Summer 7-2016

The Case for Small Data In Higher Education

Jennifer Shewmaker
jws02b@acu.edu

Berlin Fang
bxf13b@acu.edu

Intelligent Environments 2016
P. Novais and S. Konomi (Eds.)

doi:10.3233/978-1-61499-690-3-159

The Case for Small Data in Higher Education

Berlin FANG[a,1], Jennifer SHEWMAKER[b2]

[a] *Abilene Christian University, Abilene, Texas, United States*
[b] *Abilene Christian University, Abilene, Texas, United States*

Abstract. While colleges and universities seek to utilize big, complex and aggregated data to generate insights to support students, big data is limited in its use to design interventions due to limits it inherently has. Small data, or one-dimensional teaching and learning analytics, provide quality feedback to teachers and students. In this paper, we will discuss how Abilene Christian University (ACU), a private university in the western region of the United States, uses alert system to serve at-risk students, IDEA evaluations to gather feedback to improve teaching, and data from the use of a learning management system Canvas to identify ways to improve student learning by providing formative evaluation for faculty to improve their practices. Towards the end of the paper we will discuss what students can do to make informed choices about their learning.

Keywords. Small data, learning analytics, teacher evaluation, Learning management Systems, learning tools

Introduction

Big data became trendy partly due to a few engaging stories, including one published in *New York Times* about a furious father complaining to retail giant Target that his high-school daughter was getting baby coupons from Target [1]. We were told that Target, through predictive models developed by statistician Andrew Pole and his team, was able to make fairly accurate predictions about pregnancies of its customers, which enables Target to conduct rather precise and customized marketing to its customers. The story shows the public the power of big data.

Big data is a collection of large, complex sets of data from diverse sources and requiring a variety of tools for gathering, analysis and display. *Forbes* contributor Gil Press traces the evolution of the concept, which is reminiscent of a Cinderella story. "Big data" started as a problem, due to its size and complexity, which presented challenges to both computers and humans trying to analyze them. The definition from the *Oxford English Dictionary* also focuses on the thorny aspects of "big data", claiming that the large size of data "present significant logistical challenges"[2]. It became a gold mine only when organizations, especially government entities and commercial enterprises, discovered that multiple sets of data actually yield insights

1 Corresponding author. Email: bxf13b@acu.edu
2 Corresponding author. Email: Jennifer.shewmaker@acu.edu

about people, problems, and places that had been previously inaccessible. This shift of mindset made big data a big trend.

Big data provides new hope of fundamentally changing education. However, is the confidence in using big data to change education justifiable? Will people be able to find the Andrew Poles in education to connect the dots of data to make a prediction of learning outcomes for specific students? We have found that there are certain areas where big data can play significant roles in education, especially higher education, where the authors work. Alumni outreach programs make use of multiple types of data to maintain or develop relations with alumni and to increase effectiveness of fundraising. Student recruitment offices take advantage of diverse data to reach potential students. In either case, big data plays a positive role as both alumni and potential students are not immediately within reach through more conventional methods.

Big data is not as useful in student learning as it is in business settings. It is not doing anything game-changing in the classroom yet. Teachers are wary of the term as possibly one more bandwagon the rest of the world is jumping on. Our center, the Adams Center for Teaching and Learning, has organized a few sessions on big data in teaching and learning, but there was low interest in these sessions. Of all the educational trends out there, big data may be the most familiar term that people know the least about. A few problems in the concept or the diffusion process may have contributed to the lack of interest.

1. Problems with “Big Data” in Higher Education

1.1. Mystery:

The word “data” has an aurora of scientific rigor to it. That does not prevent people from using “big data” the way others use palms or tea leaves for prediction. Dots are connected in spite of gigantic holes between them. Theoretically the more data you collect, the fewer and smaller holes there will be. In actuality, collecting such data can present human and financial challenges that are not rewarded by correspondingly useful results. Therefore, there will be holes, or incomplete data, and there are currently not good ways around this problem. Users may accept having biased or incomplete data, which lead to unreliable conclusions. Some may accept this practice as rigorous simply because it is “data-driven”. This dilemma has always characterized scientific research, but big data poses a special risk as the word “big” connotes vast amounts of data, and that can project false confidence and cause readers to ignore perfunctory behaviors in the process of establishing conclusions.

1.2. Capacity:

The second problem is that institutions of higher learning, especially small ones, are rarely capable of generating large sets of data to arrive at the kind of insights Target may have. Ivy-league institutions, for instance, are capable of collecting and utilizing large data sets, but they are not desperately in need of insights for student recruitment and retention. Struggling colleges and universities, who could benefit most from such data, may not have the resources for collecting them. At such institutions reliance on big data for insights may create dependency on sophisticated tools and pools, as well as excuses for not reaching out to populations organizations are supposed to serve.

1.3. Necessity:

The third problem with big data is the necessity. The Targets of the world cannot directly ask about people's pregnancies. They have to develop insights indirectly, for instance, by analyzing purchase records rather than asking consumers directly, which make big data methods relevant. It is discouraged for companies to be too inquisitive about their consumers due to consumer rights laws and privacy culture. A company's relationship with customers is limited, to a large extent, to selling and buying. For colleges and universities operating in a different mode, it is encouraged to care for students as whole persons, to build personal relationships where appropriate. Insights about students can come more organically, through class discussions, student papers and projects, office hours, and other methods of interaction.

1.4. Majority:

The fourth problem with big data is that it may create a "silent majority". Big data is helpful for classifying and clustering [3], distributing people in groups. The radars of big data can detect individuals on the margins or heading towards margins, which is of course useful, as we will discuss below. However, the remaining individuals are lumped together so much so that their learning processes remain unknown. It is not desirable to see students as dots in large data sets, but as full human beings. Educational institutions, when obsessed with data collection, analysis and utilization, may know substantially about groups of people, but little about the individuals within these groups. Without personally reaching out to individuals, insights about groups cannot be fully used.

We strongly agree that data, not big data, is valuable to higher education. When educators claim they use big data to improve teaching and learning, often we find it is a misnomer. Educational agencies are not gathering large, complex data sets from various sources and in various dimensions as government or businesses do. Rather, when educators say they use data, they actually mean learning or teaching analytics, or rather, "small data", data existing in smaller units, not necessarily complex, but having a huge impact. There is not actually a need to favor "big" for its own sake. If small data can help student learning more, we recommend educational institutions focus on them. In the following sections, we will describe three types of "small data" to illustrate how such data can have impact.

2. Using Data for At-risk Student Management

At-risk students refers to students who risk failing in their personal lives or academic study, transferring to another school or simply dropping out of school. They are the types of students that lead to low retention rates, which in turn negatively affect college budgets and rankings, as well as reputation among potential students, alumni and other stakeholders. The United States is known for the quality of its higher education, but also for the quantity of choice and freedoms students enjoy. Such choice and freedoms make it easy for students to leave but raise the cost for their institutions. Having predictive data would help identify some of the concerns and issues at-risk students may encounter, and subsequently designed interventions can change their attitudes and behaviors to make it undesirable for them to fail, transfer or drop.

Abilene Christian University, for instance, has a program called SOAR (Student Opportunities, Advocacy and Resources) to collect multiple types of data about at-risk students. When professors, academic advisors, resident hall directors, coaches and other parties in close relationship with students suspect that a student may be at-risk for potential problems, they can send alerts to the SOAR program, which then recommends interventions.

Alerts can include such at-risk information as: 1) Academic issues such as grades, attendance, homework, test performance or anxiety, and inability to cope with academic pressures; 2) Physical health issues such as illness, lack of sleep, signs of abuse, and inability to perform required tasks; 3) Mental health issues such as thoughts, threats of suicide, needs encouragement, relationship troubles, depression, stress, anxiety; 4) Life management issues such as crisis at home, loss of family member, loss of friend, and financial trouble; 5) Major/career issues, such as undecided major or unsuitable Major; 6)Behavioral issues such as inappropriate behavior, disruptive behavior, verbal harassment, physical harassment, 6) retention issues, such as considering transferring, and being unable to return due to finances and 7) other issues [4].

Such data are collected and aggregated. ACU also recommends that alert senders leave their own names and contact information so that they can become involved at varying degrees in solving problems. This type of partnership is possible due to the campus culture of this faith-based institution to encourage faculty and staff to care for students. However, not all alert reporters, including faculty, are willing or capable of solving varied problems they help to detect due to lack of time, energy, interest or expertise. SOAR functions as a liaison to connect students to other university resources, such as the counseling center for issues like stress, anxiety, family crisis and suicidal tendencies. Financial difficulties can be referred to the office of financial aid. Problems with academic studies can be helped by such services as the tutoring program, the writing center, the speaking center, or the research desk at the library. This referral-aggregation- intervention loop makes data actionable and useful.

3. Using Data to Improve Teaching

Commercial organizations may purchase or collect large and complex data for better understanding of consumers. Colleges and universities are in unique positions to learn directly about students and how they experience their learning. It is fairly easy to collect their feedback, as all students may appear in class at the same time, or are required to visit the same class sites. A fairly universal practice for colleges and universities is to collect class feedback at the end of a semester. The drawback of this method is that many such class feedback forms are too generic to be of much use. Students fill out online or Scantron forms basically to pass a judgment on teachers and their teaching, not much different from what students post at sites like "Ratemyprofessor". Such feedback, instead of providing insights about aspects of teaching, could create fear and resentment among faculty. In other words, such feedback is summative, but not formative evaluation. The difference between the two is thus summarized by education professor Dr. Bernard D. Bull from Concordia

University Wisconsin: “Formative feedback is the annual checkup at the doctor. Summative feedback is the autopsy.” [5] As reputation, tenure or promotion are at stake, some teachers dismiss such feedback altogether or do not cooperate in gathering feedback unless they are forced to do so by the administration.

Recognizing these difficulties, in 2013, Abilene Christian University began to consider the adoption of a more useful evaluation or feedback system. The Director of Faculty Development in the Adams Center for Teaching and Learning at that time spent an academic year investigating a variety of systems to find one that would provide the University and the faculty with useful student feedback on their courses. In 2014, a pilot program was run, and then in 2015, Abilene Christian University adopted a teacher evaluation mechanism through IDEA Center, a non-profit organization dedicated to improving the quality of teaching. This feedback form asks teachers to rate instructional objectives, as the importance differ across disciplines and grade levels. Allowing teachers to rate objectives for their own classes restores the autonomy for teachers in judging which objectives are minor, important, or essential for their own courses. To use this kind of front-end measure of objective rating, an institution also projects trust in teachers. Such ratings are factored in final adjusted (versus raw) scores.

The feedback form is provided to the students three weeks before the end of the academic semester. Abilene Christian University has adopted the online version of the form, so the evaluation is administered to students via an email link that is shared by their professors. Students are then asked to rate teachers and teaching using evidence-based research findings. The questions are distributed in a number of categories covering various aspects of the student experience, such as “the instructor”, “progress”, “the course”, and extra questions teachers may ask themselves.

In the end, faculty receive a diagnostic form which includes a raw score and an adjusted score for each item, with the adjusted score being related to teacher reported importance of objectives. Teachers’ scores are entered into a larger database against which teachers can see their own ratings compared with the average rating for those at one’s institution, those within one’s department, and all of those teaching similar courses within the IDEA system. This provides the faculty member with important information about his or her own student feedback. For example, if all teachers for a particular course or type of course have received low or high ratings, then it provides a relevant set of scores with which to compare one’s own ratings.

This evaluation process helps to raise awareness among faculty about best practices in teaching by providing them with teaching practices that are aligned with the learning objectives that they have identified as important. For instance, they will be exposed to the idea that peer learning and group work help in the learning process (“Asking students to help each other understand ideas or concepts”). Teachers can then change teaching methods or materials based on the diagnosis. The feedback mechanism is also “actionable,” by which we mean that if student responses show low ratings in a particular area, there are resources from IDEA Center to guide teachers to make improvements. Alternatively, institutions can develop programs to train faculty in areas where they do not do well, according to analysis of student responses.

At Abilene Christian University, the Adams Center for Teaching and Learning has worked closely with Educational Technology to provide guidance for faculty, department chairs, and deans on how to best use the information obtained from the IDEA system. There were numerous open training sessions for faculty members and administrators to learn about the IDEA system, how to complete the faculty information form to gather the most useful data, and how to interpret results. Adams Center staff also visited each college and department by appointment to provide more personalized training. In addition, now that the University has completed one full year of use with the IDEA system, the Center for Teaching and Learning can examine the overall results to see the areas in which faculty may need the most support. This provides the opportunity for targeted faculty development based on the needs of the faculty as a whole.

4. Using A Learning Management System to Improve Learning

We have also found that powerful, actionable data can be collected through the constant use of a learning management system (LMS). Such systems help teachers gain instant understanding of how a class or a particular student is doing before it is too late to intervene. Most universities in the United States use some kind of learning management system, such as Canvas, Blackboard, or Moodle. Instructional content, such as lecture notes, outlines, audio or video content, can be posted for students to access. Learning management systems can also include various learning activities, such as discussions, quizzing, assignment collection or grading, and web conferencing. Most learning management systems have some course analytics tools but they may not always be known or used by teachers. In the following section, we illustrate the use of several analytics tools in Canvas, and the information or insights they generate.

4.1 Overall class activities analytics

"Overall class activities" tool shows student activity by date, submissions and grades. It is an overview that shows how the students interact with the course content. For instance, teachers can see which days students are most active in viewing and interacting or which assignments have the most entries of late or no submissions. This gives teachers various insights into learning processes, such as the best time to release content for greater participation, which assignments seem to be too difficult or easy (as indicated by grade distribution), or when students seem to be overloaded or not sufficiently loaded, as indicated by a higher number of late or no submissions.

4.2. Student overall activities analytics

Student activities report shows student page-views, participation frequencies, submissions (including on time, late or missing submissions) and their scores. Teachers can rank order these items and correlate their activities with their scores. The form gives teachers the tool to identify academically at-risk students in terms of their scores or behaviors. Such knowledge can be used to send alerts to them as a group (for instance, all students who have missing assignments) or contact students individually to inquire what problems they may have and how they can get help.

4.3. Individual student analytics

Canvas analytics will also allow faculty to check specifically how a particular student is performing. Teachers can tell from a student's individual reports how he or she is doing in class in terms of: 1) how active the student is in viewing class content or participating in class activities; 2) how the student manages his or her time in submissions; and 3) how this student does in grades compared to the rest of the class. This analysis gives a teacher information that can be used to design individualized intervention for a student.

4.4. Quiz analysis

Quiz summary shows teachers how grades are distributed for any tests or quizzes given in the learning management system, including average score, high score, low score, standard deviation and average time it takes to complete the quiz. This has implication for faculty on the appropriate level of difficulty that has been set, allowing teachers to make adjustments if needed.

4.5. Question analysis

Teachers can also find out about specific questions in a quiz. For instance, how many students get a particular question right, and which answers are most often chosen. It also provides information about the discriminating factor of a question, which is how a response to a question correlates to the final score of the test in general. If the discrimination index is high, it means that students scoring high for the quiz in general have a higher chance of getting the question right. If the discriminating index is zero, it means generally high or low-performing students have equal chance of getting this question right. If the discriminating factor is negative, it means students who get this question wrong actually perform better on the rest of the quiz. This may make the teachers decide to throw out certain questions or revise them depending on the purpose of the quiz (for placement or for mastery). If a question proves to be a “trick question” (students with good grades do not answer correctly), teachers may decide to rewrite it for clarity.

4.6. Test log

The LMS Canvas also has a feature called “action log” for tests, recording how students behave while taking a test. For instance, how much time a student spends on a question. A student may be shown to be going back and forth on a particular question, which may indicate that there is a knowledge or skill that he or she is still struggling with. The log also records if a student has left the test page and come back afterwards. It could be interpreted as a sign that the student has left the page to search for an answer. If such activities increase, teachers may decide to use secure testing environment, or change the format of testing from closed book to open book. Use of the LMS quizzing feature helps teachers to check digital tracks students leave behind when there is a concern. More importantly, simply telling or showing students how much data can be found will make them more accountable for their learning. Additional tools, such as Turnitin, a plagiarism detection software, can also be integrated into an LMS to generate additional data to make sure students submit original work, instead of plagiarizing from other sources.

Data from a learning management system can be instrumental for teacher decisions and actions during a course. For instance, they can rank participation activities to find out who is not participating as much in the class, and subsequently reaching out to students to prompt them to complete learning tasks. With test and question analysis, teachers can revise test banks so that new iterations of the course can become better. Teachers can make such changes within the current semester. For example, quiz questions are often reused in midterm or final exams. Information on quizzes in the earlier part of the semester can immediately change questions on midterm and final exams. Such data, when released to students, help them to determine how they are doing in their class. It is also possible to use these analytic tools to develop good study habits.

5. Guidelines for Using Data for Learning in Higher Education

Issues of generalization versus particularization are core issues in machine learning for personal artificial intelligence agents. There is not much value in using big data in conjunction with artificial intelligence to plan instruction. Using algorithms to generate student profiles, as shown in the research of Mihăilescu, Niţă & Pau may be useful for analyzing large populations [6]. However, in higher education, analytics trumps algorithms in generating insights. Analysis from larger data sets yield help that is, as the Chinese saying goes, “distant water that cannot quench immediate thirst.” Teachers can benefit more by using small data analytics. Here are some recommendations we have for educators in colleges and universities:

Table 1. Guidelines for using small data to drive learning

Student insights	Small data sources	Measurements	Interventions
Does a student really read learning materials?	-Access report from LMS	- Frequency of logging in the system - Page accessed - Time accessed - Completion of content viewing	- State in syllabus about access requirements - Use announcements to require access - Send email to individual students urging them to access
Does a student actively participate in class activities?	-Interaction report from LMS -Assignment analysis -Assessment analysis	- Number or quality of interaction - Number or quality of responses - Timely completion of assignments - Timely completion of assessment	- Set expectations and rules - Create activities with clear due dates - Sequence content with modules - Provide scaffolding to support student when they struggle - Use office hour to assist
Does the student show learning during the course?	-Assessment report from LMS - Data from campus services	- Quiz/exam results analysis - Question analysis - Individual performance analysis - Student background report, such as test grades and other courses taken	- Use frequent low-stake quizzes - Work with class on questions difficult for students - Work with individuals on specific content - Provide alternative assessment
Does the student show learning at the end of course?	- Gradebook in LMS	- Distribution of grades	-Recommend additional resources - Recommend remedial services

6. Advise Students to Use Data to Improve Learning

Educators can also advise students to make use of data to inform their learning, or to cultivate healthy study habits. Here are a few suggestions educators can give to students:

6.1. Break data bubbles

Big data can create risks not only of privacy violations, but also of forming bubbles within our comfort zones. Netflix, for instance, is capable of using our viewing behaviors to predict similar movies and shows, when it might benefit us more to diversify into other genres and products from regions with which we are not familiar. Similarly, if educational institutions or vendors claim to be able to recommend learning products based on previous records, students should be cautioned about them, as learning is about "a permanent change of behavior" due to experience (behaviorism) or mental associations (cognitivism) [7]. If students keep resorting to similar content, they will fail to change behaviors due to the lack of new experiences and associations. Instead students could gain greater experience by stretching themselves beyond familiar content and getting advice and feedback from experts in fields.

6.2. Develop healthy digital habits

Research shows that 50% of our decisions come from our habits rather than a deliberate decision-making process [8][9]. Healthy habits help us by routinizing constantly performed tasks, which helps to free mental bandwidth for complex learning habits. But such automation can also be costly as we may fail to perceive subtle changes in external cues and thereby fail to learn or notice new information. Unhealthy habits are especially worrisome. For instance, constantly checking social media may cost us time and concentration when we should be performing deliberate practice or reflection. One way of reducing this problem is to increase cues that prompt us for healthy habits, and to avoid cues prompting us towards unhealthy habits. Examples include: Turning off notifications for social media so that we do not get constantly pulled away from tasks at time or putting smart phones away while studying.

6.3. Develop quantified self

Students could be reminded of what Wired magazine journalist Gary Wolf calls "quantified self" [10]. In his TED Talk, Wolf shows how humans can use digital gadgets, especially wearable ones, to capture personal health data, such as calorie intake, steps walked, and blood pressure, which may stimulate users to develop healthy life habits. Students may try to capture data of learning habits. For instance, keeping a daily time log to record study activities and the time spent. This encourages students to cut down on activities that distract, disrupt or interfere with study goals. Computers can also automatically generate logs of time spent on different online activities. For instance, RescueTime (https://www.rescuetime.com/) catalogues online behaviors to help students understand how they spend their time. Such information provides students with "quantified self" data in order to better manage their time.

6.4. Generate their own data

Instead of using technology simply for entertainment and distraction, educators can also encourage students to study and generate data that provides information about their learning. Quizlet (https://quizlet.com) allows students to study and test themselves, as well as using games to make the learning process more engaging. The tool also generates some fairly simple data (we call them micro data, data from one activity, for one person), such as completion rate, scoring, and items often miss that may help the student to decide what to do next. A mobile tool called Brainscape (https://www.brainscape.com/) also provides practice and generates micro data, using that data to develop items for further practice. Users can rate their familiarity with a concept on a scale of 5, and the application reshuffles the deck of cards based on self-rated familiarity. Data is generated and incorporated almost simultaneously in new iterations of practice.

7. Conclusion

In an age when "big data" is the catchword, educators should not seek to resemble retail stores such as Target in its ability to use big data for consumer prediction. While large, complex data sets help in many areas in higher education, such as alumni relations management, by far the greater treasure is in "small data" gathered using simple platforms which have rich and actionable implications for the improvement of teaching and learning. Digital tools can be used to generate data, which will result in better connections between students, their learning materials, assessments and learning outcomes. Similarly, students can use tools to gather data about their habits and learning processes for better learning outcomes. We have illustrated our points by providing examples from our own institution, which of course is limited. We welcome future research to extend the discussion.

References

[1] Duhigg, C. (2012) How companies learn your Secrets. *Nytimes.com*. Retrieved 31 March 2016, from http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html
[2] Press, G. (2013). A very short history of big data. Forbes Tech Magazine, May, 9.
[3] Sin, K., & Muthu, L. (2015). Application of big data in education data mining and learning analytics -- a literature review. ICTACT Journal on Soft Computing, 5(4), 1035–1049.
[4] Abilene Christian University SOAR Program (2016). Student Referral Form. ACU SOAR Program. Retrieved 28 April 2016, from http://www.acu.edu/legacy/soar/referral.html
[5] Bull, B. E. (2014). 10 Assessment design tips for increasing online student retention, satisfaction and learning. *Faculty Focus*. Retrieved 2 April 2016, from http://www.facultyfocus.com/articles/online-education/10-assessment-design-tips-increasing-retention-satisfaction-student-learning-online-courses/
[6] Mihăilescu, M. I, Niţă, ş. L., & Pau, V. C. (2016). New big data model based on social learning environment using artificial intelligence. *Elearning & Software For Education*, (1), 428-435. doi:10.12753/2066-026X-16-062
[7] Ormrod, J. E. (1999). Human learning (3rd ed). Upper Saddle River, N.J: Merrill.
[8] Neal, D. T., Wood, W., & Quinn, J. M. (2006). Habits—A repeat performance. *Current Directions in Psychological Science*, 15(4), 198–202.
[9] Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, 83(6), 1281–1297
[10] Wolf, G. (2010, June). The quantified self. *TED Talk*, Cannes. Retrieved 1 May 2016, from https://www.ted.com/talks/gary_wolf_the_quantified_self