Kelly, Nick, Montenegro, Maximiliano, Gonzalez, Carlos, Clasing, Paula, Sandoval, Augusto, Jara, Magdalena, Saurina, Elvira, & Alarcon, Rosa
(2017)
Combining event- and variable-centred approaches to institution-facing learning analytics at the unit of study level.
International Journal of Information and Learning Technology, *34*(1), pp. 63-78.

https://doi.org/10.1108/IJILT-07-2016-0022

Combining event- and variable-centred approaches to institution-facing learning analytics at the unit of study level

Nick Kelly[1], Maximiliano Montenegro[2], Carlos Gonzalez[2], Paula Clasing[2], Augusto Sandoval[3], Magdalena Jara[2], Elvira Saurina[4], Rosa Alarcón[3]

[1]*Science and Engineering Faculty, Queensland University of Technology; and Australian Digital Futures Institute, University of Southern Queensland, Australia*

[2]*Facultad de Educación, Pontificia Universidad Católica de Chile*

[3]*Facultad de Ingeniería, Pontificia Universidad Católica de Chile*

[4]*Bibliotecas UC, Pontificia Universidad Católica de Chile*

nick.kelly@qut.edu.au

Target: IJILT international journal of information and learning technology

http://mc.manuscriptcentral.com/ijilt

http://www.emeraldgrouppublishing.com/products/journals/author_guidelines.htm?id=ijilt

Maximum 6,000 words length.

Abstract

250 word structured abstract is required with headings of:

Purpose (mandatory)

The paper demonstrates the utility of combining event-centred and variable-centred approaches when analysing big data for higher education institutions. It uses a large, university-wide dataset to demonstrate the methodology for this analysis by case study. It presents empirical findings about relationships between student behaviours in a learning management system and the learning outcomes of students, and further explores these findings using process modelling techniques.

Design/methodology/approach (mandatory)

The paper describes a two-year study in a Chilean university, using *big data* from a learning management system and from the central university database of student results and demographics. Descriptive statistics of LMS use in different years presents an overall picture of student use of the system. Process mining is described as an *event-centred* approach to give a deeper level of understanding of these findings.

Findings (mandatory)

The study found evidence to support the idea that instructors do not strongly influence student use of an LMS. It replicates existing studies to show that higher performing students use an LMS differently to lower performing students.

Research limitations/implications

The study is limited by its institutional context, its two-year time frame, and its exploratory mode of investigation to create a case study.

Practical implications

The paper is useful for institutions in developing methodology for using big data from a learning management system to make use of event-centred approaches.

Originality/value (mandatory)

The paper is valuable in replicating and extending recent studies using event-centred approaches to analysis of learning data. The study here is a larger scale than existing studies (using a university-wide dataset), in a novel context (Latin America), that provides a clear description for how and why the methodology should inform institutional approaches.

1 Introduction

The paradigm of learning analytics (LA) has become widely adopted by higher education institutions since its success in utilising available data to optimize learning, teaching and the environments in which they occur (Siemens & Long, 2011). This success has been possible due to the nature of the *big data* available from learning management systems (LMSs), which have the characteristics of being large in volume (many data points about students), high in velocity (good coverage over time), diverse in variety (different types of data) and exhaustive in scope (covering the entire population) (Kitchin, 2013). The use of LA within higher education has progressed from early attempts to predict student outcomes and 'at-risk' students using big data (Campbell, DeBlois, & Oblinger, 2007; Macfadyen & Dawson, 2010; Wang & Newlin, 2002) to becoming integrated within a number of systems (e.g. Course Signals, Blackboard Retention Centre, Insights). LA can assist institutions in understanding how students are making use of LMSs for learning (Arnold & Pistilli, 2012; Chamizo-Gonzalez, Cano-Montero, Urquia-Grande, & Muñoz-Colomina, 2015; Merceron, Blikstein, & Siemens, 2016).

In recent years however there has been recognition of the limitations of what has been labelled as the *variable-centred* approach to LA, as distinct from an *event-centred* approach (Reimann, 2009). In a variable-centred approach the focus is upon the identification of independent and dependent variables and an understanding of their relationships (analysis of variance). In contrast, an event-centred approach considers the relationship between events directly and the patterns occurring in such sequences (stochastic modelling). Whilst the two approaches are compatible (Reimann, 2009), the distinction serves to highlight the potential for higher education institutions to make greater use of process modelling to enhance their insights into student learning (Bannert, Reimann, & Sonnenberg, 2014; Reimann, Markauskaite, & Bannert, 2014).

This paper describes a study that combines these two approaches. It uses LA data within an institution to analyse, at the level of a *unit of study*, the behaviours that are occurring within a LMS. The arguments for combining event- and variable-centred approaches have been made previously (Reimann, 2009), however there is a lack of examples that extend this theoretical work to show how it can be applied to the use of big data at the level of the institution. The study described in this paper extends the work done in a number of existing small-scale studies that use the *process mining* approach (Bannert et al., 2014; Kapur, 2011; Wise, Zhao, Hausknecht, & Chiu, 2013). In addition to making an empirical contribution to understanding a Latin American university, we aim to make a methodological contribution to show how institutions can combine variable- and event-centred approaches to understand more deeply how students are making use of the LMS provided by the institution. The paper uses an exploratory approach to address questions of: What change is occurring from year to year in the use of the LMS within the university? What understanding of the basis for these changes can be found within the data? Does LMS use have a relationship to student performance and if so what can be understood from this?

These questions are addressed through a case study, using data from a university in Chile. The data used in this study are a combination of those from the LMS and from a central university database (containing demographics and unit of study results). The data span a period of two years. In answering these questions, the paper aims to make a contribution by: (i) instantiating (with a large dataset) a number of claims that have been made in recent years about the rationale for adoption of an event-based account of data (Reimann, 2009; Reimann et al., 2014); (ii) serving as a case-study for practitioners in LA who aim to understand their learning analytics data on a level beyond variable-centred approaches; and finally (iii) giving an image of the ways in which students in a high ranking university in Latin America are utilising an LMS (Sakai) at the level of courses and possible explanations for this variance.

2 Theoretical foundations

Learning analytics studies are typically categorised based upon who the findings will be utilised, as variously institution-, academic- or learner-facing analytics (van Barneveld, Arnold, & Campbell, 2012). This study is concerned with methods for producing useful *institutional analytics*, that aim to inform institutional policy in supporting student success and to improve business models (Oakleaf, 2016). For example, institutions often make use of analytics for measuring student retention and student success (Campbell et al., 2007; Hrabowski III, Suess, & Fritz, 2011). There has been a wide uptake of such analytics within institutions. However, it is common for institutions to solely make use of variable-centred approaches to analytics and to neglect event-centred approaches, even where they may be of high value (Reimann et al., 2014).

2.1 Variable-centred approach

The variable-centred approach is characterised as having one or a number of independent variables act upon one or a number of dependent variables (Reimann, 2009). Two approaches within this are the use of mining techniques to find significant correlations between variables (exploration) and selection of variables based upon semantic constructs and precedents (application). A paradigmatic case of the former consists of using mining techniques to uncover correlations between variables related to students at the course level and their final results within big data (Romero, Espejo, Zafra, Romero, & Ventura, 2013).

In an example variable-centred analytics, it has been well-established that indicators of student performance can be found within a LMS for early instructor intervention (Campbell et al., 2007; Carceller, Dawson, & Lockyer, 2013; Macfadyen & Dawson, 2010; Wang & Newlin, 2002). This approach has become a part of institutional practise to the extent that a number of software packages (e.g. Course Signals) are integrated into the LMSs of universities to make use of data (demographics, current progress within the course from testing, use of the LMS relative to peers and prior academic history) to detect at-risk students (Arnold & Pistilli, 2012).

2.2 Event-centred approaches

In contrast, event-centred approaches directly study the sequences of events that occur during student learning (Reimann, 2009). Event-centred approaches typically consider the probability of transitioning from one state to another through *stochastic models* (Bartholomew & Bartholomew, 1967), or provide a holistic view of the states that make up a process through *sequential process mining* (Reimann & Yacef, 2013; Trčka, Pechenizkiy, & van der Aalst, 2010). Process mining, the technique adopted in this paper, is useful when "event sequences can be meaningfully conceptualized as being generated by a process (or a number of processes) with internal structure" and where "processes are seen as being more than chains of atomistic events" (Reimann & Yacef, 2013, p.478). Variable- and event-centred approaches augment each other to contribute to an understanding of the significance of data.

Institutions are also not currently utilising unit of study data to the extent that they could be, i.e. they are not making comparisons from unit to unit, despite recognition that this is the level at which students can most benefit from use of LMS data, as highlighted by Dawson et al (2010):

> Our findings suggest that for the purposes of monitoring student activity and achievement, predictive models must be developed at the [unit of study] level. Furthermore, future developments of any evaluative and data visualization resources must be highly customizable to cater to instructor differences for adopting LMS tools and their overarching pedagogical intent. (p. 598)

There is thus a rationale for both disaggregating variable-centred data to consider separate units of study, as well as a rationale for looking directly at the sequences of events that occur during learning. The case study presented in this paper shows how this can be achieved.

2.3 Limitations of using LMS data

There are often significant limitations to the ability to make any claims about student learning or behaviour based upon LMS data (Reimann et al., 2014). Firstly, the data about learning within the institution are often not at a sufficiently detailed level of granularity. Whilst there may be many thousands of data points per students, the picture attainable from these data is greatly impoverished compared to, say, a video of that same students' activities. Secondly, with LMS data there is an inability to trace the learning that takes place outside the institution and occurs in the open world. Significant student learning typically takes place outside of the LMS, both in other activities on a computer and out in the physical realm. Due to both of these factors, patterns found within data are often at best to be treated as mere hypotheses resulting from exploration, and it is important that they be understood in greater depth prior to any use as a guide for changes to actions by the institution. The crucial step to link patterns in data to practical application is that of theory:

> Patterns of regular event sequences are only conceptually interesting if they are sufficiently explained in theoretical terms. Moving from exploration straight to normative guidelines without having a real explanatory account is risky and hard to justify. (Reimann et al., 2014, p. 538)

The study presented in this paper is subject to these limitations, hence the presentation as a case study rather than as applicable findings. However, through engagement with the theory in light of results, there is potential for further research and future real-world application.

2.4 Motivation for a case study

The hypothesis explored in the case study is that the combination of variable- and event-centred analysis of big data can yield valuable institutional analytics, following Bannert et al. (2014). A case study is presented here to provide an instantiation of this methodology with a large data set, with sufficient description of the methodology and outcomes for others to make use of a similar process.

A learning analytics investigation begins with specific questions to be explored, with methods tailored to suit those questions (Jones, Beer, & Clark, 2013; Kelly, Thompson, & Yeoman, In Press). The IRAC framework (information, representation, affordance, change) guided the investigation, with this framework being used over others (e.g., Greller & Drachsler, 2012) as it emphasises the potential for learning analytics to lead to change in an institution (Jones et al., 2013).

The case study began with a question, to understand the way that the institution's recently implemented LMS was being used and ways that this could be influenced. The *information* for this came from the LMS and a central repository. There was a large volume of data. The guide for what to look for within these data came from considering the affordances – what can the institution actually change? Thus there was a focus on two areas: (i) the effect of instructor upon LMS usage; and (ii) the impact of LMS usage on each of higher performing (HP) and lower performing students (LP). Taken together these areas can contribute to an understanding of how an instructor influences student use of an LMS, and whether adopting an LMS is potentially contributing to widening the student achievement gap.

2.4.1 Background for the case study

Students change their usage of a LMS over time. One explanation for this is that they are becoming accustomed to learning with an LMS and changing their usage as this occurs. Another, compatible explanation, is that this effect may be mediated by the behaviours of instructors (lecturers) within a unit of study as they change their usage of the LMS. Whilst the importance of instructors for LMS usage by students is well recognised (Al-Busaidi & Al-Shihi, 2010; Klobas & McGill, 2010; Lonn & Teasley, 2009), there has not been significant study into the effects of instructor behaviours upon students. There is thus a motivation for enquiry into the effects that instructors are having upon student usage of the LMS. The institution may be able to develop policy if indeed instructors are having a strong impact.

Further, there is significant empirical evidence that certain behaviours in an LMS, such as time of first log-in, can be associated with student outcomes (e.g., Carceller et al., 2013; Macfadyen & Dawson, 2010; Romero et al., 2013). However, there is a tendency for these studies to rely upon variable-centred approaches. Such analysis is able to reveal associations, but there can be a challenge in connecting findings to theory and thus to lead to a deeper understanding of student behaviour (Reimann et al., 2014).The institution expects that there will be a difference between the way that LP and HP students use the LMS; however it is interested, at the level of units of study, to enquire into hypotheses that may explain this difference.

3 Method

An empirical study was conducted in order to investigate these questions, situated at the Pontificia Universidad Católica de Chile (UC). The method utilised two phases of analysis. In the first, a *variable-centred* approach to the data was used to obtain an overview of LMS usage and to address the question of instructor influence on patterns. Further analysis was used to identify which units of study were most interesting for further analysis. An *event-centred* approach was then used by applying sequential process mining.

3.1 Data

Learning analytics studies of higher education in the literature are strongly concentrated in Europe, North America and Oceania (Ochoa, Suthers, Verbert, & Duval, 2014). It is thus significant that this study was conducted in a South American country, whilst recognising that UC may potentially have atypical results as it is ranked in the top 10 universities in the continent[1]. Institutional ethical approval for the study was obtained through UC. Data were gathered for Semester 2 (the second of two semesters) in each of 2013 and 2014. The first year of the LMS was excluded (2012) due to anomalies related to this being the year immediately following implementation within the university. Data were drawn from the university database and the log files from Sakai. These log files were interpreted by grouping each session by each user together. Some assumptions are made in making these groupings of activity, summarised by the question: when does one user's session end and another begin? In this research we used a cut-off of thirty minutes as indicating that a session had terminated as this is a commonly used value (Cooley, Mobasher, & Srivastava, 1999), i.e. if no activity has occurred for thirty minutes then it is assumed that a new session has begun.

The analysis of the data was conducted at the level of unit of study. Institutions are not currently utilising unit of study data to the extent that they could be, i.e. they are not making comparisons from unit to unit, despite recognition that this is the level at which students can most benefit from use of LMS data, as highlighted by Dawson et al (2010):

> Our findings suggest that for the purposes of monitoring student activity and achievement, predictive models must be developed at the [unit of study] level. Furthermore, future developments of any evaluative and data visualization resources must be highly customizable to cater to instructor differences for adopting LMS tools and their overarching pedagogical intent. (p. 598)

The analysis of the data makes reference to *low and high performing students.* Following Reimann et al. (2014), we defined these groups as those that respectively had a z-score of less than -1 or of greater than 1 in their overall course average.

[1] UC was ranked within the top ten of Latin American universities in both the QS (www.topuniversities.com) and Shanghai rankings (http://www.shanghairanking.com) in 2015

In the LMS data, there are many categories of student activity that are recorded in log files. These Sakai activity types were grouped into six categories, described in Table 1, of contents, information, read communications, write communications, test and personal. The full mapping from Sakai categories present within the data into these six categories is included in Appendix A. The category "personal" was excluded from the analysis as this activity is not related to LMS interactions within a unit of study; further, there were very few instances of this category within the data.

Table 1 Descriptions of codes used to group Sakai activity data

Code	*Description*
Contents	Reading unit of study details posted by the instructor, including unit readings
Info	Formal information about the course, including syllabus and news
Readcomm	Reading communications through the LMS
Writecomm	Writing communications through the LMS
Test	Access online tests/questionnaires within the LMS
Personal	Access and/or change personal information about the user *(excluded from analysis)*

There were 2,569 units of study in UC in 2013 and 2,616 in 2014 and the data cover over 20,000 students. The study had criteria to only include units of study that were: (i) present in *Semester 2* of both 2013 and 2014; and (ii) had either all the same instructors or all different instructors (as multiple instructors were possible). A scatter plot of this result set, looking at access of contents per student (Appendix B) showed that that were 2 units in 2014 that were extreme outliers, which were removed to increase the accuracy of results (Osborne & Overbay, 2004). This left 1,467 units included in the study.

3.1.1 Descriptive statistics

Table 2 shows descriptive statistics for use of the LMS of event type per students per year, showing that whilst the number of students per unit of study is relatively unchanged, the amount of student access increases for all types of event. This general trend of increased use of the LMS may be driven by students (who are more aware of it and how to use it), by lecturers (for similar reasons) or, most likely, by both.

Table 2 Descriptive statistics for LMS usage by year per student (n=1467 units of study)

	2013				2014			
	Min	Max	Mean	S.D.	Min	Max	Mean	S.D.
Num. students in unit	1	711	38.64	54.66	1	698	39.99	57.32
Event								
Content/stu	0	298.02	19.99	31.68	0	428.63	25.72	39.56
Info/stu	0	46.66	3.83	5.45	0	68.24	5.23	7.55
Readcomm/stu	0	43.94	1.51	3.41	0	75.00	1.83	4.37
Writecomm/stu	0	17.84	0.24	0.86	0	97.42	0.36	2.98
Test/stu	0	28.31	0.30	1.66	0	36.85	0.41	2.35

3.2 Analysis

3.2.1 Effect of instructors

The effect of instructors was assessed by utilising a variable-centred approach. The analysis resus upon four assumptions. Firstly, it is assumed that in each unit of study, from year to year, most of the students are different. There are clear exceptions to this, e.g. the case of students who repeat a unit of study. Secondly, it is assumed that where the *same* instructor teachers a unit from year to year their strategy in teaching the unit is likely to be more similar than where there is a *change* of instructor from year to year. Again, this is taken to be a rule of thumb – there will clearly be exceptions where instructors change their approach from year to year. Finally, it is assumed that there is little difference in the content and structure of a unit of study from year to year. Again, there are exceptions to this assumption, adding to limitations to the generalisability of findings.

Taken together, these assumptions suggest that a comparison between two groups of units: (i) those that have the same instructor in both years; and (ii) those that have a different instructor in both years. If the assumptions listed hold true then the expectation is that if instructors make a significant difference to student behaviours in the LMS then we expect that the changes from 2013 to 2014 in group (i) will be less pronounced than the changes in group (ii). This was analysed using a repeated measures ANOVA statistic (not reported here due to lack of significant findings) and descriptive statistics.

3.2.2 Difference between high and low performing students

The overall changes from year to year across the university were analysed to compare differences in LMS use between high and low performing students. This was done by firstly using descriptive statistics to identify units of interest, and then use of process mining to view event sequences within these units of study.

3.2.3 Inclusion of time

It has begun to be widely recognised that in analysing LMS data, time is important (Reimann, 2009). For example, within an event log from an LMS, one approach is to simply count the number of times an event occurs for a student; time-on-task is widely recognised to be a better indicator of the learning that is actually occurring (Bloom, 1974; Stallings, 1980). There are many ways in which to calculate time-on-task that each have different results (Kovanović et al., 2015). In this work we adopt the metric of time until transition to the next task, within a session. Process mining analysis was conducted using the software Disco.

4 Results

4.1 Variable-centred analysis of units

4.1.1 Is change driven by instructors or students?

Units of study that had the same instructor from year to year were compared with those units of study that had a different instructor. The results, Table 3, show that regardless of whether a unit had the same or a different instructor, there was an increase in use of the LMS. The results have been included for the sake of completeness, showing that this exploration revealed no significant findings. This may indicate that the instructor did not have an influence upon student use of the LMS, or it may indicate that one of the assumptions behind the question was not valid. Table 4 shows the results from a repeated measures ANOVA from year to year, with treatment of a unit having the same/different instructor. The results show no significant effects from the treatment; whilst the frequency of contents appears to be significant, the low partial η^2 indicates that it has a low effect size.

Table 3 Influence of instructor

		Same instructor				Different instructor				
		2013		2014			2013		2014	
	N	Mean	S.D.	Mean	S.D.	N	Mean	S.D.	Mean	S.D.
num stu	1225	39.22	52.47	40.61	53.89	242	35.69	64.67	36.83	72.27
contents/stu	1225	21.26	33.17	27.49	41.36	242	13.58	21.60	16.80	27.17
info/stu	1225	3.88	5.47	5.25	7.28	242	3.54	5.36	5.13	8.81
read/stu	1225	1.58	3.57	1.88	4.45	242	1.16	2.38	1.60	3.95
write/stu	1225	0.25	0.91	0.40	3.26	242	0.19	0.57	0.19	0.58
test/stu	1225	0.33	1.74	0.47	2.56	242	0.15	1.08	0.08	0.53

Table 4 Results of repeated measures ANOVA for all units of study, for each type of LMS activity, for time points of 2013ii and 2014ii and a treatment of same/different instructor for the unit of study.

	F	*P*	Pt η^2
Contents	14.450	.000	.010
Info	.030	.862	.000
Read comm	2.206	.138	.002
Write comm	.291	.590	.000
Test	2.441	.118	.002

4.1.2 Usage by high/low performing students

Table 5 shows a comparison of low performing (LP) and high performing (HP) students in each of the two semesters studied. The results show that for every type of access, HP students used the platform (on average) significantly more than LP students, the gap increasing from

2013 to 2014. For example, there was no difference in the average number of times that LP and HP students wrote to their instructor in 2013, but by 2014 the two values were 0.37 and 0.53 for a significant difference between the two, with HP students writing to their instructor more frequently. It is not possible to know from the data if this is the beginning of a trend (of a widening gap increasing in later years) or if this is an anomaly – further investigation is required.

Table 5 Change in use of LMS between years in high and low performing students

	Low performing		*High performing*	
	2013	**2014**	**2013**	**2014**
Unit count	1100	1084	971	964
Count/stu (mean)				
Contents	28.90	37.32	37.76	49.30
Info	4.58	6.50	5.55	7.99
Readcomm	2.35	2.93	2.52	3.17
Writecomm	0.23	0.37	0.23	0.53
Test	0.22	0.30	0.28	0.40

4.1.3 Units of study with significant difference in LMS use between low/high performers

The variable-centred analysis in this way allowed for identification of the units of study in which this effect was most pronounced. Table 6 shows the units that were identified for further analysis, as those that had the greatest differences between the LMS use of LP and HP students. These units were selected using criteria of: (i) units that had a large disparity in LMS use between HP and LP students in *both* semesters (defined as over 200% increase in both semesters); and (ii) where there was a non-trivial usage of the LMS in both groups (defined as having mean frequency of at least 2 for all categories in both years). Two of these units were selected for demonstration of event-centred analytics at the level of unit of study, on the basis of best meeting these two criteria.

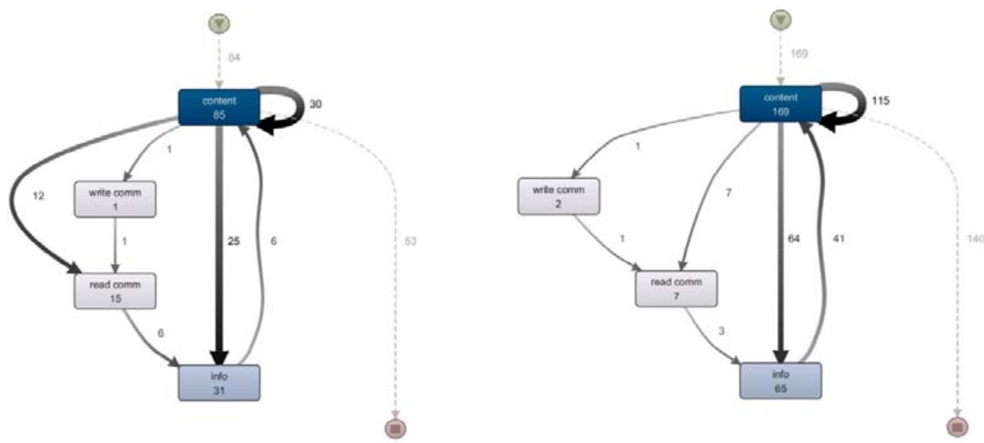
Table 6 Extract from comparison of high performing and low performing students by unit. Bold indicates units chosen for event-centred analysis.

Unit of study	*2013 L.P. content/stu*	*2013 H.P. content/stu*	*percent increase*	*2014 L.P. content/stu*	*2014 H.P. content/stu*	*percent increase*
UNITA	21.57	96	445.06%	60	121.7	202.83%
UNITB	10	42	420.00%	19	40.5	213.16%
UNITC	6.75	27	400.00%	5	20	400.00%
UNITD	**12.85**	**45.44**	**353.62%**	**17.13**	**56.53**	**330.01%**
UNITE	**6**	**22.06**	**367.67%**	**41**	**168.5**	**410.98%**
UNITF	7	23.6	337.14%	8	16.17	202.13%
UNITG	4.5	14	311.11%	6	85	1416.67%
UNITH	44.2	133.67	302.42%	73	156.5	214.38%
UNITI	11.57	34.71	300.00%	8.75	21.43	244.91%
UNITJ	11.5	33.09	287.74%	6.14	19.33	314.82%

4.2 Event-centred analysis of unit detail

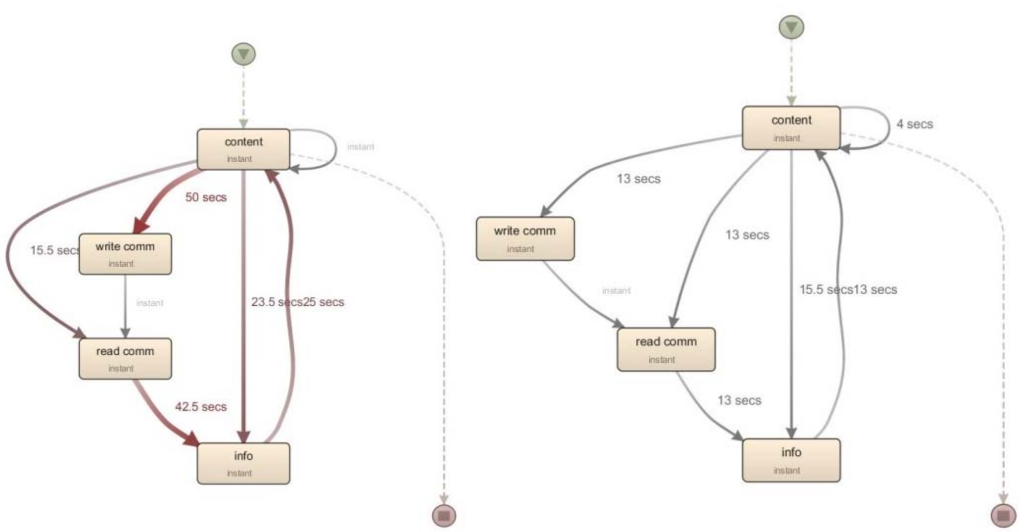
The two units selected for event-centred analysis were UNITD and UNITE. Analysis used process modelling that included both *sequence* and *time.* Figures 1 and 2 show a comparison between the process models of LP and HP students for 2013 UNITD. Figure 1(a) shows the frequency of transition between activities for low performing students and Figure 1(b) shows the same results for high performing students. Figure 2(a) and Figure 2(b) show the time taken for each transition rather than the frequency for the same LP and HP students respectively.

These results for UNITD in 2013 show that high performing students that moved from looking at *contents* to looking at *info* were far more likely to continue the session and come back to looking at *contents* than were low performing students. One possible interpretation of this kind of recursive activity is that these HP students are taking more actions to explore information on the LMS. Also, the one activity in which LP students had more activity was in reading communications. There are many possible interpretations for this, e.g. they may have returned to read communications multiple times, or the instructor may have communicated more with them to try to assist them in their studies. Also, in this unit HP students are spending far less *time* in each of the transitions on the site – in every instance the time taken (e.g. from contents to info) occurs faster.



(a) (b)

Figure 1 Process mining results for UNITD for 2013 frequency of transitions: (a) low performing students; and (b) high performing students. Numbers indicate median number of transitions per session per student, where breadth of arrow is proportional to the number. Green circle indicates start of session, red circle indicates end of session.



(a) (b)

Figure 2 Process mining results for UNITD for 2013 median time for transitions (per session on the LMS) for: (a) low performing students; and (b) high performing students.

In 2014 for UNITD, Figure 3(a) and 3(b), similar findings can be observed, in that the frequency of transitions is higher for HP students, whilst the time taken for the transitions, Figure 4(a) and 4(b) is lower. An interpretation of the process model is that HP students were more likely to directly reply to communications (direct transition from reading communications to writing communications), whilst LP students returned to contents or wrote communications in a subsequent session.

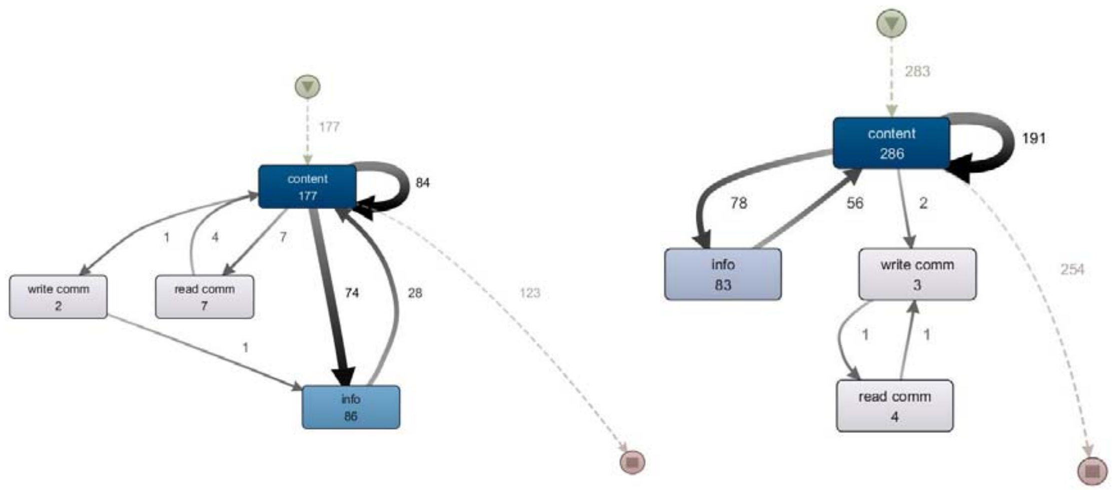


Figure 3 Process mining results for UNITD for 2014 frequency of transitions (per session on the LMS): (a) low performing students; and (b) high performing students.

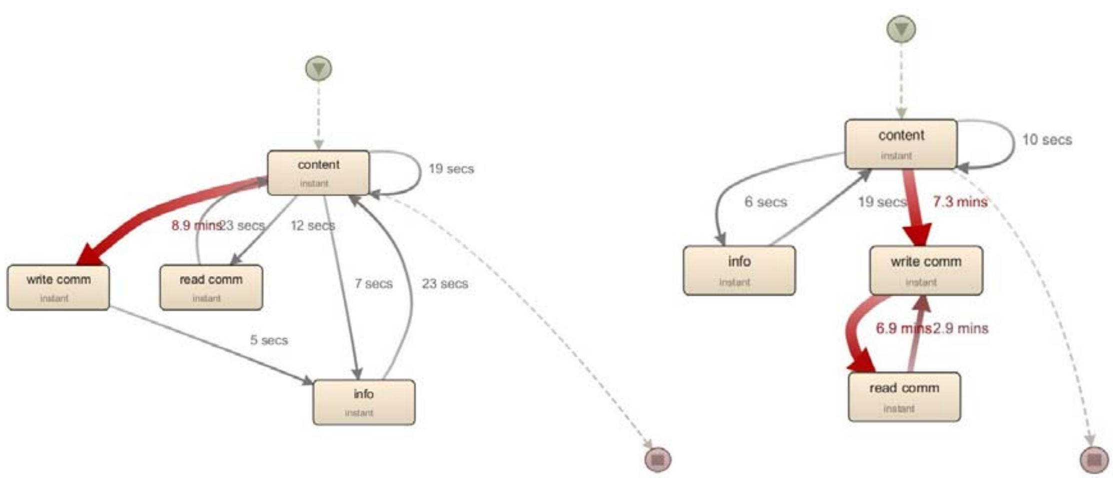


Figure 4 Process mining results for UNITD for 2014 median time for transitions (per session on the LMS) for: (a) low performing students; and (b) high performing students.

UNITE frequencies for 2013, Figure 5(a) and (b) for LP and HP respectively, show a similar and unsurprising finding that there are far higher frequencies of transitions for HP students. The process model allows for a deeper analysis, to see that there were a number of cycles in the HP students of reading content, reading communications and looking at information – such cycles are not present in the LP students. Similar to UNITD, in UNITE in 2013 HP students are faster to move from content to information. However, there is a strong difference in the 8.6 minutes taken to move from information back to content, where contrary to prior results HP students spend far more time. In 2014 LP students are faster to move from content to information.

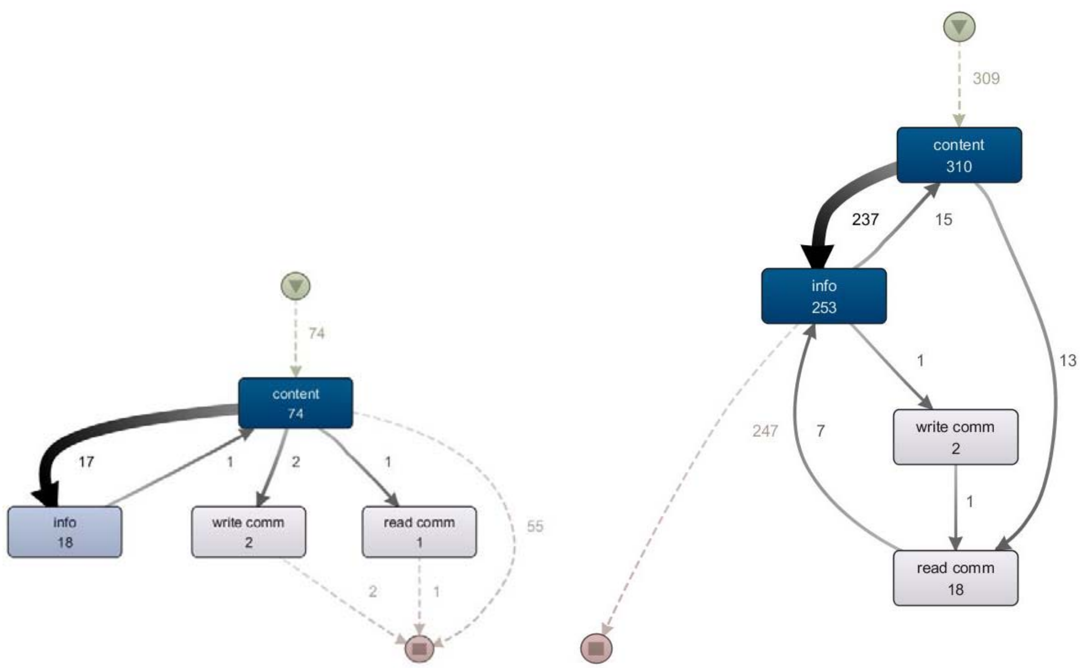


Figure 5 Process mining results for UNITE for 2013 frequency of transitions (per session on the LMS): (a) low performing students; and (b) high performing students.

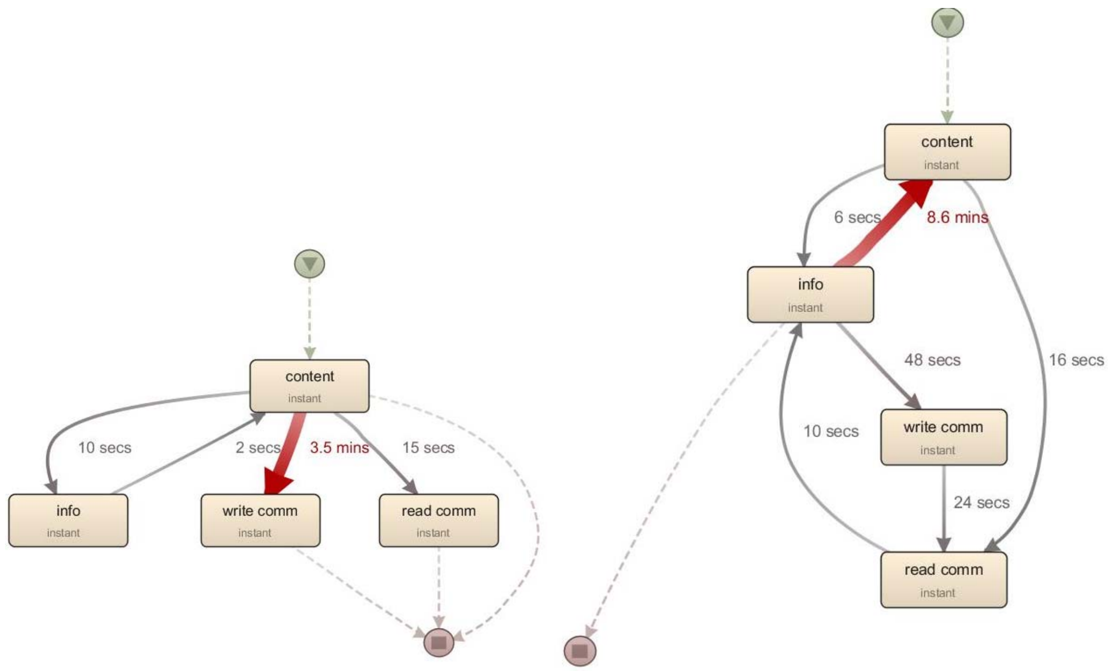


Figure 6 Process mining results for UNITE for 2013 median time for transitions (per session on the LMS) for: (a) low performing students; and (b) high performing students.

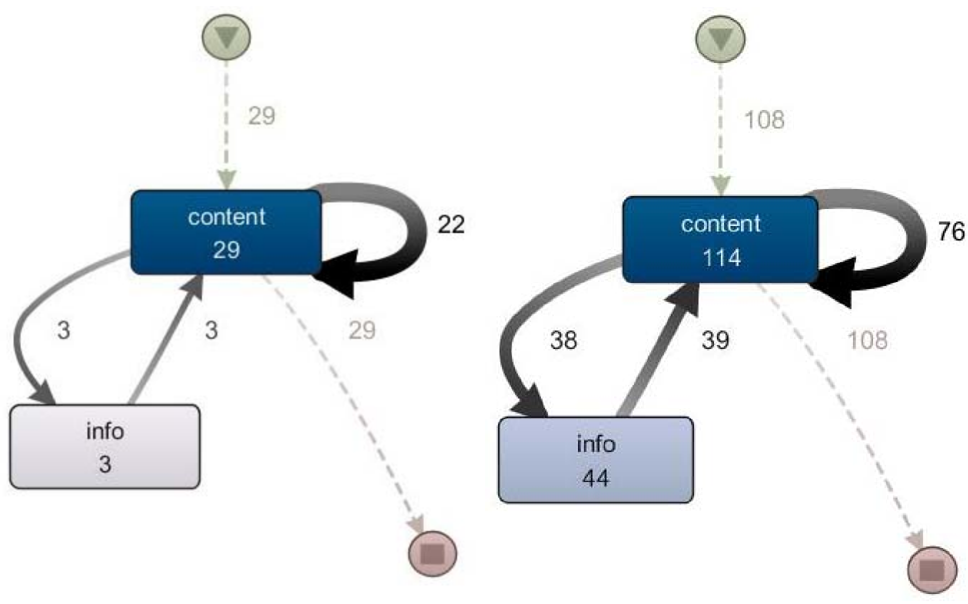


Figure 7 Process mining results for UNITE for 2014 frequency of transitions (per session on the LMS): (a) low performing students; and (b) high performing students.

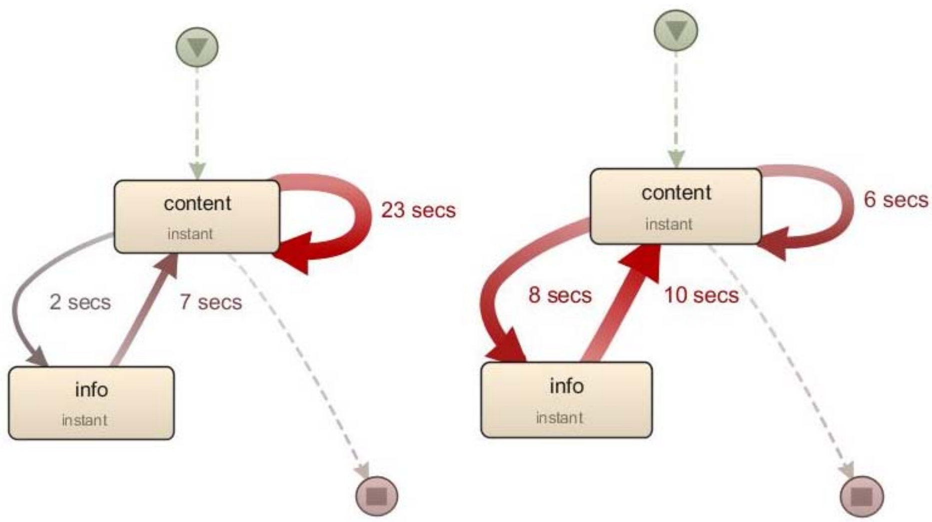


Figure 8 Process mining results for UNITE for 2014 median time for transitions (per session on the LMS) for: (a) low performing students; and (b) high performing students.

5. Discussion and conclusions

5.1 Summary of findings

The institution, UC, saw a general increase in the usage of the LMS from 2013 to 2014, and in all activity types within the LMS. An attempt was made to distinguish whether this was due to the influence of instructors, however the findings were inconclusive.

Variable-centred analysis comparing HP and LP students showed that, in general, HP students were using the LMS more than LP students, replicating prior findings. An event-centred analysis using process mining was conducted into two units of study. The results can suggest that HP students are, in general, moving faster from content to information, but occasionally spending longer reading information. Two possible interpretations for this are that these students are faster at synthesising information or these students are more comfortable with the technology, however many other interpretations are possible.

5.2 Limitations

There are significant limitations on the study that prevent these findings being used as generalizable evidence. Firstly, the context of the dataset limits the findings. The data cover just two specific years, where many years would be needed to make claims about year to year trends. The study was carried out in a specific South American university and it is likely that specific institutional factors influence results.

A second limitation is that there was no qualitative study to follow on from the event-centred suggestions. A deeper analysis of the units of study to follow up on the hypotheses, such as one involving videos of LMS use and interviews with instructors and students, would facilitate the kind of understanding required to adequately investigate these hypotheses.

5.3 Conclusions

The paper has presented a case study of the combination of event- and variable-centred approaches to institutional learning analytics, including results from a large scale learning analytics study in a Chilean university. Whilst the applicability of the findings are limited, the paper makes a methodological contribution to the field. Further, these findings include hypotheses that can lead to further investigation. It gives an image of the ways in which students in a high ranking university in Latin America are utilising an LMS (Sakai) at the level of courses and possible explanations for this variance.

Acknowledgements

The work in this paper was supported by FONDECYT grant number 1131109.

References

Al-Busaidi, K. A., & Al-Shihi, H. (2010). Instructors' Acceptance of Learning Management Systems: A Theoretical Framework. *Communications of the IBIMA, 2010*, 2010.

Arnold, K. E., & Pistilli, M. D. (2012). *Course signals at Purdue: using learning analytics to increase student success.* Paper presented at the Proceedings of the 2nd International Conference on Learning Analytics and Knowledge.

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and learning, 9*(2), 161-185.

Bartholomew, D. J., & Bartholomew, D. J. (1967). *Stochastic models for social processes*: Wiley London.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *Educause Review, 42*(4), 40.

Carceller, C., Dawson, S., & Lockyer, L. (2013). Improving academic outcomes: does participating in online discussion forums payoff? *International Journal of Technology Enhanced Learning, 5*(2), 117-132.

Chamizo-Gonzalez, J., Cano-Montero, E. I., Urquia-Grande, E., & Muñoz-Colomina, C. I. (2015). Educational data mining for improving learning outcomes in teaching accounting within higher education. *International Journal of Information and Learning Technology, 32*(5), 272-285. doi:doi:10.1108/IJILT-08-2015-0020

Cooley, R., Mobasher, B., & Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Knowledge and information systems, 1*(1), 5-32.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society, 15*(3), 42-57. http://hdl.handle.net/1820/4506 Retrieved from http://hdl.handle.net/1820/4506

Hrabowski III, F. A., Suess, J., & Fritz, J. (2011). Assessment and Analytics in Institutional Transformation. *Educause Review, 46*(5), 14.

Jones, D., Beer, C., & Clark, D. (2013). The IRAC framework: Locating the performance zone for learning analytics. *Proceedings of the 30th Australasian Society for Computers in Learning in Tertiary Education Conference (ASCILITE 2013)*, 446-450. http://eprints.usq.edu.au/24352/ Retrieved from http://eprints.usq.edu.au/24352/

Kapur, M. (2011). Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *International journal of computer-supported collaborative learning, 6*(1), 39-56.

Kelly, N., Thompson, K., & Yeoman, P. (In Press). Theory-led design of instruments and representations in learning analytics: Developing a novel tool for orchestration of online collaborative learning. *Journal of Learning Analytics*.

Kitchin, R. (2013). Big data and human geography Opportunities, challenges and risks. *Dialogues in human geography, 3*(3), 262-267.

Klobas, J. E., & McGill, T. J. (2010). The role of involvement in learning management system success. *Journal of Computing in Higher Education, 22*(2), 114-134.

Lonn, S., & Teasley, S. D. (2009). Saving time or innovating practice: Investigating perceptions and uses of Learning Management Systems. *Computers & Education, 53*(3), 686-694.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588-599.

Merceron, A., Blikstein, P., & Siemens, G. (2016). Learning Analytics: From Big Data to Meaningful Data. *Journal of Learning Analytics, 2*(3), 4-8.

Oakleaf, M. (2016). Getting Ready & Getting Started: Academic Librarian Involvement in Institutional Learning Analytics Initiatives. *The Journal of Academic Librarianship*.

Ochoa, X., Suthers, D., Verbert, K., & Duval, E. (2014). Analysis and Reflections on the Third Learning Analytics and Knowledge Conference (LAK 2013). *Journal of Learning Analytics, 1*(2), 5-22.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation, 9*(6), 1-12.

Reimann, P. (2009). Time is precious: Variable-and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 239-257. http://link.springer.com/article/10.1007/s11412-009-9070-z Retrieved from http://link.springer.com/article/10.1007/s11412-009-9070-z

Reimann, P., Markauskaite, L., & Bannert, M. (2014). e-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology, 45*(3), 528-540.

Reimann, P., & Yacef, K. (2013). Using process mining for understanding learning. *Handbook of design in educational technology*, 472-481.

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education, 21*(1), 135-146.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 30-32.

Trčka, N., Pechenizkiy, M., & van der Aalst, W. (2010). *Process mining from educational data*: Chapman & Hall/CRC.

van Barneveld, A., Arnold, K. E., & Campbell, J. P. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE learning initiative, 1*, 1-11.

Wang, A. Y., & Newlin, M. H. (2002). Predictors of performance in the virtual classroom: Identifying and helping at-risk cyber-students. *THE Journal (Technological Horizons In Education), 29*(10), 21.

Wise, A., Zhao, Y., Hausknecht, S., & Chiu, M. M. (2013). Temporal Considerations in Analyzing and Designing Online Discussions in Education: Examining Duration, Sequence. *Assessment and evaluation of time factors in online teaching and learning*, 198.

Appendix A

Table A.1 Coding for Sakai categories present within the data

code	*sakai category*		*code*	*sakai category*
content	content.read		writecomm	chat.new
content	melete.section.read		writecomm	forums.newtopic
content	webcontent.read		writecomm	forums.response
info	news.read		writecomm	mail.create
info	syllabus.read		writecomm	messages.forward
personal	prefs.add		writecomm	messages.new
personal	prefs.del		writecomm	messages.reply
personal	prefs.upd		writecomm	poll.vote
personal	profile.new		writecomm	wiki.new
personal	profile.prefs.new		writecomm	wiki.revise
personal	profile.privacy.new		test	asn.read.submission
readcomm	forums.read		test	asn.revise.assignment
readcomm	messages.read		test	asn.revise.assignmentcontent
readcomm	poll.viewResult		test	sam.assessment.revise
readcomm	wiki.read		test	sam.assessment.submit
			test	sam.assessment.take

Appendix B

Figure B.1 shows a scatter plot of the mean number of times that each student accessed the contents, for each unit of study. The chart shows the two outliers that were removed for the study.

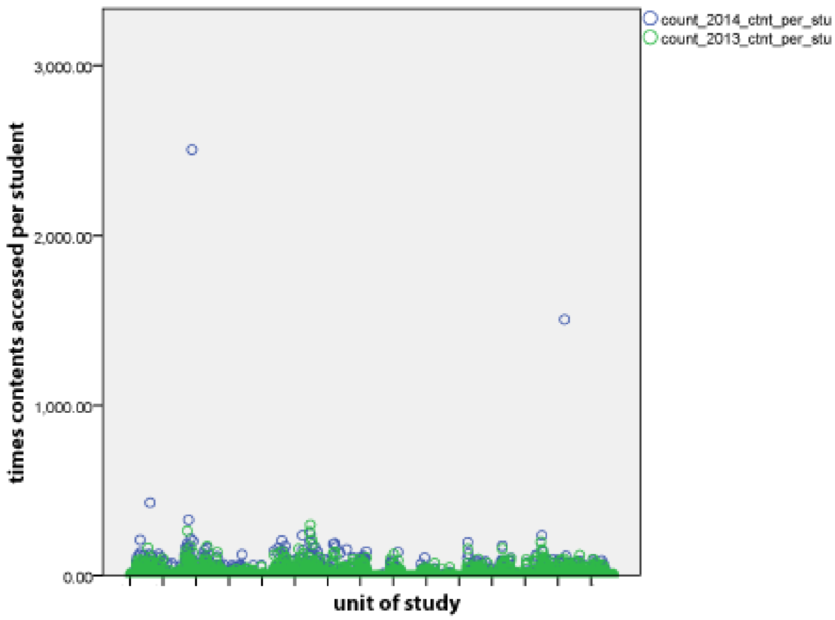


Figure B.1 Outlying units of study as determined by mean number of times students accessed contents for a unit of study