Supporting faculty and staff to make better use of learning analytics data

Alexis V. Knaub,1 Benjamin Koester,2 Charles Henderson,1 and Timothy McKay2
1Department of Physics, Western Michigan University, Kalamazoo, MI, 49008
2Department of Physics, University of Michigan, Ann Arbor, MI, 48109

Placing educational data in the hands of faculty and allowing them to explore these data has been of interest for many stakeholders in the past few years, with the hopes that faculty will use these data to improve the educational experience for students. Although these data have the potential to be powerful for improving STEM education, data are only as good as their analyses and application. Using 14 interviews conducted at one institution, we studied what compels faculty to seek out university-collected educational data, what challenges they face, and suggestions they have for improving data usage. Our sample consisted primarily of those already involved with education reform and were using learning analytics data to inform those projects. Challenges included unfamiliarity with quantitative education data and unawareness of the possibilities for learning analytics. We recommend that strategic, thoughtful advertising and already-existing resources be used to help mitigate challenges.

I. INTRODUCTION

We are now in an era when institutions of higher education have the ability to collect and store massive amounts of longitudinal data on their students. These include traditional transcript information such as grades, course registrations, and demographics. In addition, the increased use of online course systems and use of swipe cards to access many services has significantly increased the types of data available. Such data include what areas of course management software students visit and for how long and the use of swipe cards to access tutoring services. The data are not only collected with little effort but also can provide very accurate information about student behavior and performance.

With the massive number of variables collected from numerous students, these rich data sets can be used to uncover obscure information. Examples include the potential for peer interactions based on course enrollment patterns and whether student activity in online course spaces (e.g., participation on discussion boards) is linked to any outcomes such as grades or persistence in the major. Such information can be used to identify trends and potentially predict student success. In recent years, this use of data in educational settings has been referred to as learning analytics or academic analytics.

Researchers and administrators alike are beginning to understand the potential of these data to serve a variety of purposes such as informing student choices (e.g., allowing students to know when they should consider seeking academic support), teaching practices (e.g., informing faculty on how their students perform), and institutional policies (e.g., determining whether and to what extent standardized tests should be used for admissions). However, because learning analytics is a relatively new area [1], we know that education innovations need support to be used productively [2] and users have a variety of backgrounds with educational data, it is important to understand how to best support faculty and staff who use learning analytics data.

To further understand the use of such data in higher education settings, we interviewed faculty and staff regarding their use of analytic data at one institution. Our main research questions were:

1. Why do faculty and staff seek out learning analytics data?
2. What challenges do users encounter?
3. How could the experience of using learning analytics data be improved?

The answers to these questions can help learning analytics advocates understand who is using the data and how to best support them.

II. BACKGROUND

A. What are analytics in education? How are these data being used?

Although a few papers were published pre-2000, the use of big data in higher education truly began in the early 2000s [3]. As a young and emerging field, there is no standardized terminology to describe the use of these large data sets on students in higher education and thus, there are different terms to describe similar work. These terms may also have different definitions depending on the person asked.

One popular term is “learning analytics.” The current definition of learning analytics, as defined by the 1st International Conference on Learning Analytics, entails collecting, analyzing, and reporting student data with the goal of understanding and improving learning conditions [1]. Another term that may be considered a subset of learning analytics is “educational data mining”, which refers to the development of methods and tools to explore these data sets [4]. A third term is “academic analytics” which is described as using these data for statistical purposes and predictive modeling and is often undertaken by institutional research offices as results are aimed at the institution [3, 4]. Some researchers make the
distinction that academic analytics is focused on the institutional and national level interests, whereas learning analytics examines course and department-level interests [5]. For ease of discussion, we will use learning analytics. This is not only how the institution we studied describes this work but also appears to be what they hope to do. Regardless of which term is used or its precise definition, researchers in this field come from and draw from many disciplines such as data mining, social network analysis, and e-learning [1, 3, 6]. These fields helped shape learning analytics data collection and analysis techniques.

The purpose of learning analytics data is to support users in making data-driven decisions [1, 3, 4, 5, 6]. The decisions vary by population. For students, a course management system could provide feedback and suggestions for students based on learning analytics data with the goal of supporting the student [1, 5, 7]. For instructors, learning analytics data might reveal that student performance on certain activities is not correlated to final grades [7]. Instructors may consider modifying the activity or how grades are weighted. For departments, longitudinal data on students could be used to determine if a course-wide intervention was effective.

**B. What are the challenges to using learning analytics data?**

Those working with learning analytics data face several key challenges regarding student privacy and how data are used. Although there is a general awareness within society that data are constantly collected, students may not realize this is happening within educational settings [7, 8]. De-identifying data before institutional use is recommended, as is allowing student to opt-out of being included in the data [8]. However, there are no standard method across institutions for students to opt-out [3]. This may mean that opting out methods are cumbersome or confusing for students at some institutions.

While the data can have positive impacts, it also can have negative impacts if not used responsibly. There is the potential to oversimplify issues and use inappropriate interventions [7, 8]. This is important to note because not all users will have experience with quantitative education data. One ethical dilemma that researchers may encounter is how much information to provide to students and instructors if the information could be discouraging [9]. Additionally, the data provide information on a student at a particular time and with some demographic variables attached [8]. Students have multiple demographic identities so the demographic variables may not fully depict the student. Individual students may experience difficulties in one year that are not present in the next. Also, students can grow from their pasts. A student who struggled in their first year of college may flourishing by the end of their sophomore year. Institutions are encouraged to have experts work with faculty on data interpretation [6, 9]. Learning analytics data are used as a tool to help guide understanding of how to best serve students and should not be used as a substitute for expert input [6].

**III. METHODOLOGY**

The institution used for this case study hired a staff member to focus on learning analytics as part of a larger grant-funded project dedicated to undergraduate education reform in STEM departments. The reform project involved a team of staff members (including a learning analytics staff member) dedicated solely to the reform project and faculty from STEM departments. One of the key goals of this reform is to support STEM faculty use of data to drive teaching decisions. The learning analytics staff member offers an array of services, from simply providing learning analytics data to analyzing the data. Currently, potential users contact the staff member to make a request for data or analysis.

For this project, the staff member provided the research team for this project with a list of 45 learning analytic users. The list included all faculty and staff who received data or had requested analysis. Notes on the users were vague and the staff member did not recall much about interactions or the users’ projects.

We chose interviewees based on their department and involvement or lack thereof with the reform program. Thirty-five users were contacted; the 10 who were not contacted were involved with the reform, and we did not wish to oversample a group that had more opportunities to interact with the staff member. Fourteen (40%) agreed to an interview. Five interviewees were in a STEM department and involved with the reform program, 6 were in a STEM department but not involved with the reform, 3 were in a non-STEM department The sample consisted of faculty (N = 8), staff members (N = 3), postdocs (N = 2), and a graduate student (N = 1). Six interviewees had data analyzed by the learning analytics staff member.

Interviews were semi-structured with questions pertaining to why they sought learning analytics data, what they did with the data, whether they were satisfied with the experience of obtaining and analyzing learning analytics data, and how they would improve this experience. Each interview took place either over the phone or in-person and was recorded. Notes were taken while listening to the recordings, then coded using emergent codes that were refined throughout the study.

**IV. RESULTS**

**A. Why do faculty/staff seek out analytic data?**

Most (N = 10) users sought out the analytics staff member to inform an already existing project, such as a grant-funded program. A minority (N = 4) were beginning to explore research questions with three of the four looking to inform teaching improvements of individual or multiple instructors.
Of the four, three were active in educational reform. This suggests that current learning analytics users are already involved with improving higher education.

Of these four interviewees who were not working on an existing project, two did not receive data or the data they wanted. One felt that it took too long to receive data and was able to obtain it through another channel. The other said that the staff member did not have the exact data for the topic of interest. However, they received data although not what they had in mind.

Each user was interested in different data to inform their particular project. The types of data they were interested in include: grades, demographic data, entry and exit points for majors, and persistence in introductory courses. The data analyses of interest varied among the users. Four interviewees were only interested in descriptive statistics (e.g., frequency of student demographics in a course). For the other 10, the analyses ranged from comparative statistics (e.g., t-tests) to correlations to data mining in order to inform practice. These ten users were using the data as learning analytics intended: to make data-driven decisions [1, 3, 4, 5, 6].

B. What challenges do users encounter?

Interviewees encountered both individual challenges with learning analytics data and challenges with the learning analytics program itself. Seven interviewees indicated individual challenges they have experienced or observed with others (one interviewee indicated two challenges). The majority of individual challenges were due to lack of experience with quantitative education data and research (N = 6). These challenges include learning how to work with social science data (e.g., controlling for variables, using statistics in the social sciences). They were, however, able to teach themselves or find resources on campus to help. One interviewee noted that they struggled with issues related to their human subjects institutional review board (IRB). Because these users were involved with reform efforts, demonstrating an interest and use learning analytics data and may not be aware of the same resources that a user more actively involved with reform would.

Two interviewees did not have time to analyze, interpret, and use learning analytics data. One reason is that they were more novice-level so it would take more time. They also noted that they had responsibilities, such as their non-education research, that had to take priority over working with these data. We again suspect that this challenge will be greater for more casual users. Perhaps providing users with a list of resources (e.g., a statistical center on campus, websites that explain IRB processes) that may help them get acquainted with social science techniques and issues that arise.

However, having help from an expert familiar with learning analytics would be more ideal as an expert would already know the data and may be familiar with how the particular data have been used (e.g., using data to argue for course reform). The literature also suggests having an expert available to support learning analytics research [6, 9]. Although the literature recommends having an expert available to support the interpretation of data, the expert could also help mitigate these challenges related to time and IRB issues. As identified in the next paragraph, though, having an expert on hand may not be adequate if the users do not seek out help.

Eleven interviewees identified challenges they found with the learning analytics program (two interviewees noted two challenges). Most (N = 6) of the interviewees noted that the learning analytics staff member was busy. Half (N = 3) of the interviewees found that the time to get data was too long. They indicated that it could take a few weeks. The other half noted that they did not want to burden the staff member with additional work.

Some (N = 4) interviewees were unsure what data were available and the possible uses of the data. These interviewees had been told that data are available but did not understand the scope of the data (e.g., what variables exist) or what could be done with the data. Even if they knew all the variables available, they might still not understand various ways that the data could be used. This is probably not surprising given the number of fields learning analytics draws from [1, 3, 6]. For example, one interviewee recalled how the learning analytics staff member could infer information about a student’s background and pre-college experiences based on information about students’ home zip codes, which the interviewee did not realize could be done.

A few (N = 3) interviewees felt that the availability of the learning analytics staff member was not advertised well. These interviewees noted that people simply have not heard that such a service was available at the institution. One interviewee described the advertising for learning analytics as “grassroots” and said that they probably would not have heard about the data if they were not involved with the project that hired the learning analytics staff member.

C. What suggestions do users have to make learning analytics data more readily used?

Twelve interviewees offered suggestions about how to promote more use of learning analytics data. The primary suggestions involved increasing the visibility of the learning analytics staff member and the learning analytics work (N = 10). We note that in the previous section, the staff member was already perceived as busy so acting on these suggestions without care may exacerbate that issue.

Increasing the visibility of the learning analytics work could simply be letting others know specifically what data are available and possible scenarios for data usage (N = 5). Some (N = 3) suggested that the learning analytics staff mem-
ber should be placed in a more general institutional unit, such as a teaching and learning center. The rationale is that such places have visibility, the ability to highlight additional programs and services that could support learning analytics use, and is a place where faculty and staff seek similar resources.

Others (N = 5) suggested that creating a web-based data explorer where faculty and staff could play with data would be helpful. They recognized the student privacy issues noted in section II.B. and suggested that a synthetic data set that is similar to the real data could be made available. The learning analytics staff member noted that prospective users are hesitant to go through the IRB process without knowing whether they will receive the data they want. Synthetic data in an explorer would help them know what data are available and determine if it is worth getting real data; the data of interest may not be currently available. This may also help the staff member manage their time. Users could have a clearer idea of what is possible with these data so the staff member would not need to refine a user’s project.

One interviewee emphasized that the advertising must be meaningful. This interviewee felt that this program has not “made the case” for why learning analytics data are useful for faculty and staff. They gave the example that Amazon or Target can reasonably deduce whether someone is pregnant based upon purchases; they would like to see an education equivalent of that such as using learning analytics data to determine whether a student is likely to succeed the class. This is what the learning analytics research is supposed to do [6, 7, 8, 9]. Perhaps more guidance and examples are needed to achieve this goal. However, as noted in the research in section II.B, we emphasize that predictive and inferential uses of these data should be done with care.

The other main suggestion was that interviewees (N = 4) felt additional staff should be available for learning analytics work. This includes an additional learning analytics staff member to handle the sheer volume of data requests and analyses, as well as someone who could support faculty and staff in the development of research questions that could be studied using learning analytics data. While having additional staff is ideal, it is not always feasible. If a learning analytics program cannot hire additional staff, we recommend that staff members manage users’ expectations by communicating a time frame. Perhaps the request process should be streamlined and the steps made clear to users. Users could fill out a form describing what they need. The staff member could then draft a timeline, so the process is clearly communicated and users know that their request will be fulfilled at a specific time rather than simply waiting.

V. LIMITATIONS

The primary limitation to this study is that it is a case study at one institution where the learning analytics researcher was hired to support a reform. Other institutions may see different uses and experience different challenges with learning analytics usage. However, we suspect other institutions experience similar challenges to the ones noted by the interviewees in this study. The challenges are reminiscent of other education innovations in that support is needed for them to be used effectively.

VI. CONCLUSIONS

At this time and institution, learning analytics data are primarily used by those already involved with education reforms. The users in this study noted that learning analytics use faces three primary challenges: the availability of a busy learning analytics staff person, user familiarity with education data, knowing that the data exist. The last two issues may be worse for those who are more casually involved in education reform issues, because they are noted by those who are actively involved in reform. Suggestions to overcome these challenges include having the staff person articulate the process with a timeline to manage user exception, thoughtful advertisement that engages users, and providing resources to support the understanding use of these data.