6. CONCLUSION. FURTHER RESEARCH

Peer-assessment has many advantages for students (for example development of metacognitive skills) and for teachers (for example saves teacher’s time) but there are several challenges related to their implementation such as calculation of final grade, reliability check and awarding an evaluator for peer-assessment.

In this paper we propose new methods for calculation of the grades in peer-assessment. We propose a measure for reliability and a method for grading peer-evaluations in a peer-assessment exercise. These metrics are based on two distinguished scenario analysis that takes into account a number of possible evaluators and evaluator expertise (domain knowledge and evaluation skills). We pursue an approach to model assessment LA analytics with a geometric model.

In [4] we analyzed a case study based on the master level Project Management course at the University of Zagreb. Our analysis has confirmed the need for deeper analysis of reliability in peer-assessment. Further exploring of data related to the peer-assessment learning analytics in MOOCs is expected. Having additional data should result in improvement of the model and recommendations on the applicability of scenarios, parameters and analysis of the acceptable error of the assessment set.

Also, we intend to implement our model (algorithms and the supporting recommendation system) as a peer-assessment plug-in for the Moodle LMS.

Finally, we conclude that a well founded mathematical modeling, based on not just descriptive statistics, should be used more often in learning analytics.

7. REFERENCES


**ABSTRACT**

In this paper, we present some of our recent experiences with a data visualization tool and offer some use cases where the visualization tool can potentially drive programmatic change in universities. The Ribbon Tool provides an interactive visualization of student flows through academic programs, progressing over time to either successful completion (graduation) or attrition. Through effective use of the Ribbon Tool by those who can effect curriculum change, their ability to generate persuasive arguments for change are enhanced. This paper presents some use cases and commentary on actual usage of the Ribbon Tool to call for programmatic change across a university.

**Keywords**

Learning Analytics, Visualization, Programmatic Change, Ribbon Tool

1. INTRODUCTION

Academics pride themselves on evidence-informed decision making, but when it comes to making changes in their teaching practices, curricula, or academic programs, data and evidence seem to hold little sway. Perhaps this stems from the belief that as an expert in a subject area, one is automatically an expert in how, where, and what of the subject area should be taught. Perhaps this stems from the outdated “mini-me” assumption that students are either faculty in training or destined for attrition. Or, or perhaps it stems from the discipline-based belief that teaching practices, curricula, and academic programs were carved in stone tablets by the ancestors and never meant to change.

Instigating change in university programs is difficult, in part because it is easy to throw sand in the wheels of change, but also in part because the agents of change and the influencers are rarely the same people. Sadly, evidence-informed arguments to justify changes in teaching or curriculum often have no more persuasive effect, or perhaps even less effect, than anecdotal stories about “in my day”, or “my son or daughter experienced”. While skepticism can be healthy when evaluating evidence gathered from others’ observations and statistical analysis, it can also be used to stonewall or stymie change.

Confronting academics and administrators with cold facts, such as “One third of your students from certain diversity groups are leaving your program within the first two years” or “One quarter of your students are failing their required first mathematics course” are met with retorts like “Tell me something I haven’t heard before!” or “Bring me some evidence that is actionable!”.

Studies of decision-making indicate that sometimes decisions are made very quickly based on instinct, ignoring the actual deeper problem underneath [1].

With learning analytics and data visualization tools it is now easier to put into the hands of academics more powerful interactive tools to dig into data, to discover for themselves the facts and relationships that matter to them, to experiment with models that can answer some of their questions, and to develop persuasive arguments that can support the case for change. We have found that interactive data visualizations can support academic leaders in initiating data-driven and evidence-informed change.

2. DATA-DRIVEN VISUALIZATIONS WITH THE RIBBON TOOL

A data visualization tool called the “Ribbon Tool” has been developed at UC Davis (http://t4eba.com/ribbon/) building upon the Sankey Diagram functionality with the Data-Driven Documents (D3) data visualization library [2]. This tool has been utilized for visualizing student flows through academic programs in universities, with groups of students represented as coloured ribbons as they move from admission to graduation or attrition. An example of a Ribbon screenshot is shown in Figure 1.

Vertical bars within the tool indicate the status of students in a particular year and term of an academic program. The ribbons that flow from bar to bar correspond to the number of students moving from state to state. For example, in Figure 2, the three bars indicate September snapshots in 2011, 2014 and 2015. The red ribbons show numbers of students who began in Engineering in September 2011 and continued tracking them as they move forward in time.

In the Ribbon Tool, a “mouse-over” in the diagram will reveal a text box showing the number of students in a particular ribbon. The textboxes in Figure 2 show that of the 351 students who began in Engineering in the fall of 2011, some 240 were still enrolled in the fall of 2014. Another 33 students had transferred in from Arts and Science. Some students had transferred out of Engineering, to Arts and Science or another faculty. Some had dropped out of the University, and a few were on a “stop-out”. By fall 2015 (after 4 years), one can see that 88 students had graduated with an Engineering degree. A few others had degrees in other faculties and 177 were still enrolled for their 5th year. Note that a substantial number of Engineering students complete a
one-year paid internship, which naturally extends the degree to a minimum 5-year duration.

The vertical bars represent a hierarchy of temporal information. In the above example, the top level of the hierarchy represents whether students were enrolled, had been granted a degree, or had left the institution. In the next level, we show the college or school in which they had been enrolled (or had granted them a degree or from which they left or stopped-out). If one were to drill down to a third level, the data shows the department (Electrical, Mechanical, Civil, etc.) where the student is enrolled or awarded a degree. Expanding or collapsing the hierarchy gives a more or less refined view. The interactive visualization allows the user to isolate a particular group in the hierarchy (for example students who were enrolled in Mechanical Engineering in 2014) and project backward to see where they came from and forward to see where they went next. Moving through the hierarchy and isolating views allows the user to focus in on areas of potential interest.

Along with the visualization, the user is provided a set of filters. For example, if one were interested in examining gender differences in student flows through Engineering, one could filter to obtain separate diagrams for female and male students. These can be quickly visually compared to see if proportions of degrees granted, attrition, time to graduation, departmental breakdowns are impacted by gender. Other filters based on any set of categorical demographic or academic characteristics can be added. For example, the program flow-through for female students entering Engineering directly from high school with SAT scores in the top decile can be examined with a few mouse clicks.

This flexible and powerful visualization tool has been used extensively at UC Davis and is now being disseminated to other universities through the “Tools for Evidence-Based Action (TEA) Community” [3], funded in part by the Helmsley Foundation. The Ribbon Tool has been greeted with great enthusiasm by deans and other administrators at our University as a tool to augment their other data analysis efforts and as a means to explore elements of their academic programs.

Figure 1: Screenshot of the UCDavis Ribbon Tool
3. POPULATING THE RIBBON TOOL WITH DATA

The Ribbon Tool requires two chunks of data, a set of filter values and a data hierarchy. There can be an arbitrary number of filter variables, each with a label and a set of nominal values. The data hierarchy can have an arbitrary depth and at each level of the hierarchy a value must exist for each student. The branching factor at each level of the hierarchy must be fixed in terms of its subcategory options. Each student represented in the visualization must have a full set of values corresponding to the filter variables. Further each student must have a value for each level in the hierarchy. Data can be imported into the Ribbon Tool from either a pair of csv files or from a JSON file.

In the datasets we have prepared for our institution, students are not individually identified, other than by a sequential index. As a result, the data held in Ribbon, although hosted in the Amazon Cloud, has low risk of abuse. Nevertheless, efforts are underway to offer a local data storage option for some universities hesitant to store even de-identified student data off-site.

4. SOME USE CASES AND EXPERIENCES

We have been using the Ribbon Tool at the University of Saskatchewan for only a few months now. During that time the tool has been further enhanced in its capabilities and features and improving in its reliability and robustness, thanks to the development team at UC Davis. Below are some use cases that have proven useful in our experience to date.

4.1 Examining Degree Completion and Time to Graduation

Timely degree completion is a key component of enrolment management. For example, university funding is often associated with 6-year completion rates in undergraduate programs. Students stuck in a program for an extended time can reduce the number of available spaces in critical courses, and can face compounded delays due to rigid, prerequisite-bound course sequences.

Using the Ribbon Tool it is easy to see degree completion times and to determine the number of students in a cohort who are completing degrees within 6 years or who are embarking on a 7th or 8th year. Furthermore, it is possible, using filters to see if the students failing to complete within 6 years have had stop outs, academic probation actions, internships, etc. It is possible to differentiate completion rates for students with different demographic factors, such as first-in-family (first-generation) students, international students, under-represented minority students, etc. It is possible to display student GPAs within the hierarchy to determine if students slow to graduate have or have not reached certain academic achievement levels.
The combination of filters and hierarchy refinements has permitted our Engineering School to discover some new insights and bottlenecks regarding time to graduation.

4.2 Retention and Attrition

Analyzing student attrition and retention factors is an interest in some parts of every university. Universities focused on broad access in Arts and Sciences are often faced with retention challenges. Students unprepared for the change in culture of university life and those with academic deficiencies are not the only students who sometimes leave the institution. Established retention risk factors such as lower socio-economic status, being a first-generation student, being a member of an under-represented minority all need to be considered. But when comparing different academic programs, such as Engineering and the Humanities, there may be different factors leading to attrition. For example, belief in the benefit of completing a university degree may be a factor in some areas whereas the rigor of completing the degree may be a factor in others [4].

We have used the Ribbon Tool to track attrition and to differentiate students moving to a different program versus drop-outs versus drop-outs. Furthermore, in areas where there are various entry points into programs, it is possible to examine retention factors for students who have entered through different paths. In doing a retention analysis with Ribbon, demographic filter variables corresponding to expected causes of retention can be quickly examined to see which factors or combinations of factors seem to make a difference. Being able to isolate a particular collection of students (e.g. those who drop out after sophomore year in a program) to further investigate their demographic makeup and their pathways has proven useful. Ribbon can also be used to determine whether the flow of students through academic programs has been affected by changes in demographics of entering students, whether as a result of changes to the feeder system or changes in admission policies.

The Ribbon Tool has enabled our Engineering School as well as our Faculty of Arts and Sciences to study retention issues (in STEM and elsewhere) more closely and to get a better understanding of attrition patterns, particularly of under-represented minority students.

4.3 Program Innovation, Monitoring and Evaluation

As academic programs evolve and as new learner supports are introduced there is a need for ongoing program monitoring and evaluation. The Ribbon Tool provides a mechanism for supporting the early phase of program evaluation through its rapid means of detecting differences across cohorts of students. For example, it is easy to compare student flows before and after the implementation of some program change. It is also possible to differentiate with a filter those students who were selected for participation in a pilot program and further to filter those who did or did not engage.

We have begun to explore the impact of changes to our academic advising processes, the introduction of a freshman learning communities program, and the impact of increased academic support services in mathematics and writing. In such programs, where the macro effects may take many years to be realized, where effects may be differential across the different student demographics, and where levels of participation and engagement are vital indicators, the Ribbon Tool is helping us to us develop and refine program-impact hypotheses that can then be tested statistically.

5. ACTIONABLE DECISIONS

Of course, all of these kinds of comparisons and descriptions presented in the use-cases above can be achieved with a comprehensive set of reports, bar charts or tables or with the facile use of a statistics package. The difference with the Ribbon Tool is the speed with which one can mock up a scenario and try different filters and breakdowns to get an impression of where problems may be lurking or where impact may be seen. Furthermore, with the Ribbon Tool, an Associate Dean or Department Chair can take the reins and drive the visualization tool to explore exactly what is interesting - to follow a hunch or to confirm or deny a commonly held view. Putting a powerful visualization tool in the hands of agents of change can empower them to make more persuasive cases for change with their colleagues. We have seen how visualizations that show scenarios with no perceptible difference, when conventional wisdom would predict a difference, does help people to confront and question their biases. These are precisely the kinds of evidence that can change minds, and actionable decisions arise from changed minds.

6. CONCLUSION

Our experiences with the Ribbon Tool confirm that visualizations of student progression can be highly informative and powerfully persuasive in moving administrative staff to action. Uncovering the factors affiliated with undesired outcomes and discovering those connected with positive outcomes sets the stage for change.

The Ribbon Tool is one tool that can help with moving people to action, but like any tool it has its limitations. It is best suited for analyzing historical patterns and flows and is not well suited for forecasting or modeling the future effects of potential changes. It is also a tool that readily looks over relatively longer time scales we have not yet produced data to explore a more granular time scale. Finally, like any other tool, it can be mis-used to oversimplify relationships or to mis-represent realities. Just as with any power tool, much persuasive power is placed in the hands of the tool operator.

REFERENCES

Promoting Instructor and Department Action via Simple, Actionable Tools and Analyses

Marco Molinaro
University of California, Davis
mmolinaro@ucdavis.edu

Matthew Steinwachs
University of California, Davis
mksteinwachs@ucdavis.edu

Qiwei Li
University of California, Davis
qwli@ucdavis.edu

Alberto Guzman-Alvarez
University of California, Davis
aguzmanalvarez@ucdavis.edu

ABSTRACT
In this paper, we present some of our ongoing, as well as more recent work designing, implementing, and improving three tools to help university instructors and department leaders make evidence-based improvements to instruction. The first tool, Know Your Students, is a very early prototype that helps instructors tailor their instruction based on characteristics of the students they would not otherwise be aware of in their courses. The other two tools, the Departmental Diagnostic Dashboard and Ribbon Tool, help department chairs, curricular chairs, and/or advisors identify and make sense of student patterns they may be trying to minimize or enhance within individual courses, course series, and/or throughout their entire program. This paper illustrates examples of these tools and some of the actions they have inspired as a means of improving student outcomes.

Keywords
Learning Analytics, Visualization, Programmatic Change, Ribbon Tool, Departmental Instructional Dashboard

1. INTRODUCTION
University faculty members, staff and administrators often pride themselves on making decisions in a collaborative manner that takes multiple viewpoints into consideration. They view their decision-making as highly informed by evidence though the evidence is often more opinion-based than quantitative in nature. When data is brought into the equation it is often insufficient, not timely, nor tailored to questions that have the potential to impact student outcomes. Class size, student grades, and time to degree sliced and diced by gender, ethnicity and class standing are often the primary measures and offer outcome information after the term. Student evaluations of general course satisfaction are also universally applied and tend to be the only “actionable” information provided to instructors, albeit at the end of the course. There is a fundamental need to better understand our students, and the patterns that exist within our instructional system at the course, department and institution levels, to bring about effective instructional improvement.

Often, when instructors, curricular chairs and others involved in student instruction are asked how to improve student outcomes, a few common replies are heard: “we need to be more selective”, “students were better in the past”, “our students are not prepared”, “when I was a student I was expected to work so much harder”, “we need to fail more students and increase our rigor”, and the list goes on. There is a tendency to blame the students while reality suggests that it is not so simple. Students, faculty members, administrators and staff form a complex system that has to work together and respond to changing student demographics, societal needs, economic challenges and more.

A growing body of literature primarily from the STEM (Science, Technology, Engineering and Mathematics) educational arena [1], [2], [3] suggest that much is known about improving student instructional outcomes. Attempts to institute such approaches on a large scale are often met with “why should I change, I know what I do works”, “my classes are too large, only thing I can do is lecture”, or “my one class is not the determinant of student success”. While data and tools cannot solve all instructional issues, they can make people understand what is, and what’s not, happening within their course and department and connect student outcomes between instructional experiences.

With learning analytics and data visualization tools we can put more meaningful, actionable data in the hands of those responsible for ensuring a quality educational experience. We have found that faculty members, administrators and staff care greatly about students and have often not been able to act to improve student outcomes primarily because they lacked the needed information. We are embarking on providing useful, actionable, and timely information at multiple instructional levels through a variety of “dashboard-like” tools to promote positive change.

2. HELPING FACULTY KNOW THEIR STUDENTS
When instructors receive their class roster at many universities they are usually presented with student names, identification numbers and codes corresponding to students’ field of study (4 letter major codes at our institution). The assumption is often made that the students have met all preparatory requirements and are ready to effectively engage in the course content. A student’s failure to grasp course content adequately is usually assumed to be due to lack of time on task, lack of interest, lack of capability, and/or poor preparation. Very few instructors bother to take the time to check on one or more of these assumptions due to time pressure to “cover” all needed material, especially when teaching courses of 50 or more students over a 10 week course period (our standard “quarter”).

We posit that there is actionable data that can be shared with the instructor prior to course start that can lead to a better instructional experience and increased student learning. This data...
is gathered by our Center for Educational Effectiveness, a division of undergraduate education, and presented in aggregate fashion. This aggregation helps alleviate potential privacy concerns were the data to be presented is by individual student. Some of the information that can be considered, and sample actions that have been attempted, include:

**Basic demographics** This can contain aggregate information on gender, year in school, international status, primary language spoken, need for English remediation, first generation status, socioeconomic status, extent of testing accommodations for learning challenges and more. Such information can shape the form of writing assignments and grading rubrics, the types of examples brought to the classroom, the types and form of classroom activities or assignments chosen.

**Preparation** Background expertise brought into the course from prior course experiences at the university or at prior institutions is important. This not only includes grades received but also time gaps with material, course repetition, learning objective achievement (only available where measured as part of standard course practice – currently only available for one department’s introductory courses), pre-requisite completion, motivational survey data and more. Such information can greatly influence course content and emphasis as well as examples chosen.

**Motivation and load** Students’ needs and interest in a specific course is based on a variety of factors such as course of study being pursued, credit load in quarter, number of STEM courses currently enrolled in, course difficulty load and more. Awareness of student motivation can greatly influence course examples chosen, course workload and expectations.

These are the three areas that are currently being pilot tested with multiple first and second year courses that vary from 70 to 600 students. Data are currently aggregated manually into a multiple page report shared with the instructor. The information and analyses are being created using Tableau, SPSS and R, as most convenient. The envisioned final product would be an automatically generated analysis with suggested actions for student success and links to a Shiny dashboard providing more detailed information.

### 2.1 Prototype Example

One instructor currently prototyping Know Your Students information dashboard teaches a first course in organic chemistry for physical science majors. Based on the data she has discovered that a substantial number of students had not met pre-requisites or had performed very poorly on prior introductory chemistry courses, most of her students were first generation which usually is indicative of lack of knowledge of support structures within the university. Over 40% of her students had not had chemistry for over a year with some having had as long as a 2-year break. She also received information pointing out the level of mastery of her class on 27 learning objectives covered in the introductory chemistry year (3 courses; Chemistry 2A,B,C). Some of her information can be seen in Figure 1. Based on this information, she has altered various course sessions as well as expanded the range of information she would like to see from the product in the next iteration.

![Figure 1](image-url)
Welcome to PROTOTYPE Departmental Diagnostic Dashboard!
Here, you will find instructional information about your department

<table>
<thead>
<tr>
<th>Student Information</th>
<th>Course Information</th>
<th>Teaching Assistants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students</td>
<td>Graduate: Freshmen</td>
<td>Graduate: Transfers</td>
</tr>
</tbody>
</table>

Choose which course to visualize:
- LCS 120A
- LCS 120B
- CSCE 100

Choose splitting factors:
- ADMIT LEVEL: URM
- Exclude Summer Terms

Choose years to plot:
- 1996
- 1997
- 1998

See Visualization!

**Figure 2:** Screenshot of student information screen of the Departmental Diagnostic Dashboard for the computer science major separated by admission level (freshman vs. transfer) and under-represented minority status (URM). X-axis represents time in primary quarters (3 per year), for example 200703 represents spring 2007, 200901 is winter 2009 and 201010 is fall 2010. Y-axis is number of students.

**Figure 3:** Screenshot of the course information screen showing the courses most often taken by computer science majors that graduate in four years with course designations on left and year and term on bottom (1.F = first year Fall term, W = Winter, S = Spring, SS1 = Summer Session 1). Number in cells represents the average yearly number of students in a specific major receiving a D, F or W = withdrawal grade. Adding all numbers in a row gives an indication of the barrier posed by a specific course to students in the specific major. Other information can be selected in the left hand options and all university degree programs (majors) can be explored.

3. DEPARTMENTAL DIAGNOSTIC DASHBOARD

The Departmental Instructional Dashboard was initially conceived as a unified place where each of our 100 plus programs of study could be better understood by department administration and advising staff. The dashboard provides information on a quarterly basis rather than the traditional once every seven year program review cycle. With an ever-growing pressure to improve graduation rates, this prototype tool was created in Fall 2015 and now contains student, course, and graduation information for all undergraduate programs at our university since 2000. Once one or more programs are selected, the following types of information are currently available in the prototype:

**Student Information** This information shows numbers of enrolled students by quarter that can be separated by multiple different demographic variables. In Figure 2 you can view the growth of our computer science major since 2006 as well as gauge the numbers of under-represented minorities in the program and whether they entered as freshmen or as third year transfer students.

**Course Information** This area shows information related to course popularity at the various stages of a student’s timeline on our campus, and can point out courses that are particularly...
difficult or challenging for students in a particular course of study. In Figure 3 you can see which courses were taken and when they were particularly troublesome for the computer science graduates that started as freshmen and completed their degree within 4 years.

**Graduation Information** These tabs contain information about an initial cohort of students who enrolled in a given major and the numbers and percentages of those students that were able to graduate, graduate within four years, or almost graduate within four years. Additionally there is information about “forgetters” – those that graduated one term after 4 years but took no units in their last term, and “almosters” – those that finished one term past four years. This information is presented via simple bar chart outlining numbers of students and courses taken.

The dataset used by the dashboard is a cleaned dataset supplied by our registrar and enhanced with multiple additional fields and substantial calculations, analyses and visualizations completed in the R programming language and delivered via a Shiny dashboard interface. The tool was initiated as a summer project for one of our Master’s students in statistics and is now slated to be delivered via password-protected Shiny interface to all of our program administrators and lead advising staff.

### 4. SEEING THE BIG PICTURE WITH THE RIBBON TOOL

We have developed a data visualization tool called the “Ribbon Tool” (http://t4eba.com/ribbon/) building upon the Sankey Diagram functionality with the Data-Driven Documents (D3) data visualization library [4]. This tool is utilized for visualizing flows of many kinds, primarily student flows between academic programs within universities, with groups of students represented as colored ribbons as they move from admission to graduation or attrition (dismissal or departure). An example of a Ribbon diagram is shown in Figure 4 below.

Vertical bars within the tool indicate the status of students in a particular year and term of an academic program. The ribbons that flow from bar to bar correspond to the number of students moving from state to state. For example, in Figure 4, the engineering discipline, a set of multiple degrees within an academic program. The ribbons represent colored as they move from admission to graduation or attrition (dismissal or departure). An example of a Ribbon diagram is shown in Figure 4 below.

**Figure 4:** Screenshot of Ribbon Tool comparing engineering and math and physical science student paths after 1, 2, and 4 years at university. Column indicators show year and term – 200810 is Fall 2008.
5. ACTIONABLE DECISIONS

All of the ideas and visualizations presented in this paper can be created via specific requests to institutional research staff, the Registrar, Admissions and/or department analysts – the difference is that the request, turn around time, and likely need for multiple iterations can be mostly bypassed once useful datasets are obtained or created. The tools presented allow for a great deal of local exploration and generation of powerful visuals that can help communicate ideas to others in position to make positive changes. The dashboard approach taken in the Departmental Dashboard and Know Your Students suites of analyses allows for easy central updating, secure use, and quick iterative improvement. The common language across departments that is gained from using the same tools also facilitates discussions across departments and colleges allowing the potential for decisive actions to occur on a faster timescale. Additionally, the collection of tools fosters informed discussions and decisions at the scale of the individual instructor, the department, the school/college, all the way to the institution- or system-wide level.

6. CONCLUSION

Our extensive experience with the Ribbon Tool, growing use of the Departmental Diagnostic Dashboard and beginning work with the Know Your Students collection of data are setting the stage for widespread use of data to improve instructional outcomes for all students. As with all tools, there is still much room for improvement as well as ongoing opportunities for misinterpretation. Ribbon Tool and the Departmental Diagnostic Dashboard are best for looking for patterns and helping uncover areas for potential improvement while the Know Your Students tool can help break the patterns from the start and guide, hopefully, useful interventions. It is yet to be seen how many types of data and visualizations thereof can inform effective actions. Showing a trend does not clarify where and how action should be taken. Still, ongoing experimentation with these tools continues to uncover new ways for their use and inspires more thinking about what instructors, administrators and staff can do to create the most effective educational experiences for our students.

REFERENCES