

Data Extraction from Web-Based Learning Management Systems

M. Austin Creasy

Assistant Professor, Mechanical Engineering Technology, Purdue University

Email: mcreasy@purdue.edu

Abstract

Many schools and universities are using learning management systems (LMS) as a means of providing students with course material. Many of these LMSs are web-based and allow students the opportunity to access course material from numerous locations and from numerous devices. Students are able to review notes, see homework, submit homework, prepare for tests, review course grades, contact other students, chat with their professor, etc. These LMSs are designed to assist students with learning course material and assist the instructor in providing course material. The features of these LMSs provide the course instructor with numerous tools that allow them to assist students in mastering course material that have not been previously available in education. The programming tools allow instructors to vary the course content in numerous ways so that students are able to learn at different rates and in different ways. These features also provide the instructor with data that had not been previously available. Such data can be used to assess learning pedagogy and other details of learning. Depending on the LMS, data can be extracted from the students' actions and responses and used to assess the activities performed. This paper will review the literature on the use of LMS in academics and review the literature on data extraction and share some of the experiences that the author has had in using this tool for academics and learning assessments.

Introduction

Learning management systems (LMS), also called course management systems (CMS), are integrated computer systems that assist in the presentation of courses to students.¹ These systems have changed the way that courses are taught and presented to students¹ and have changed the experiences and way that instructors teach courses.² LMSs provide numerous tools to instructors that were not previously available. Many of the tools are unknown to instructors because they are course content specialist and do not have training or experience in the software or pedagogy that the LMS uses. The tools typically include communication, material delivery, assessments, and user management.¹ All of these tools can be used in educational research if the information is extracted from the LMS.

Data extraction, also known as data mining, is an emerging discipline that studies how to explore data from educational sources.³⁻⁵ This field of research has grown in conjunction with the increase of new technologies that record different types of data. Romero³ reviewed educational data mining and explained how large volumes of data are collected from web-based educational tools that can be used to assess anything from student learning to pedagogy. Some researchers have stated that data mining is "discovery driven"⁶ and not hypothesis driven. This statement

means that the data found from the extraction activity is what finds the “patterns and tendencies”⁴ from the LMS.

Learning management systems are numerous and come in many forms.⁷⁻¹¹ Some are commercial closed-source systems like Blackboard and others are open-source systems like Lon-Capa. Each system has different components that are available to the instructor as previously explained and therefore data extraction from different systems is varied. This work will overview data extraction from the Blackboard system, as that is the system used by the author of this paper. The Blackboard LMS components include file sharing, grade tracking, on-line assignment submission, announcements, on-line discussions, student help tools, and many other assistant tools for students. In 2007, reference¹⁰ reported that over 70 percent of colleges and university were using the Blackboard LMS. Therefore, a majority of academics have this LMS available for data extraction for different types of learning and pedagogy assessments.

The author of this paper uses the Blackboard LMS for every class taught, which averages about 8 courses per year. These courses have included traditionally taught classes, flipped courses¹², and a distance course. The majority of the courses uses similar components of the LMS and does not include a comprehensive use of the all of the components of the LMS. The major components used are file sharing, grade tracking, and on-line assignment submissions. The author will focus on showing examples of data extraction from these components used in the courses offered.

Data Extraction Examples

Data extracted from the Blackboard LMS system includes details of usage of the LMS system and the LMS is capable of recording individual student usage. Details of when students download specific files or reviewed files are recorded, on-line assignment times are recorded, on-line assignment answers are recorded, etc. This section will show examples of data extracted and what information that data provides to the assessor.

Student usage of specific files is a key piece of information that a LMS can store. Figure 1 shows an example of when students are viewing a specific file available to them in the LMS as processed by the LMS system. The abscissa shows the dates of the extracted data and the ordinate shows the number of times that the file was viewed by someone in the class. The results show that on January 18, the students were using this file significantly more than any other date during the time range. Now this data can be assessed with other information that was occurring in the course to answer why Sunday, January 18 was an important day to view this file. Additional information not shown about this specific file includes the time of the day that students typically view the file and which days of the week students are viewing the file. This data can be extracted for the entire class or for a specific student. This extraction allows you to make generalizations about the entire class or view the progress of individual students that may need additional assistance. This data can be used as evidence for the actions of students that may have questionable academic honesty standards. This data may be extracted⁷⁻¹¹ in numerous formats that include spread-sheet applications for additional processing.

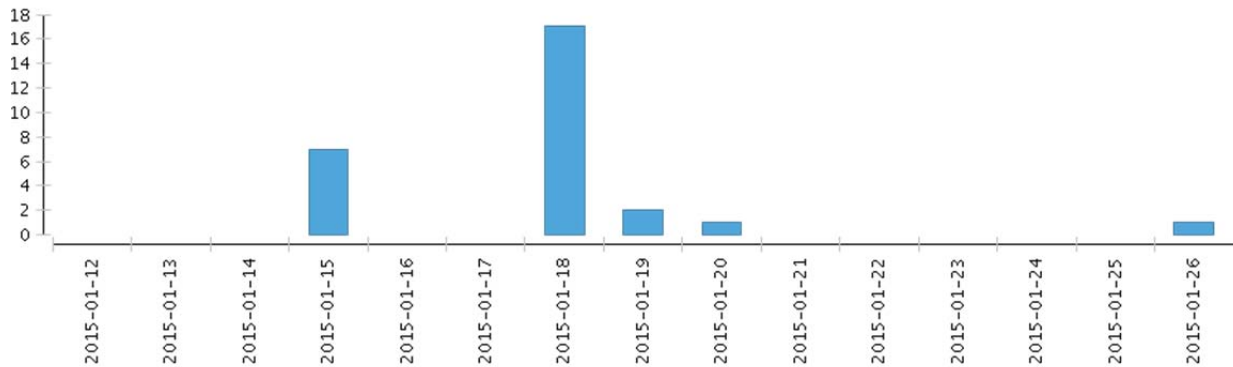


Figure 1. File usage data extracted from Blackboard LMS over a specific time range with number of views per day recorded.

In reference¹², the Blackboard LMS is used with flipped mechanics courses. These classes provide students with notes to be reviewed before class. Class time is used to practice additional example problems and answer homework questions. Homework is assigned online with unlimited submissions and students gain access to the solutions once 80 percent of the problems are answered correctly. One issue discussed in the reference is that one student may be able to complete the assignment and provide the answers to the remaining members of the course. Data extracted from the LMS system is able to show which students are benefitting from their classmates by extracting the scores, times, and answers of the on-line submissions. Figure 2 shows some of the information that may be extracted from this data without viewing individual answers. This data shows when a student attempted an assignment, their score, the start and end times, and how many attempts are made. The submission answers for each attempt can be viewed as well in order to see where students are having problems with specific concepts.

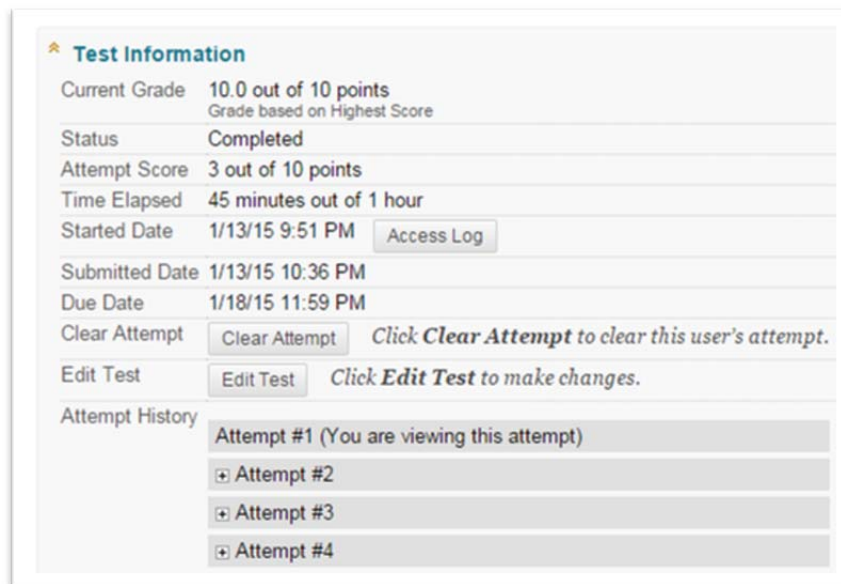


Figure 2. Information extracted from Blackboard that provided details of student work.

Discussion

Tsantis⁶ stated that data mining is not hypothesis driven. This statement is true if a researcher is searching for unknown patterns or tendencies in the data. A hypothesis can be formulated and tested using data mining from LMS of courses that have been taught. Once the data is collected, different hypothesis can be formulated and tested on the data in order to see if there are any relationships present. Many users of LMS have huge amounts of data that can be analyzed for the various teaching styles, different pedagogy usage, or different student usage. One of the largest hurdles may be extracting the data from the LMS without knowing what data points you are initially interested in analyzing. The Blackboard LMS provides the user with the ability to extract numerous types of data into spreadsheets. But, not all of the data is extracted into the spreadsheets. The author has analyzed what students do during different homework exercises that are submitted online. The LMS provides the information seen in Figure 2 and the LMS allows you to download the responses of all of the students into an excel spreadsheet, but it does not include all the time stamp information in the spreadsheet as seen in Figure 2. These time stamps could be beneficial in analyzing the data depending on what hypothesis that you formulate to test. Therefore, if your hypothesis depends on this data, the extra data needs to be extracted for post-processing.

Using data mining without a hypothesis and details of other information within the data may cause false findings. Figure 1 shows the usage of the homework solutions file from a class taught like reference¹² for the first week's homework. Figure 3 shows the usage of the homework solutions file from the same class for the second week's homework. This figure shows that the major usage of the file was on Sunday, January 25. For Figure 1, the major usage of the file was on Sunday, January 18. Data mining without context or a hypothesis could conclude that students access homework solutions for this class typically on Sundays. Homework sets are due at 11:59 pm on Sundays and as previously explained, the solutions are released to the students after they score 80 percent on the assignment. Therefore the students are typically finishing their homework on Sunday and looking up the last couple of answers to obtain 100 percent on the assignment. Without the context and the hypothesis, strict data mining as explained by Tsantis⁶ may cause incorrect conclusions to data obtained from a LMS.

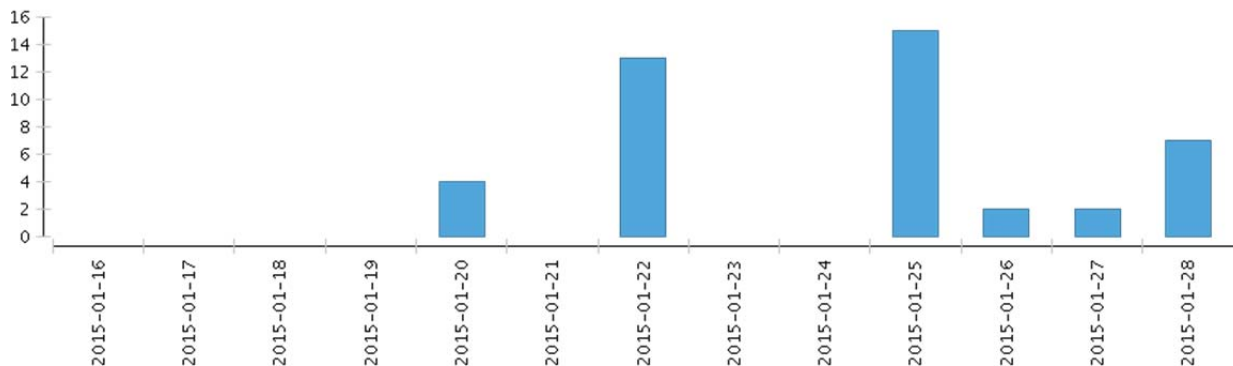


Figure 3. File usage data extracted from Blackboard LMS over a specific time range with number of views per day recorded.

Data extraction has also been used to verify the work performed by students. Some students use classmates to finish assignments where credit is granted and used toward the final grade. Sometimes determining when this action has occurred is difficult without witnesses. A LMS stores information that will provide details of when this type of action has occurred. The time stamps of the submissions and the submitted answers are excellent sources of data for determining this action. These time stamps allow the professor to view the time duration used by each student to finish the assignment and the professor can view the number of times that a student submitted an assignment. For some students, the time stamp would show numerous hours of work and numerous attempts as they worked to solve the problems. Other students would contain time stamps below 10 minutes with few attempts and obtain a perfect score.

Data extraction has also been used to verify student complaints about homework length. A new faculty member using parts of the class similar to the flipped class of reference¹² has several students complaining about on-line submissions and the time to finish the problems. This student stated that the work was taking between two and three hours each night to finish. The data extracted showed that student was working on the homework for less than an hour. Therefore this data can be used to understand student behaviors in their classwork or to provide data for students that complain about specific class activities.

Currently, data extraction from the Blackboard LMS is being used to assess several different hypotheses about learning. The hypotheses relate to feedback on on-line submitted assignment, time dedication to homework assignments and how that relates to overall performance, how frequently students use provided notes and to what extent, and the differences between freshman and senior students in performance for the specific courses. The hypotheses for each research question vary, but the data that is extracted is dependent on the hypothesis. Therefore the hypothesis needs to be defined in order to make sure that the data is extracted from the LMS to answer that question.

Conclusion

Data extraction from LMS can assist in assessing learning and pedagogy. The Blackboard learning system is used to extract data for different types of assessments. Some researchers use this type of data in order to define trends and let the data precede the hypothesis. The problem is that some data may not be included in the mined data and therefore incorrect trends will be found. Data extraction should follow a hypothesis that will lead to the needed data based on all of the variables available. Once the hypothesis is defined, the data can be extracted that is needed to verify the hypothesis. LMS record vast amounts of data and many users of these systems have vast amounts of data that can be analyzed with some digging and questioning of the teaching activities, the student performance, and the teaching styles used.

References

1. Coates, H., James, R., and Baldwin, G. (2005) A critical examination of the effects of learning management systems on university teaching and learning, *Tertiary education and management* 11, 19-36.
2. West, R. E., Waddoups, G., and Graham, C. R. (2007) Understanding the experiences of instructors as they adopt a course management system, *Educational Technology Research and Development* 55, 1-26.
3. Romero, C., and Ventura, S. (2007) Educational data mining: A survey from 1995 to 2005, *Expert systems with applications* 33, 135-146.
4. Romero, C., Ventura, S., and García, E. (2008) Data mining in course management systems: Moodle case study and tutorial, *Computers & Education* 51, 368-384.
5. Romero, C., Ventura, S., Pechenizkiy, M., and Baker, R. S. (2011) *Handbook of educational data mining*, CRC Press.
6. Tsantis, L., and Castellani, J. (2001) Enhancing learning environments through solution-based knowledge discovery tools: Forecasting for self-perpetuating systemic reform, *Journal of Special Education Technology* 16, 39-52.
7. Kortemeyer, G., Kashy, E., Benenson, W., and Bauer, W. (2008) Experiences using the open-source learning content management and assessment system LON-CAPA in introductory physics courses, *American Journal of Physics* 76, 438-444.
8. Steif, P. S., and Dollár, A. (2009) Study of Usage Patterns and Learning Gains in a Web-based Interactive Static Course, *Journal of Engineering Education* 98, 321-333.
9. Larkin, T. L., and Belson, S. I. (2005) Blackboard Technologies: A Vehicle to Promote Student Motivation and Learning in Physics, *Journal of STEM Education: Innovations & Research* 6.
10. Bradford, P., Porciello, M., Balkon, N., and Backus, D. (2007) The blackboard learning system: the be all and end all in educational instruction?, *Journal of Educational Technology Systems* 35, 301-314.
11. Yueh, H.-P., and Hsu, S. (2008) Designing a learning management system to support instruction, *Communications of the ACM* 51, 59-63.
12. Creasy, M. A. (2014) Hybrid Class Experiences: Flipping Mechanics Courses and Homework Feedback, In *ASEE Illinois/Indiana Section Conference*, Terre Haute, Indiana.