Issues, Challenges, and Lessons Learned When Scaling up a Learning Analytics Intervention

Steven Lonn  
University of Michigan  
USE Lab, Digital Media Commons  
1401 B Duderstadt Ctr, 2281 Bonisteel  
Ann Arbor, MI 48109-2094 USA  
+1 (734) 615-4333  
slonn@umich.edu

Stephen Aguilar  
University of Michigan  
School of Education & USE Lab  
Suite 4215, 610 E University Ave  
Ann Arbor, MI 48109-1259 USA  
+1 (734) 764-8416  
aguilars@umich.edu

Stephanie D. Teasley  
University of Michigan  
School of Information & USE Lab  
4384 North Quad, 105 S. State St.  
Ann Arbor, MI 48109-1285 USA  
+1 (734) 763-8124  
steasley@umich.edu

ABSTRACT
This paper describes an intra-institutional partnership between a research team and a technology service group that was established to facilitate the scaling up of a learning analytics intervention. Our discussion focuses on the benefits and challenges that arose from this partnership in order to provide useful information for similar partnerships developed to support scaling up learning analytics interventions.

Categories and Subject Descriptors
J.1 [Administrative Data Processing] Education; K.3 [Computer Uses in Education] - General

General Terms
Management, Measurement, Performance, Design.

Keywords
Learning Analytics, Design-Research, Scale, Higher Education.

1 INTRODUCTION
Over the past few decades, most of the new learning technology innovations in higher education have been tested in relatively small, localized settings [1]. Given the personal and financial costs of transferring these technologies to new contexts, and providing training as well as hardware, it is (unfortunately) very difficult for these technologies to scale beyond the initial scope of the research project. Ideally, recent innovations in learning analytics should be able to overcome these challenges and make the leap from the focused and particular to the broad and general. At their core, these technologies leverage rich and massive data sets (i.e., "big data") that have the potential to generalize across disciplines and individual learners. Purdue University, for example, has scaled their Course Signals innovation to over 100 courses thus far, providing formative grade feedback to more than 23,000 students [2].

Our prior research mined Learning Management System (LMS) data to better understand the influence of students’ “in-system” behaviors on educational practices [3]. Building on this research, we developed an Early Warning System to support just-in-time decision-making around students’ academic performance for the academic advisors within two specific learning communities [4]. As we worked with academic advisors to identify what data they required to increase and inform the academic support they provided, we intentionally developed processes and displays that could be adapted later to serve the needs of other academic advisors and potentially other classes of users as well. To achieve our vision of a robust system that facilitated these support activities, we needed to identify a way to automate the data extraction and transformation processes within an online environment.

Although our research team has access to the LMS system logs and the student information system (SIS), we do not maintain these systems directly nor do we have the necessary infrastructure within our research lab to host an institutional web service. Given this, we concluded that it made sense to partner with an internal organization tasked with building, running, and supporting the highly technical infrastructure—at a broad scale. The university’s Information and Technology (IT) Service was the ideal partner because they had the personnel and technical capacity to address database design, storage, load testing, documentation, user support, and other issues that would inevitably arise when building a system meant to be widely used. We benefitted from an existing relationship with IT based on the development and prior research of the LMS that we were able to leverage and extend for this project. Moreover, IT has been included and is committed to supporting the institutional decisions pertaining to the emergent technologies and processes related to learning analytics.

IT personnel, moreover, had an established track record of maintaining and linking databases within the SIS and related databases (e.g., the university’s data warehouse). Our product would require access to such databases since the nature of the analytics project specified that connections to various student records databases would be necessary. Specifically, our design required that student demographic and course history information would eventually flow into an analytics database. IT personnel had the capacity to devote time and resources to a new project, and we were fortunate to partner with them so that we did not need to hire outside expertise.

Through our partnership with IT, we gained the capacity to extend our work to other academic advisors and to further investigate how, when, and why student performance, effort, and demographic data can inform engagement with students who are in need of academic assistance. Long-term, the partnership will allow our system to scale in size as well as sustainability, meaning
that operational staff can manage the system without the involvement of the researchers who developed the initial design.

Intra-university partnerships can be an important and critical step in the scalability for learning analytics innovations; yet these partnerships are not free from challenges and complex issues. Scholars have identified learning sciences research, in particular, as being well positioned to unpack the different processes and interconnected issues that permeate partnerships established for the purpose of bringing technology innovations to broad scale [5]. This paper is therefore intended to share the benefits and drawbacks of such partnerships and to provide a case study that can inform the broader learning analytics community.

2 BACKGROUND
We developed an Early Warning System, Student Explorer [4], to satisfy a tangible, small-scale problem for the academic advisors in the M-STEM and M-BIO Academies. These academies provide an integrated student development program for first- and second-year undergraduate engineering and biology students [6]. The advisors in these academies required a way to track student performance that would allow timely intervention during the semester, rather than relying on final course grades. Our Student Explorer system provides the program advisors with weekly updates on their students’ academic performance and effort. These updates are presented through dashboards that provide easily interpretable presentations of data that allow the advisors to identify students who are in need of immediate support. Our long-term goal is to scale this functionality broadly so that the “feedback loops” between learners, teachers, and academic advisors can be reduced in time and effort [7].

Our design-based research program is a collaborative effort among the researchers and practitioners who are also our target users [8]. By beginning with a relatively small set of users (the academic advisors) and the related set of data (students' performance and effort), we believed that we could improve the data processes and displays at a relatively low cost in terms of time and effort, while shaping user expectations for continuous iterative development [9].

2.1 Initial Design of Student Explorer
In order to support the academic advisors and their decisions about student engagement, we manually queried information from the LMS Gradebook and Assignments tools in order to track students' performance. We also used LMS course website login events to act as a proxy for student "engagement" that is consistent across all LMS websites. Two linked dashboards (an overall summary screen and individual student/course combination details) included figures displaying students' developing grades and use of the LMS in comparison to their peers, students' performances on specific assignments, and week-to-week classifications of students (Figure 1, left). The institutional LMS' data structure is similar to popular LMS platforms (e.g., BlackBoard, Moodle) and could, in theory, be adapted and deployed using data from those systems.

Based on specific combinations of students' grades and course site login frequency, Student Explorer displays whether advisors should take one of three actions: "encourage" students to keep doing well, "explore" students' progress in more detail, or immediately "engage" students to assess possible academic difficulties. These classifications are generated using three rules: (1) whether a student's percent of points earned is at or above the thresholds of 85%, 75%, or 65%; (2) whether a student is 10% or 5% below the course average in percent of points earned; and (3) whether a student is below the 25th percentile in number of logins. Absolute and relative percentages of points earned are given predominance in our categorization algorithm while each students' "effort" or percentile rank for course logins is used to classify students close to the boundaries of the first two rules.

We opted to design the initial implementation of Student Explorer within Microsoft Excel, which allowed us to present the data in a relatively sophisticated fashion with minimal coding or technical development. After manually extracting and transforming the data into the spreadsheet displays, we distributed the Excel file to the academic advisors on a weekly schedule. While this procedure did, for the first time, allow advisors to view and act on student performance and effort data, the processing time required effectively made the included data approximately 6-7 days old by the time the advisors received the information.

2.2 IT's Selection of BusinessObjects
Before we began developing dashboards, our preliminary exploration of how to conduct Extract, Transform, and Load (ETL) processes for the LMS and SIS data and translate those processes into user-facing dashboards included an investigation of a variety of tools, including Tableau and Pentaho. Such analytics-specific tools are important in order to effectively scale learning analytics systems and achieve long-term sustainability [10]. IT personnel also reviewed the functionality of Pentaho, but in order to move the project forward, they suggested that the existing institutional practice of building an Oracle data structure with ETL processing power would suffice. Layering on that database infrastructure (i.e., the data "universe"), our IT partners suggested that BusinessObjects software could be utilized as a way to replicate most of the functionality of the Excel spreadsheets previously developed by our research team. Collectively, the driving belief was that this solution offered the fastest way to explore the potential of online analytics-powered dashboards with the least effort and at relatively low financial cost.

Before the partnership was established, however, IT explored the feasibility of our project and placed it in their timeline of priorities. This formal process was expected, and once we were folded into their workflow, our partnership with IT allowed us to leverage their existing infrastructure, staff skills, and institutional software licenses of Oracle and BusinessObjects. The affordances created by this arrangement allowed us to adapt our pilot system of Student Explorer and prepare it for a larger audience, while IT managers and technical staff gained experience working in the emergent learning analytics domain [7]. This freed up our team members considerably—we no longer had to spend 5-7 hours per week (on average) manually conducting an ETL process. Instead, these processes were automated against the LMS production and archive servers, and relevant external data (e.g., grading data that resided outside of the LMS) was included as well.

3 CHALLENGES
As with any partnership, various challenges can arise due to unforeseen circumstances as well as different priorities and sensibilities—and our partnership with IT was no exception. In the sections that follow, we detail some of these issues.

3.1 Usability Gaps: BusinessObjects
BusinessObjects is a software tool designed to allow users to use a Graphical User Interface (GUI) to identify columns and related criteria in relational databases in order to construct reports. These
reports are generated in a manner that is reminiscent of Microsoft Excel’s pivot table feature, and produce similar deliverables. The current iteration of Student Explorer simply automates this process. It should be noted, however, that BusinessObjects is not designed with the intention to produce analytics dashboards; it is designed primarily to build tables from database queries.

The IT team has adapted BusinessObjects reports to duplicate and in some cases, expand the functionality of the original Excel spreadsheets our team had previously handcrafted for academic advisors (Figure 1, right). In one example of the expanded functionality, the advisor can now click on any given week in a student’s performance history and see a detailed view of the student’s course performance snapshot as of that date. This allows the advisor to identify any changes on key assessments (e.g., a grade that got changed on a midterm exam).

However, there are several limitations to BusinessObjects in terms of user interface. The IT design team could not, for example, replicate the behavior of our Excel sheets where clicking on the course name opened the student detail report in a new tab. Instead, BusinessObjects could only be used to create two separate reports—this requires the opening of a new browser window or tab. Consequently, when a advisor is using the system to look at data from multiple students, the potential for confusion and frustration with the multiple windows and tabs open (none of which return to the summary page) may prove cumbersome.

Additional quirks about the BusinessObjects interface include the user interface for advancing through multiple pages being available only though a small button located at the top of the window and not at the bottom. Therefore, this button is easy to miss, especially when advisors scroll to the end of the page. Timeout limits from computers accessing the system through wireless connections has also proven to be a difficult challenge to overcome. Consequently, when a advisor is using the system to look at data from multiple students, the potential for confusion and frustration with the multiple windows and tabs open (none of which return to the summary page) may prove cumbersome.

When IT automated the ETL process in the online version of Student Explorer, they encountered the same challenge but the system could not “hand-code” the variations. We arrived at a stopgap solution which involved only counting entered zero values against the student and ignoring all null values. It should be noted, however, this solution masks the issue of students who do not turn in assignments, are absent from class, etc., so some information is treated inaccurately. This troublesome automated treatment of certain types of data risks making the tool less useful overall for the end users, and might ultimately impede the kinds of learning and teaching interventions that the original design made possible.

In a related issue, our research team discovered after working with the academic advisors that some instructors were using their LMS Gradebook to record assessment grades, but chose to remove those scores from the automatic calculation of the course grade (we uncovered a variety of reasons for this behavior including manual calculation of grades, extra credit, and grade curving). In the first (manual) iteration of our system we were able to adjust our database query to include these grades and display them for the academic advisors, while also retaining their exemption from the overall calculation of formative course performance (i.e., points earned divided by points possible). After several months of investigation and testing, the IT team was able to create an additional column in the student detail view indicating whether the individual item was included in the overall class grade.

3.2 Calculation Gaps: Errors in Manipulating Gradebook Data

One technical challenge in scaling up an analytics application is that undergraduate courses often involve careful manipulation of grade curves, optional assignments, extra credit, weighting, and other nuanced ways in which students are assessed. When instructors’ assessment decisions are translated into the LMS Gradebook tool, the resulting data appears to Student Explorer in unexpected ways. These edge cases are easily identifiable by the advisors who are very familiar with specific courses and instructors, but an automated system cannot differentiate between edge cases and standard cases as accurately.

For example, in a high enrollment Chemistry course, there are three subsections of students that share one overall LMS site, and certain assignments (e.g., lab work) apply only to individual sections while other assessments (e.g., the midterm) apply to all students. This leads to gaps in the data (appearing as null values or zeros) that can be caught and re-coded by a knowledgeable human coder. When our team conducted the ETL process into Excel, we were able to hand-code the variations in this course so that the student data could be parsed and aggregated properly by the system.

When IT automated the ETL process in the online version of Student Explorer, they encountered the same challenge but the system could not “hand-code” the variations. We arrived at a stopgap solution which involved only counting entered zero values against the student and ignoring all null values. It should be noted, however, this solution masks the issue of students who do not turn in assignments, are absent from class, etc., so some information is treated inaccurately. This troublesome automated treatment of certain types of data risks making the tool less useful overall for the end users, and might ultimately impede the kinds of learning and teaching interventions that the original design made possible.

In a related issue, our research team discovered after working with the academic advisors that some instructors were using their LMS Gradebook to record assessment grades, but chose to remove those scores from the automatic calculation of the course grade (we uncovered a variety of reasons for this behavior including manual calculation of grades, extra credit, and grade curving). In the first (manual) iteration of our system we were able to adjust our database query to include these grades and display them for the academic advisors, while also retaining their exemption from the overall calculation of formative course performance (i.e., points earned divided by points possible). After several months of investigation and testing, the IT team was able to create an additional column in the student detail view indicating whether the individual item was included in the overall class grade.

![Figure 1. Screen Shots of Student Summary Data Displays in Student Explorer. "Excel" version (Left) and "BusinessObjects" version (Right) Examples.](image)
Our conclusion from these coding issues is that while BusinessObjects is a reasonable cost-savings solution, it may not be nimble enough to be responsive to idiosyncratic cases like the one outlined above. This may end up being an irreconcilable challenge to the ability of the system to scale beyond the M-STEM community of users. Nonetheless, the academic advisors who have used the system have been appreciative of the ability to access the system online and view weekly data from any computer with an Internet connection.

3.3 Access Gaps: Two-Factor Authentication
In order to better secure sensitive data in the institutional data warehouse, such as Family Educational Rights and Privacy Act (FERPA)-protected information, our institution has implemented a two-factor authentication system that utilizes a username/password combination as well as a 6-digit random number stored on a remote keychain assigned to an authorized individual. Since BusinessObjects is the adopted interface used to access data of this sort, it can only be accessed by using the two-factor process outlined above.

Two-factor identification access is not automatically granted to all university staff and must be approved at the management level. Initially, two of the four M-STEM and M-BIO advisors did not have this level of access. Getting the approvals, multiple authorizations, and the physical keychain device took time and energy on the part of the advisors. This authentication process may prove to be a barrier against scalability, as faculty and undergraduates do not routinely have this kind of access. In order to deliver this kind of dashboard to faculty and/or students, it will be necessary to find a technical alternative in the future to be able to deliver the data displays to these users.

3.4 Performance Gaps: Impact on Enterprise Systems
One of the primary reasons we decided to partner with IT is that Student Explorer, by necessity, interacts with both archival as well as production LMS data. As we investigated the possibilities of scaling up our pilot design, IT made sure to include load testing against the production servers as part of the project timeline in order to mitigate the risk of an adverse system event.

The extraction of the Gradebook, Assignments, and login data had a marginal impact on LMS performance in our manual ETL processes, largely due to our limitation to only extract data for the courses in which M-STEM and M-BIO students were enrolled. However, as IT developed the analytics data universe, they decided to include all Gradebook and Assignments data for the current term to accommodate an eventual comprehensive scale for systems like Student Explorer that would access the databases. In October 2012, this larger extraction caused a system failure and unintended shutdown of the production LMS when the servers ran out of allocated memory. While the root cause of this shutdown was quickly rectified, this event highlights the challenges of learning analytics solutions that utilize interrelated systems and processes to produce timely data displays for end users.

3.5 Automatization Gaps: Manual Maintenance of Cohort and Advisor Information
Although critical activities (such as ETL processing) were automatized by the IT team during the evolution of our pilot project, not all components were automatized. For example, importing which student cohorts (e.g., M-STEM and M-BIO students) are included in Student Explorer continues to be a manual process in the current IT-maintained data universe. First, students to be included are identified by the individual programs using admissions metrics from the data warehouse, which are forwarded to IT using spreadsheet files. These files are then uploaded into the database, thus populating a student table. The ETL process then uses the imported lists to match students to their courses and LMS sites. In order to scale this solution in future iterations, the system will have to be able to identify all known groups of students and their corresponding academic and/or programmatic advisor, or we will have to find a way to manually enter this information in an efficient manner.

Finally, many (but not all) students have their assigned academic advisor listed in the data warehouse, but any program affiliation, including the growing number of learning communities and special mentoring programs, like M-STEM and M-BIO, are often missing from this central database. These kinds of gaps in student data can impede scaling of analytics systems, particularly data that serves to group and sort students in order to present relevant analytics to the desired end user.

4 DISCUSSION
The iterative and fluid nature of design research projects often leads to sporadic and non-linear modes of development for new innovations when the work culture of collaborators are dissimilar. For example, for faculty, a casual conversation with a colleague may lead to a new insight and generate ideas to test in a pilot project. By contrast, most institutional technology departments are often run like a-for-profit business, where products and processes are managed in strict pipelines and outcomes are measured in number of users served and terabytes used. These divergent cultures and work processes can be reconciled, but the process of doing so can also lead to misaligned incentives and development delays. Overall, our research team's partnership with IT allowed us to explore and innovate new approaches to leverage learning analytics to provide a pathway to bringing them to scale. However, there were several times when our organizational work processes did not align.

First, IT's implementation of our project was delayed due to existing projects (e.g., LMS hardware replacement) that were a higher priority for the institution. Later, the IT version of Student Explorer was “locked” in terms of additional data inclusion and query changes due to the stringent requirements guidelines by which IT projects are organized. Because IT operates with multiple projects prioritized in their long-term pipeline, a delay in their deliverables is potentially very costly. In the current term, IT is approaching the BusinessObjects delivery of analytics-powered displays as a high priority pilot project. To that end, our research team is routinely contacted to help address technical problems that are necessary to solve in order to scale the system more broadly.

While using BusinessObjects may not be the ultimate solution for scaling-up Student Explorer, IT's commitment to addressing usability and technical challenges has produced a powerful and collaborative partnership that can serve as a model for future learning analytics projects at our institution and beyond. A summary of the challenges described in this paper is provided below (Table 1). We also suggest possible solutions that other learning analytics projects should consider and explore when establishing partnerships between researchers and technologists.
The partnership between our research team and IT is one example of how focused learning analytics innovations can be scaled within an institution. Appropriate skills from both partner organizations have been applied to allow for growth in terms of size and functionality, as well as broader and more generalizable research investigations. Recognition of the challenges and processes highlighted in this paper may be useful to other institutions who are planning to scale learning analytics innovations either within their own infrastructure or by utilizing an outside vendor.

Furthermore, it is important to note that our learning analytics innovation, Student Explorer, utilized student data that is captured and maintained within our home institution. There may be additional challenges when the source data resides with an external vendor, such as whether there is access to non-aggregate data that can be reliably integrated with the institutional data warehouse. Institutional leaders need to be mindful of these issues and challenges when negotiating new contracts educational technology vendors—these contracts need to acknowledge not only the service level agreement, but a data access agreement as well.

As learning analytics solutions serve to break down the technical barriers between "silos" of data, it is important to recognize that technical, cultural, and process-oriented challenges may be unavoidable as different groups of professionals work together toward a common goals of improving teaching and learning in education. Addressing these challenges directly between partners with complimentary expertise has the potential to lead to more successful and broadly applicable learning analytics solutions overall.

5 ACKNOWLEDGEMENTS

We would like to acknowledge Andrew Krumm and Joseph Waddington for their contributions to the original design of Student Explorer. We also thank Gierad Laput, Amine Boudalia, and SungJin Nam for their work in the USE lab, and to the continued participation and feedback from our partners in the M-STEM and M-BIO Academies. Finally, many thanks to Daniel Kiskis and his fellow analytics team members from IT.

6 REFERENCES