

研究ノート

A Preliminary Look at Online Learner Behavior: What Can the Moodle Logs Tell Us?

オンライン受講者の行動観察: ムードル・ログで何が分かるか。

Peter PARISE

概要:

本稿では、神奈川県立国際言語文化アカデミアの英語教員研修講座「英語教育アドヴァンスト研修」及び「英文ライティング添削講座」のなかで使用している、オープンソースの e-ラーニングプラットフォーム「Moodle」について、受講者の活用のしかたに関する調査報告をする。e-ラーニングに関する文献研究を踏まえ、受講者のシステムへのアクセス記録であるログ・ファイルの蓄積から、2016 年度のものを精査し、受験者が研修プログラムのかなで、どのように、どの程度、この e-ラーニングシステムを活用しているかを探る。データ・サイエンス・ライフサイクル (Dietrich, Heller, & Yang, 2015)の手法を用い、ログ・ファイルから得られたデータを分析し、受講者の活用状況を精査した。分析の結果、受講者は頻繁に e-ラーニングシステムにアクセスしたものの受講者同士のインタラクションより一方向的な情報消費のほうに興味があり、理想的な e-ラーニングの活用方法には至らない状況が観察された。今後の研究として、受講者同士のインタラクションをどのように促進するか、そして、そのことが学習にどのような効果を与えるか、という点について調査することを考えている。

Introduction

The purpose of this report is to perform a preliminary investigation of the log files obtained from the e-learning portion of the Kanagawa Prefectural Institute of Language and Culture Studies for two courses in the Teacher Training Division via a Moodle website on group1training@kanagawa-ilcs-training.com. For over 5 years the division has amassed a bulk of data from log files obtained from these courses which details the activities of learners. For this report, we will look at the 2016 cohorts for two courses: The Advanced Program and the Support Writing Course, to determine what sort of behaviors are exhibited while participants interact with the website. This paper will briefly review the literature on e-learning, Moodle, and how data science, the process of analyzing large collections of data, can be adapted to an educational context, and examine the choices participants make on the site. This report then will conduct a preliminary study that will provide a descriptive analysis of the data for both courses and based on these results directions for future research will be discussed.

Literature Review

E-learning has numerous benefits for teachers in training: improved access to

materials, providing examples of activities that do not rely on a textbook, and offer opportunities to practice and review skills in a place that goes beyond the context of the training workshop. Another benefit is that the participants of the workshops, Japanese teachers of English who serve junior and senior high schools in the Kanagawa prefecture are exposed to an entirely English reading and listening environment. An e-learning environment can also enable teachers to work on the skill of digital literacy, a vital part of the development of a larger set of 21st century skills. Digital literacy is integral to several skill initiatives around the globe such as the Partnership for 21st Century Skills, Assessment and Teaching of 21st Century Skills, Center for Public Education, the Organization for Economic Co-operation and Development, and Metiri/NCREL (Greenstein, 2012, pp. 6–9). Such skills are more than just the ability to read texts on the Internet, they include: The ability to select core information; effectively utilize keywords and search strategies; manage the flow of information which come from various sources; critically evaluate and verify online resources; evaluate, compare and synthesize information from a variety of sources; appreciate the purpose and pervasiveness of media messages, and consider the effect of those messages on beliefs, behavior, and values (Greenstein, 2012, p.115). An e-learning environment provides a host of affordances that aid in learning such as anonymity, asynchronous, (i.e. communication not bound by time), mobile, connected, rapid, global, text-based, persistent (i.e. the interaction is stored for later use), and is multimodal in that messages can be offered in a variety of media. (Haythornthwaite & Andrews, 2011, pp. 14–27). In addition, participants are encouraged to collaborate online in an e-learning environment, and this is a key to survival in the 21st century (Mason & Rennie, 2008;Rudd, Sutch, & Facer, 2006). This emphasis on collaboration is due to shifts in technology from a consumption of information model of Web 1.0 to a Web 2.0 infrastructure that facilitates social media type interfaces that encourage interaction (Dudeney & Hockly, 2012, p. 538).

This approach to learning is not without its challenges and that is ensuring that the participants in an online course are engaged in the process. How is this construct defined in regards to e-learning? Engagement is the “learners time on task and willingness to participate in activities as well as contribute to the learning of others” (Beer, Clark, & Jones 2010, p. 76) and this is not necessarily a novel idea, there are numerous ways for naming engagement in the conventional classroom such as *active learning* (Conrad & Donaldson, 2011, p. 1). Most e-learning literature has focused on how to facilitate greater engagement by providing suggestions on how to structure online activities and promote interaction between participants (Boettcher & Conrad, 2010; Conrad & Donaldson, 2011; Ko & Rossen, 2014). While the development of the syllabus is one factor in insuring engagement in e-learning, the program used to manage the site is also an important consideration.

Moodle

With the rise of Web 2.0 as mentioned above, the technology for facilitating interaction and engagement is included in the development of Moodle as an e-learning platform. Moodle is an open source learning management system (LMS) which “is designed to provide educators, administrators and learners with a single robust, secure and integrated system to create personalised learning environments” (“About Moodle,” 2014). The software is freely available for use and the intention of its development was to promote constructivist learning online (Dougiamas & Taylor, 2001, 2003). The initial research question that prompted the development of this software was to ask:

How can internet software successfully support social constructionist epistemologies of teaching and learning? More specifically, what web structures and interfaces encourage or hinder participants’ engagement in reflective dialogue within a community of learners - by reading openly, reflecting critically and writing constructively in a way that engages their personal experiences? (Dougiamas & Taylor, 2003, p. 1)

To answer such questions Dougiamas and Taylor provided Moodle with three features: improving skills for using the Internet for distance learning, providing teachers with skill development by making the software open source and from that open source base enable Moodle to be improved with the software contributions from this community. It is these aspects that allow the software to expand to become a major LMS with about 55,031 registered sites throughout the world in 232 countries (“Moodle Statistics,” 2014). While its initial development was for general education, it can be used to promote language learning. In comparison with other LMS, a teacher using Moodle can set up quizzes with ease in a variety of formats such as True/False, Multiple Choice, Matching, Description, and Cloze as well as provide learners feedback on their performance. In addition, Moodle can allow the teacher to author interactive activities as well as upload mp3 sound files for listening tasks (Robb, 2004).

Moodle and Data Science

With this overwhelming support for e-learning for both general and language education, determining the effectiveness of Moodle is vital. Dougiamas and Taylor describe the process of data collection for their study in the form of the qualitative approach of Guba and Lincoln (1989) by incorporating a variety of data sets produced from eight participants that include questionnaire responses, journal entries and “20,000 log entries were recorded (each entry denoting an ‘action’ taken by a participant)” (2003, p. 5). It is worth noting that the log

A Preliminary Look at Online Learner Behavior

entries are mentioned here as evidence of the behavior the participants exhibit when logged into the Moodle. Log files are obtained from the Moodle website itself and detail what sort of actions are taken by a participant over time. The data can be downloaded as files in Microsoft excel (.xlsx), open document (.ods), or comma separated values (.csv). Figure 1 shows the layout of a typical log file and how the data is organized: the leftmost column is the date and time of access. Each row of the log file measures the amount of time a participant is on the Moodle minute by minute. The next two columns which are hidden are the names of who is logged in and who is effected by the actions of that participant, for example if they send a message, or check the profile page of another participant or instructor. The final three columns hold the data of interest. The middle two columns show us the context, classified as the “event.context”, i.e. where in the site was the action performed. The rightmost column shows what the participant has done on the Moodle site classified as the “event.name”. The data from this column will be the focus of this analysis.

71	29/12/16, 23:33			Course: Support Writing Course A Spring 2016	System	Course viewed
72	29/12/16, 23:32			Course: Support Writing Course A Spring 2016	System	Course viewed
73	28/12/16, 11:35			Course: Support Writing Course A Spring 2016	System	Course viewed
74	28/12/16, 11:35			Quiz: Common Sentence Errors	Quiz	Quiz attempt reviewed
75	28/12/16, 11:34			Quiz: Common Sentence Errors	Quiz	Quiz attempt reviewed
76	28/12/16, 11:33			Quiz: Common Sentence Errors	Quiz	Quiz attempt reviewed
77	28/12/16, 11:33			Quiz: Common Sentence Errors	Quiz	Quiz attempt summary viewed
78	28/12/16, 11:32			Quiz: Common Sentence Errors	Quiz	Quiz attempt viewed
79	28/12/16, 11:28			Quiz: Common Sentence Errors	Quiz	Quiz attempt viewed
80	28/12/16, 11:24			Quiz: Common Sentence Errors	Quiz	Quiz attempt viewed
81	28/12/16, 11:22			Quiz: Common Sentence Errors	Quiz	Quiz attempt viewed
82	28/12/16, 11:19			Quiz: Common Sentence Errors	Quiz	Quiz attempt viewed
83	28/12/16, 11:17			Quiz: Common Sentence Errors	Quiz	Quiz attempt viewed
84	28/12/16, 11:14			Quiz: Common Sentence Errors	Quiz	Quiz attempt viewed
85	28/12/16, 11:01			Quiz: Common Sentence Errors	Quiz	Quiz attempt viewed
86	28/12/16, 10:59			Quiz: Common Sentence Errors	Quiz	Quiz attempt viewed
87	28/12/16, 10:59			Quiz: Common Sentence Errors	Quiz	Quiz attempt preview started
88	28/12/16, 10:58			Quiz: Common Sentence Errors	Quiz	Course module viewed
89	28/12/16, 10:58			URL: Transition Signal Learning	URL	Course module viewed
90	28/12/16, 10:57			Course: Support Writing Course A Spring 2016	System	Course viewed
91	20/12/16, 11:45			Quiz: Punctuation Quiz	Quiz	Course module viewed
92	20/12/16, 11:44			Quiz: Punctuation Quiz	Quiz	Quiz attempt viewed
93	20/12/16, 11:44			Quiz: Punctuation Quiz	Quiz	Quiz attempt preview started
94	20/12/16, 11:44			Quiz: Punctuation Quiz	Quiz	Course module viewed
95	20/12/16, 11:43			Course: Support Writing Course A Spring 2016	System	Course viewed
96	16/12/16, 23:12			Course: Support Writing Course A Spring 2016	Logs	Log report viewed
97	22/11/16, 09:30			Course: Support Writing Course A Spring 2016	System	Course updated
98	19/11/16, 11:48			Assignment: Submit Assignment Three	Assignment	The status of the submission f
99	19/11/16, 11:48			Course: Support Writing Course A Spring 2016	System	Course viewed
00	19/11/16, 11:48			Assignment: Submit Assignment Three	Assignment	The status of the submission f
01	19/11/16, 11:48			Course: Support Writing Course A Spring 2016	System	Course viewed
02	19/11/16, 11:47			Assignment: Submit Assignment Four	Assignment	The status of the submission f
03	19/11/16, 11:47			Course: Support Writing Course A Spring 2016	System	Course viewed
04	19/11/16, 11:39			Assignment: Submit Assignment Five	Assignment	The status of the submission f
05	19/11/16, 11:38			Course: Support Writing Course A Spring 2016	System	Course viewed

Figure 1. A sample log file.

Other information, such as I.P. addresses are located in columns further right. These are not shown in Figure 1. to protect the anonymity of the participants.

These logs obtained from an LMS such as Moodle can be examined by a quantitative approach of analysis called: *data science* and *educational data mining*. Data science is defined as an interdisciplinary field which extracts data from various sources in a structured or unstructured state (“Data science,” 2016). What makes data science stand out from business analytics is its exploratory nature rather than provide only explanations (Dietrich, Heller, & Yang, 2015, pp. 12-13). Educational data mining (EDM) can be described as the “big data”

analysis of educational data. For example, students’ interaction with an institution’s LMS is examined to find patterns and this process is used to identify at risk students and improve student retention on a more departmental or institutional level (Huebner, 2013, pp. 3-4).

While the description of data science and educational data mining are exploratory in nature, the focus must be on the science part of the equation. In other words, a prescribed process is necessary to conduct a principled, hypothesis-driven analysis. Dietrich et al. (2015) provides one comprehensive model for data science that follows a circular analytical procedure called the *Data Analytics Lifecycle*. This cycle is a model for the collection, processing, and analysis of large data sets for business contexts. This concept though can be adapted to an educational context, for example the Institute’s more modest set of data. The Data Analytics Lifecycle consists of six phases: *Discovery*, *Data Preparation*, *Model Planning*, *Model Building*, *Communicate Results*, and *Operationalize*. These phases can be a linear process or if necessary, the cycle returns to the Discovery phase and the process continues. The description of these phases is shown in Table 1. The focus of this report will be only on the first three steps: Discovery, Data Preparation, and Model Planning due to the limited scope of this paper. Discovery and Data Preparation will be discussed in the Methods section. Before taking on a larger analysis, it is best to understand what the data set consists of by looking at the descriptive results first. The Model Planning phase will be addressed in the Discussion section.

Table 1. Description of the Data Science Lifecycle based on Dietrich, Heller, and Yang (2015, pp. 29-30)

Phase	Description
Discovery	Understanding the context and any problems in an organization, division, or even a single course. Initial development of hypotheses. The beginning of “learning” the data, which means to consider the details about the data.
Data Preparation	Creation of an “analytic sandbox” which is a space to process and investigate the data for statistical model building.
Model Planning	Further exploration of the data to determine appropriate variables for a statistical model.
Model Building	Datasets are developed for testing, training, and production purposes of the statistical model. The results at this point lead the researcher or team of researchers to determine if the tools available are sufficient to run the model.

Communicate Results	The results are communicated to major stakeholders (administrators, teachers, students) of the success or failure of the analysis and summarize the results.
Operationalize	Final reports are given and if necessary a further pilot project can also be implemented, thereby returning to the Discovery phase to repeat the process.

Through the utilization of this framework, large portions of data can be organized, analyzed, and interpreted in a principled manner. Having a clear interpretation enables the instructors to make the right pedagogical choices where it concerns e-learning.

Research Questions

Based on the prior discussion in the literature review, this report asks the following research questions:

- How do the participants in the Advanced Program and the Support Writing Course in terms of their behavior manifest engagement and interaction on the Moodle site?
- What are the differences and similarities of the two groups?
- What kind of insights or questions emerge from the first three stages of the Data Science Lifecycle that can lead to a deeper analysis for future research?

Method

Course and Participants

The Advanced Program is described as a yearlong training program in which 13-25 Japanese English teachers from public schools in Kanagawa take up an action research project with the goal of becoming expert educators who can take the initiative in promoting effective and communicative English teaching in their respective schools through enhancing their expertise in English, English education methods, multicultural understanding, and critical thinking. The process involves two observations by the Teacher Training Division staff, one in May before receiving training and in November after substantial training has transpired. In addition, the program asks the participants to give a presentation in English at the end of the year and submit a written report in Japanese about their action research project. These reports are then published online on the Institute's main website. Between these two observations 9 sessions of monthly face-to-face workshops are conducted. The current 2016 cohort of participants consist of fifteen individuals from a variety of schools. Most are early in their careers as teachers and some are seasoned teachers as well.

What is the role of Moodle for this course? The online aspect functions concurrently with these face-to-face sessions in the form of a *blended course* where part of the instruction is conducted in a physical classroom and the other is mediated by computer based instruction online (Bonk & Graham, 2006). For the Advanced program half of the interaction is in a workshop context and half is online. The idea behind the use of Moodle is to review the content of the sessions and to keep the teachers focused on the action research process. This latter part is quite important because prior participants in the Advanced program seem to have a “localized” relationship with the program. When they are in the context of the training workshop they participate, but when they return to their respective schools, there has been a noted tendency to neglect their action research. The participants are selected by their principal to attend the program, and so some might be resistant to the idea of training for a long period. In order to facilitate review, the native quiz function of Moodle is used by offering discrete item questions (i.e. multiple choice), productive tasks in English such as writing a passage about a participant’s reaction to the content of the program, and tasks which ask them to clarify their action research, for example asking the participant to write a research question in their native L1, Japanese.

The Support Writing Course in contrast focuses on writing production. The course is held in two iterations, a Spring and a Fall session with two different cohorts of participants in each. The duration of each session is 14 weeks. The max number of each cohort is ten participants. This year the total number of participants of both cohorts is 12 members, with 7 in the Spring and 5 in the Fall. One main difference between the Advanced Program and the Support Writing Course is that it admits both junior high and high school English teachers on a volunteer basis. The purpose of this course is to provide opportunities for in-service teachers to improve their writing skills. Part of this mission is to teach the basics of academic essay writing as well as build confidence. Most participants have learned English by the grammar translation method where the focus is more on knowledge and preparation for exams rather than on improving writing skills. This course provides that lost opportunity for them. In addition to the improvement of writing skills, it is also believed with an increased confidence in writing, teachers would adopt writing activities in their junior high and high school English classes.

Like the Advanced program described earlier, the Support Writing Course is an example of a blended course, except the ratio of face-to-face meetings is 1/3 and online interaction 2/3 of the course time. There are only two face-to-face class meetings, one at the beginning as an orientation and one at the end as way to reflect on the course. All the other activity in-between is conducted entirely on the Moodle. On this site the students view and download content related to the lesson, take quizzes related to that content, and upload their written assignments. There are multiple benefits of this approach to instruction over a conventional class. One such benefit is the fact that teachers have multiple responsibilities to

their own work as teachers along with family commitments as well. A blended course offers more convenience for this kind of learner along with other benefits such as increased access and flexibility, cost effectiveness, and greater self-regulation (Bonk & Graham, 2006, pp. 8–10). A blended online course allows for asymmetric communication, in that learners can access, and respond to the online course in conjunction with their schedule, which is ideal for the teacher who is managing multiple responsibilities.

Data Collection and Analysis

Discovery phase.

As referred in Table 1 above, the first phase of this process involves understanding the context of investigation, collecting and “learning the data” (Dietrich et al., 2015, p. 12). This phase is also a time to develop a hypothesis about the context as well. Since the previous section has adequately explained the rationale of the two courses: The Advanced Program and the Support Writing Program, there is no need to reiterate it again here. In terms of hypothesis making, our interest is how and if participants manifest engagement and interaction on Moodle, but we are also open to other observations since the spirit of the discovery phase is to explore the data. What is necessary is to obtain the log files which will be the object of this study. The log files can be downloaded from Moodle and this is done by going to Site Administration > Reports > Logs. Here the logs are downloaded as a .csv file to be processed by a variety of programs with ease.

Data preparation.

After obtaining the log files for both courses the files must be prepared for analysis and this involves both manual or automatic processes. The first step is to open each respective file via LibreOffice Calc, an open-source office suite to sort the rows to distinguish the entries between the instructors and the participants, since both are active in the Moodle during the coursework. The entries relating to the instructors are removed leaving only the rows which hold the participants. After this processing, the resulting log files consist of 2102 rows for the Advanced Program, 2398 rows for the Spring Support Writing Course, and 1617 for the Fall Support Writing Course. It is worth noting that the Fall cohort is still active as of the writing of this report and so the number of entries are limited. Once the files are prepared, in order to perform further data preparation and a descriptive analysis, R, an open source statistical package and software language (R Core Team, 2016) was utilized to extract as well as visualize the data. In this phase the data in the “events.name” column which show what actions the participants performed on Moodle will be extracted and made into a separate variable for analysis. It is at this point that the frequencies of these actions can be observed. In addition, the program can

visualize the data using the plot function of R to display these frequencies. This allows for further comparison between the courses.

Results

Following the extraction of the events from the main log files for both courses using R, we can obtain a list of actions of the participants from each course as shown in Table 2 and 3. The frequencies are listed in brackets next to each action. In Table 3, each iteration of the Support Writing Course is divided into both Spring and Fall sessions.

Table 2. The List of Events from the Log Files for the Advanced Program

Advanced Program
"Course module viewed", (764)
"Course viewed"(576)
"Quiz attempt viewed"(219)
"Quiz attempt summary viewed"(86)
"Quiz attempt reviewed"(78)
"Quiz attempt started"(59)
"Quiz attempt submitted"(55)
"User list viewed"(44)
"User graded"(30)
"Grade user report viewed"(22)
"Discussion viewed"(12)
"Recent activity viewed"(6)
"User profile viewed"(6)
"Grade overview report viewed" (5)
"Course user report viewed"(4)
"Other" (11)

Table 3. The List of Events from the Log Files for the Support Writing Program

Support Writing Program	
Spring	Fall
"A file has been uploaded."(62)	"A file has been uploaded."(52)
"A submission has been submitted."(54)	"A submission has been submitted."(48)
"Chapter viewed"(27)	"Chapter viewed"(39)
"Comment created"(4)	"Comment created"(11)
"Course module viewed"(443)	"Comment deleted"(1)
"Course searched"(1)	"Course module viewed"(323)
"Course viewed"(633)	"Course viewed"(473)
"Discussion subscription created"(4)	"Discussion subscription created"(1)
"Discussion viewed"(72)	"File"(1)
"Grade user report viewed"(1)	"Grade overview report viewed"(1)
"Post created"(5)	"Grade user report viewed"(7)
"Post updated"(2)	"Quiz attempt reviewed"(31)
"Quiz attempt reviewed"(68)	"Quiz attempt started"(29)
"Quiz attempt started"(63)	"Quiz attempt submitted"(24)
"Quiz attempt submitted"(60)	"Quiz attempt summary viewed"(27)
"Quiz attempt summary viewed"(71)	"Quiz attempt viewed"(87)
"Quiz attempt viewed"(246)	"Submission confirmation form viewed."(3)

"Some content has been posted."(7)	"Submission created."(27)
"Submission created."(35)	"Submission form viewed."(74)
"User graded"(90)	"Submission updated."(25)
"User list viewed"(2)	"User graded"(48)
"User profile viewed"(9)	"User list viewed"(10)
"User report viewed"(1)	"User profile viewed"(3)

Casual observation shows us that there are far more entries for the Support Writing Course in comparison to the Advanced Program since they were required to upload submissions for writing assignments onto the Moodle. The Advanced Program yielded 16 keywords relating to actions online in comparison to 23 keywords related to each cohort of the Support Writing Course. We can see that “viewed” and “attempt” are quite frequent in these courses. “Viewed” implies that the participants are consuming content on the Moodle site and for both courses this is necessary since material is posted to enable the participants to complete quizzes, or to complete a writing assignment. “Attempt” is another keyword that reveals that the participants act on the Moodle and this is in regards to the quiz, which is a feature common to both courses.

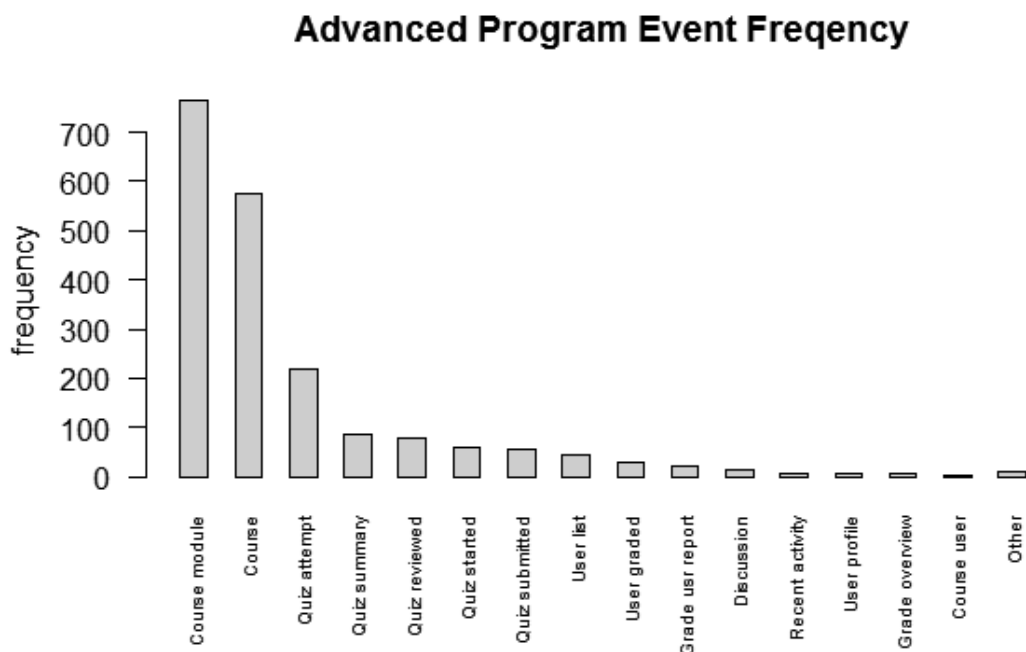


Figure 2. Bar graph of the Advanced Program events.

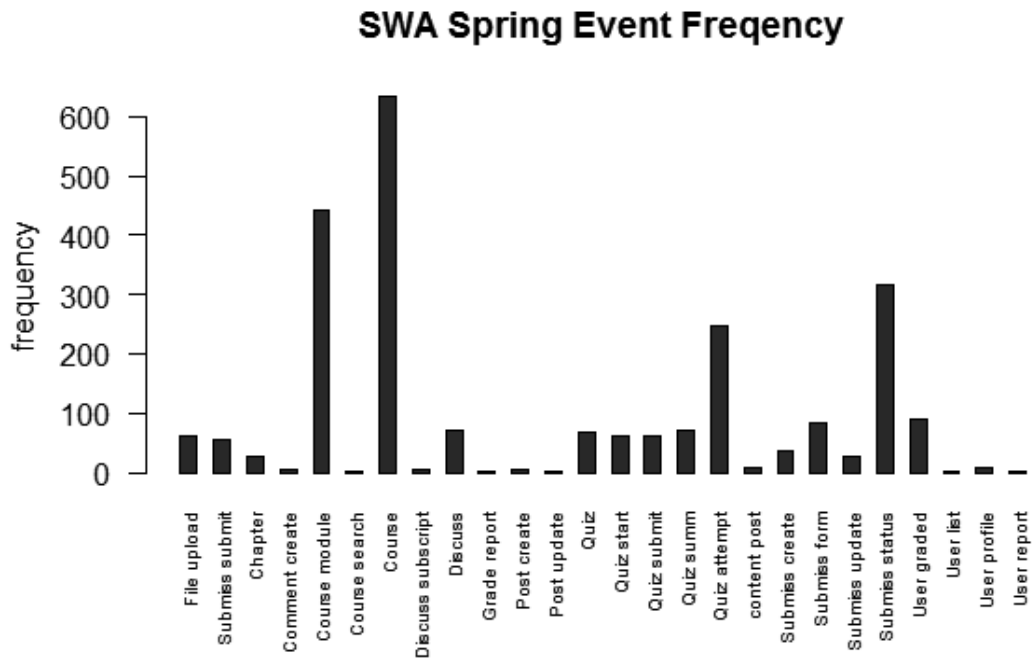


Figure 3. Bar graph of the SWA Spring events.

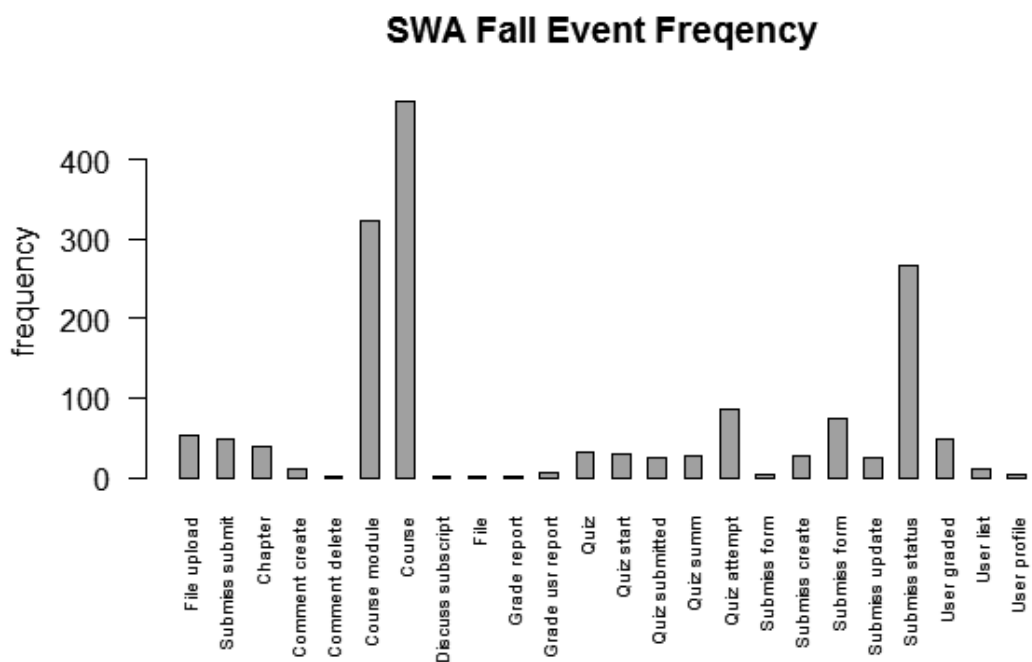


Figure 4. Bar graph of the SWA Fall events.

In Figure 2, 3, 4, we have a visual representation of the frequencies of these actions in order to make comparisons between the two courses. What is instantly recognizable is that participants view their respective courses quite frequently. Viewing the “Course module” and

the “Course” means that they are actively consuming the content of the Moodle to make use of the materials uploaded here as mentioned above. It is worth noting though that the frequencies of the Advanced Program are reversed compared to the Spring and Fall cohorts of the Support Writing Course. Another similarity is that the event called “Quiz attempt viewed” is quite frequent for both courses, yet is somewhat less for the Fall SWA in scale due to the fact that the course is still in session as of the writing of this report. This implies that the participants from both courses are invested in the results of the quizzes. In terms of the Support Writing Course in contrast with the Advanced Program, “Submission status viewed” is high for both iterations of the course, which shows that the participants are viewing the status of their writing while the instructors are checking it. One difference between the two cohorts is the “Discussion viewed” event. This is only seen in the Spring cohort but not present in the Fall.

Discussion

The results mentioned above, while descriptive in nature, allow us to make a more confident observation of how the participants use Moodle in our courses based on the plethora of data obtained from the log files. In addition, seeing how they behave on the site also informs the instructors about whether the activities set up on the Moodle are making their desired effect or not.

Addressing the first research question, which asked how the participants manifest engagement and interaction on the Moodle site, these two courses show a fair amount of engagement in terms of consuming or viewing the content of their respective courses. This is also evident from their engagement in the quiz activities as well. In the Support Writing Course, the high frequencies of viewing their submission status also show an eagerness to see the feedback from the instructors, but also the responsiveness of the participants to the Moodle. When a writing submission is checked, an email message is sent to the participant notifying them that it is ready for them to view. These high measures show that this is quite effective in garnering engagement. Interaction on the other hand is quite lacking, since the only measure available is the Spring cohort of the Support Writing Program which shows that the participants are viewing the discussion forum. What this entails is that the engagement aspect of e-learning is active, but in terms of interaction on the Moodle there is a hesitancy for the participants to participate in online interaction. Returning to Dudley and Hockley’s description of the history of computer aided language learning, Moodle is used here in a Web 1.0 rather than in the contemporary Web 2.0 manner since the participants seem more willing to consume materials by viewing, rather than interact with each other and the instructors. Further investigation into this is warranted.

The second research question asks about the similarities and differences in behavior

between the two courses. In terms of similarities, both groups of participants devote a great deal of time to engaging with the site, but there is one difference. The Advanced Program, after of almost a year of activity, view more of the course modules rather than the entire course. This is interesting in that the modules for the Advanced program are marked as “days” as the participants progress through the program. This might be facilitated by the face-to-face meetings where the participants log in to the Moodle, thereby encouraging a more module focused consumption of the site in comparison with the Support Writing Course, which tends to view the entire course over looking at just separate modules. In order to catch up on writing assignments and complete quizzes, a participant will need to look at different modules. Another difference as mentioned earlier is the noted lack of discussion in the Advanced and Fall iteration of the Support Writing Course in comparison to the Spring. This contradicts the very purpose of Moodle which was created to enable an environment where learners collaborate.

Regarding the final question, what other insights or questions arise due to the use of the first three steps of the Data Science Lifecycle? The first steps allowed us to process the data and prepare it in a methodological way. The collection, cleaning and preparation of data are vital processes which tend to be overlooked. Failing to prepare the data can impact the flow of the research, which might require backtracking in the process. (Dietrich et al., 2015). Another benefit from utilizing this approach is that it enables us to get a broader, “birds eye view” of the data available, allowing the researchers to make more accurate decisions about the direction of the training program.

In the third step of the Data Science Lifecycle, the Model Planning phase is defined as “further exploration of the data”. At this stage, it seems best to think of two things which can facilitate this process: thinking about the limitations of this report and what can be done for future research. In terms of limitations a few can be mentioned here. One is that only one aspect of the logs have been investigated, the “event.name” column of data denoting the actions of the participants, we have ignored other sections such as “event.context” which shows where that behavior transpired. There is much more to log files than reporting a single column and more data can be extracted to provide further insights into our e-learning program. Another limitation is that this report made use of keywords and frequencies, but no attempt has been made to provide descriptive nor inferential statistics, so the data presented should be interpreted with caution. Finally, the results here are only looking at the courses as one undifferentiated whole, and it does not consider the variation between individuals in a course.

Future research with log files obtained from Moodle allow a variety of choices. One example is to inquire about the relationship between the actions of the participants and the context where they perform in Moodle. Looking into the difference between say “course viewed” and “course module viewed” as well as look at their connection with the data in the

“event.context” can offer greater clarity as to the choices participants make based on the content of the course. Another important concept behind the use of log files for research is to look at how the e-learning program influences the instruction off-line. One potential exemplar study makes use of the data of logs correlated with the grades of students to determine if there is a relationship between what is learned online and classroom performance (Casey & Gibson, 2010). For the Advanced Program, we can look at correlations between the ratings made during the classroom observations, the quality of their reports, the online quiz scores, and the logs of the Moodle to find patterns. The same can be done with the Support Writing Course. What is the relationship between the writing performance of the participants and their online activity?

Conclusion

Using the log files obtained from the Moodle website, researchers have another resource to empirically observe of how learners behave in an e-learning environment. We can know what sort of choices they make, how engaged they are and what sort of material engages them. For teachers in training, exposing them to an English only environment online helps with their skill development in language by having them perform productive tasks as with the Support Writing Course, or review vital material for improving their teaching repertoire as with the Advanced Program. In addition to these, Moodle can prepare them with digital literacy skills that are vital for the 21st century. What has been revealed in this modest study is the extent to which the learners have been engaged in their respective courses on the Moodle.

For future research it is worth looking into how interaction can be facilitated and how this contributes to learning. Based on the observations of the logs, the behavior of participants is patterned after older ways of using web based material. This form of engagement that we have viewed seems more in line with the model of “information consumption” rather than the more current “engagement through interaction” model advocated by the developers of Moodle as well as other e-learning researchers. This begs the question: Should interaction be facilitated for its own sake or should it be allowed to manifest organically? The influence of the instructor in the e-learning environment should not be underestimated since the teacher’s presence in a forum for example fosters more participation from the participants (Beer, Clark, & Jones, 2010). This is another research avenue worth investigating: Does interaction online facilitate gains in terms of language learning and methods adoption? I think one benefit that would come from facilitating interaction is the development of a community of teachers which support and encourage each other to improve their craft.

Returning to the analysis of log files, the Data Science Lifecycle has allowed us to consider only a part of the data to make some interesting observations that would have been easily overlooked. The three remaining steps: Model Building and Communicate Results allow

for greater statistical models to make deeper interpretations of the data, along with sharing those results with those who have a stake in the e-learning project. The final step: Operationalization, is where the actual statistical model becomes a way to continuously monitor and enhance the teacher training program to the point where research and pedagogy become one unified process.

References

- About Moodle. (2014, December 2). Retrieved December 2, 2014, from https://docs.moodle.org/28/en/About_Moodle
- Beer, C., Clark, K., & Jones, D. (2010). Indicators of engagement. In *Curriculum, technology & transformation for an unknown future. Proceedings Ascilite Sydney* (pp. 75–86).
- Boettcher, J. V., & Conrad, R.-M. (2010). *The online teaching survival guide: simple and practical pedagogical tips* (1st ed). San Francisco: Jossey-Bass.
- Bonk, C. J., & Graham, C. R. (2006). *The handbook of blended learning: global perspectives, local designs*. San Francisco: Pfeiffer.
- Casey, K., & Gibson, P. (2010). Mining Moodle to understand student behaviour. Presented at the International Conference on Engaging Pedagogy 2010 (ICEP10), National University of Ireland Maynooth.
- Conrad, R.-M., & Donaldson, J. A. (2011). *Engaging the online learner: Activities and resources for creative instruction, updated edition*. Jossey-Bass. Retrieved from <http://www.josseybass.com/WileyCDA/WileyTitle/productCd-1118018192.html>
- Data science. (2016, December 16). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Data_science&oldid=755121899
- Dietrich, D., Heller, B., & Yang, B. (Eds.). (2015). *Data science & big data analytics: discovering, analyzing, visualizing and presenting data*. Indianapolis, IN: Wiley.
- Dougiamas, M., & Taylor, P. (2001). Improving the effectiveness of tools for Internet based education. In *Teaching and Learning Forum*. Curtin University of Technology,. Retrieved from <https://dougiamas.com/writing/tlf2000/>
- Dougiamas, M., & Taylor, P. (2003). Moodle: Using learning communities to create an open source course management system. In *Proceedings of the EDMEDIA 2003 Conference*. Honolulu, HI.
- Dudeney, G., & Hockly, N. (2012). ICT in ELT: how did we get here and where are we going? *ELT Journal*, 66(4), 533–542. <https://doi.org/10.1093/elt/ccs050>
- Greenstein, L. (2012). *Assessing 21st century skills: a guide to evaluating mastery and authentic learning*. Thousand Oaks: Corwin Press.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, California: Sage.
- Haythornthwaite, C. A., & Andrews, R. (2011). *E-learning theory and practice*. Los Angeles: Sage.
- Huebner, R. A. (2013). A survey of educational data-mining research. *Research in Higher Education Journal*, 19.

- Ko, S. S., & Rossen, S. (2014). *Teaching online: a practical guide* (4th ed). New York, NY: Routledge.
- Mason, R., & Rennie, F. (2008). *E-learning and social networking handbook: resources for higher education*. New York: Routledge.
- Moodle Statistics. (2014, December 2). Retrieved December 2, 2014, from <http://moodle.net/stats/>
- R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Robb, T. (2004). Moodle: A virtual learning environment for the rest of us. Retrieved from <http://www.tesl-ej.org/wordpress/issues/volume8/ej30/ej30m2/>
- Rudd, T., Sutch, D., & Facer, K. (2006). Opening education: Towards new learning networks. *Futurelab* (Vol. 1). Bristol, UK: Futurelab.