

Heathcote, Elizabeth and Dawson, Shane (2005) Data Mining for Evaluation, Benchmarking and Reflective Practice in a LMS. In *Proceedings E-Learn 2005: World conference on E-learning in corporate, government, healthcare & higher education*, Vancouver, Canada.

Data Mining for Evaluation, Benchmarking and Reflective Practice in a LMS

Liz Heathcote
Queensland University of Technology, Australia
e.heathcote@qut.edu.au

Shane Dawson
Centre for Learning Innovation
Queensland University of Technology, Australia
sp.dawson@qut.edu.au

Abstract: Large amounts of user and systems information is tracked by Learning Management Systems (LMS), affording the evaluation of online learning and teaching behaviour from a teaching unit level through to an enterprise level. At Queensland University of Technology, data from the in-house built LMS was mined over the period 2001- present for trends and usage information in order to provide a valuable perspective to evaluating systems usage, learning and teaching behaviours and benchmarking. This case study investigates the role of quantitative data in evaluating teaching and learning at various levels from unit of study to the enterprise through the establishment of benchmarks. Analysis of asynchronous discussion forum usage is discussed as an example.

Introduction

Within the Australian Higher Education system there has been a proliferation of commercial and proprietary Learning Management Systems (LMS) adopted and integrated into an institutional technical infrastructure in order to enhance and supplement the teaching and learning domain (Kitto & Higgins 2003). Coinciding with the rapid evolution of these technologies researchers have tended to adopt either a quantitative or qualitative evaluation methodology to assess the impact of LMS usage on learning objectives (Oliver 2000). Quantitative methods have focused more on usability and specific technical criteria, and qualitative methods have been predominantly implemented to evaluate the achievement of stated learning outcomes (Oliver 2000). Traditional evaluative methodologies assessing the integration of technologies with teaching practices have focused on unit-level qualitative studies to provide an indication of the impact on the learning domain. Ladyshevsky (2004) suggests that current researchers need to move beyond the quantitative and qualitative debate and develop additional methodologies to assess the impact of electronic mediums on learning and teaching. Similarly, Oliver (2000) suggests that contemporary evaluative research is in the process of shifting focus from advocating the adoption of either a quantitative or qualitative methodology to the promotion of a mixed methods approach in order to triangulate results, thereby enhancing the credibility of stated conclusions.

The integration of LMS has afforded the capacity to capture extensive amounts of quantitative data regarding individual user and designer behaviour within the online domain. Most commercial LMS have the ability to capture data about the amount and type of usage of the system such as hits, unique users, uptimes, downtimes, number of resources integrated into the site and the overall web site size. Where the system requires user login, it potentially also tracks navigational pathways through the online material, time spent in various areas of the sites and numbers of visits over a certain time period. Given the technical nature of these statistics, the captured data has been traditionally used by the information technology support staff to input into systems planning and monitoring. However, given the amount of user

information that is stored, and the advantages of tracking actual behaviour as opposed to recollections or perceived behaviour, it appeared to the authors that the integration of quantitative data derived from learning management systems may also provide a valuable perspective to evaluating learning and teaching practices and the establishment of specific organisational benchmarks. Furthermore, the initial interrogation of the quantitative data affords a focused qualitative analysis of sites of specific interest in order to verify stated conclusions.

This paper discusses a quantitative methodology adopted to mine data from Queensland University of Technology's in-house LMS to establish usage patterns and online learning designs occurring within the various organisational levels operating in the university – enterprise, faculty and unit. The paper provides specific examples to illustrate how quantitative data can be mined and interpreted by university administrators and teaching staff.

The Case Study

Queensland University of Technology (QUT), a large Australian university of approximately 40,000 students, has an in-house developed Learning Management System (LMS), called Online Learning and Teaching (OLT). This in-house LMS currently hosts approximately 5000 websites. Beginning in 2004, an investigation into the patterns of user behaviour, and examination of specific OLT-developed functions was conducted through quantitative data mining to evaluate how OLT was predominantly integrated with learning and teaching practices. Fayad (1996) describes data mining as a method for discovering patterns in databases that can be integrated with and guide the decision making process. The QUT OLT data mining project developed methods of collating and analysing data to ascertain trends occurring within the learning and teaching domain, in addition to trends used by information and technology systems and support staff to guide decisions on technical infrastructure. For example, Dawson *et al.* (2004), in discussing the requirements for the OLT data mining project, suggests that quantitative surveillance of discussion forum interactions occurring among staff and students provides an indication of potential sites demonstrating characteristics of a sense of community and of implemented collaborative learning activities. The authors maintain that the initial quantitative approach affords a more focused future qualitative analysis to verify and expand on initial findings. Similarly, Burr & Spennemann (2004) argue that the examination of patterns of user behaviour occurring in the online environment can be utilised for the purposes of evaluating both technical and pedagogical contexts through:

- supporting predictions for network and server traffic load;
- aiding the reflection of pedagogical practices in alignment with usage statistics; and
- affording the planning of software release, and maintenance of system developments during times of minimal activity and therefore potential disruption.

Method

Data was mined from two main databases, the page tracking database and the OLT system database (see Figure 1), and stored in a separate reports database which provided bi-semester snapshots of activity on the OLT system (becoming fortnightly from January 2005). The page tracking database contains information on user behaviour (such as viewing, adding, editing, or deleting content) and opted navigational paths through the website. This data permits an insight into the way learners and designers interact with the OLT system, such as who (student or teaching staff) is posting and replying to items on the discussion forum, or how long users spend in various areas of the website. Although the system captures individual user's details, only the summaries of user interactions are reported on from the data mining exercise to protect confidentiality and to simplify the results into a useable format (unit, faculty or enterprise level). The OLT system database contains information on the number and types of functions (such as chat rooms) available on the website as well as the number of users accessing the site.

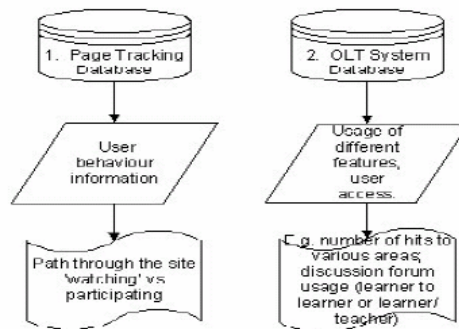


Figure 1: Data mining from usage tracking databases on OLT system (from Heathcote & Prakash: 2004)

Linking Activity Domains and LMS function types

Table 1.0 describes the model adopted for linking OLT functions to common pedagogical activities such as communication, information dissemination and assessment. The development of user/designer domains guided the codification of accessible data relating to specific OLT functionalities and the subsequent alignment with learning and teaching contexts. The systems interrogation of this data distinguishes between designer and user activity by identifying (i) the number of resources provided by the site as being a designer activity, and (ii) the number of hits/contributions made within the OLT site to or within specific resources as an artefact of user activity.

User/Designer activity domains	Examples
<i>Communication and collaboration</i> The degree to which the OLT environment is being used to support student communication and collaboration	Chat rooms Discussion forums Email archives
<i>Information dissemination</i> The degree and manner to which the OLT site is being used to deliver content.	Files, Media such as video & audio Links to a website or e-readings Text
<i>Information design/navigation</i> The manner in which the OLT site is structured	Dynamic menus/ navigation bars Pages and page styles Topics and templates
<i>Assessment</i> The degree and manner in which OLT is being used to support and manage formative and summative assessment items.	Assignment upload areas Multiple Choice, Short answer, scenario etc. Quizzes

Table 1: Alignment of OLT resources with user and designer domains

A word of caution: the limitations of data mining

A systems scan of designer and user behaviour within OLT can never describe in full how designers and users are engaging with the use of online environments for teaching and learning. Policy interventions, staff development activities and discipline culture all contribute to shaping designer and user behaviour within the online environment. Therefore, utilising a systems view to codify designer and user behaviour is 'indistinct', but can play in the refinement, ratification and benchmarking of broader evaluation strategies. This assumption is analogous to the justification of the "Flashlight" methodology of evaluation (Ehrman, 1995), which posits that by initially focusing or limiting the areas of investigation, broader trends and patterns can be identified.

In addition, the integrity of mined data for a system as large as an institutional LMS can be affected by

- technical issues such as corruption of databases; and
- change in access settings for websites during the time period (as this affects the level of user detail that is reported on).

Findings

Mined data was collated to reflect the enterprise, faculty or unit level activities occurring within the OLT system. Ongoing reportage of data is harvested on a fortnightly basis, hence the system records activity at specific timeframes to allow for future comparative analyses and benchmarking. The collation of mined data has the ability to scale upwards from the individual unit to encompass all units operating within a faculty and at an enterprise level. The interface collating mined data permits individual units to conduct ongoing comparisons with previous time periods or against other teaching units. Consequently, unit, faculty and university administrators can monitor ongoing faculty, course and unit level activity to gauge usage patterns at the various organisational levels to feed into the reportage against implemented strategic initiatives.

Overall Usage Findings

Findings at the enterprise level impact upon technology and infrastructure; planning and support; and assist in the evaluation and substantiation of university wide policies and strategic plans. For example, the “flexible

delivery” strategies begun in the late 1990s at QUT emphasized the convenience of study at the student’s own time and place, and thus can be reported in the amount of usage of the system and access timing over the week. In terms of overall access to the OLT system, figures 2 and 3 illustrate an explosive increase in usage however, little change has occurred in the pattern of user behaviour for days of the week (Figure 2) and hourly (Figure 3) over the 3 years (2001-2003). This span of access indicates achievement of strategic enterprise flexible delivery objectives in terms of span of student access outside of traditional lecture/ study periods. The observed pattern of user behaviour is analogous to that discussed by Burr (2003) in his examination of discussion forum usage by a large rural University predominantly focusing on the provision of distance education. Burr (2003) maintained that the online behaviour of distributed education students drives the subsequent pattern of designer behaviour through the timing and implementation of teaching and learning activities such as forum contributions and assessment.

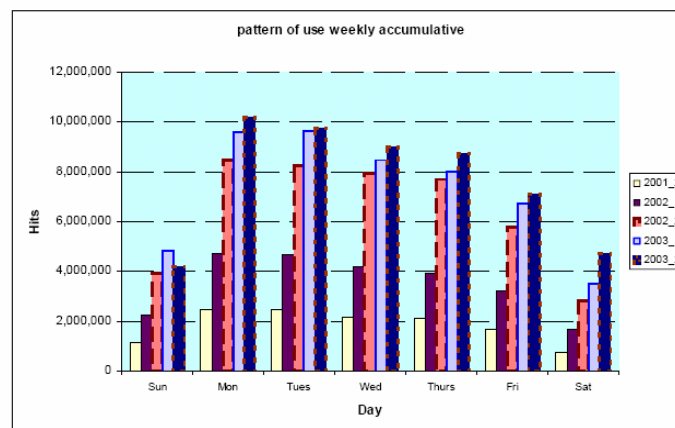


Figure 2: Pattern of weekly OLT system usage over a 3 year period.

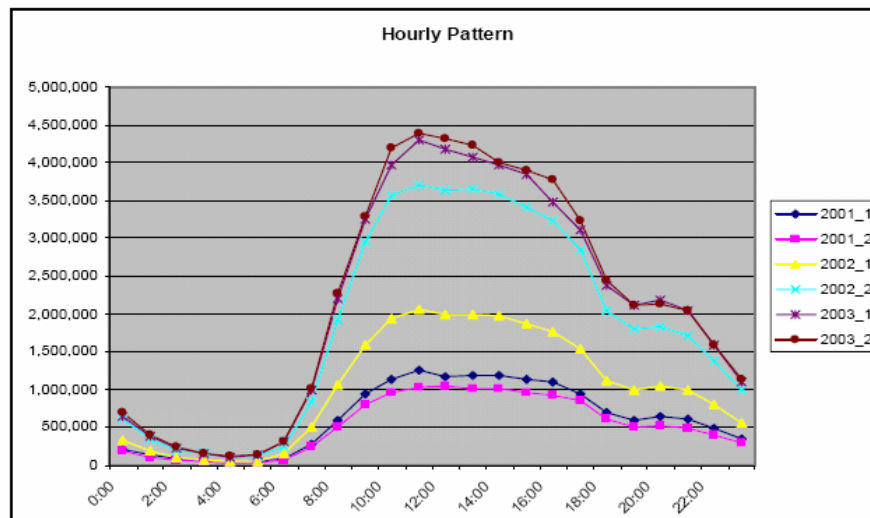


Figure 3: Hourly pattern of usage over a 24 hour period –comparison between semesters and years.

The pattern of load, in addition to feeding into planning on a technical level (such as timing of planned systems outages or predicting hardware requirements), also has learning design implications such as:

- timing of online chat/ help sessions (for example, the data indicates enterprise-wide learning support offered online may be most useful to QUT students between the hours of 10am and 12noon; while unit coordinators may wish to view the unit’s peak times to schedule the online consultation hour);
- rate at which the designer may reach the greatest number of learners with new content (at an overall enterprise level, Mondays receive the most visits to OLT; so content or activities added just prior may be seen by the same number of students as Friday night content will have reached over the weekend).

Snapshots of systems usage and learning and teaching behaviours allow for comparison against strategic intent. QUT’s top level plans specify such goals as “facilitating optimal student learning outcomes by seeking out and capitalising on emerging technologies and integrating information and communications technology into our teaching” and specify the dual measures of student satisfaction and benchmarking of IT usage in teaching. Enterprise-level data on usage of the OLT system thus allows for the benchmarking of IT usage in teaching against previous timeframes. In the following sections, we examine results of data mining system usage of discussion forums as a specific example of both evaluating current learning and teaching behaviour online as well as benchmarking the usage of communications technology in teaching.

Findings and comparisons at the Faculty level assists in the identification of the styles and practices implemented across a particular faculty and how to best exploit the affordances of online learning environments. Faculty data provides an insight into the achievement of objectives on faculty learning and teaching plans. For example, the Faculty of Education has specified as a goal “exploring new uses of learning spaces and technology-enhanced learning environments, as indicated by: a Faculty-wide project completed on technology-assisted innovations in pedagogy, resulting in higher level of technology use across all units”. The quantitative data harvesting on a fortnightly basis will allow the faculty to monitor the level of technology usage over time in its units, to gauge progress towards stated outcomes.

Example of a Communication/ Collaboration function: the Discussion Forum

The integration of discussion forums into learning activities is considered essential for the development and support of online communities (Palloff & Pratt 1999). With increasing time constraints for students and a focus on group work and collaborative practices the use of discussion forums assists in alleviating the barriers associated with group work, such as available mutual meeting times and appropriate venues. Website coordinators decide precisely how discussion forums will be displayed and implemented. The resource permits designers to either create one discussion page with multiple discussion threads or conversely create multiple forum pages with specific discussion threads embedded on each forum page.

Through examination of how designers are constructing forum pages, forums and topics we can refine our understanding of how communicative spaces are being designed with and informed by the OLT environment.

Within the OLT system, contributions to the discussion forum resource are “tagged” as either a post or as a reply to a previously posted message. Interrogation of this data provides a snapshot of the degree of discussion occurring within a forum. By determining the frequency and ratio of postings versus replies, it is possible to estimate how the faculty is predominantly utilising the discussion forum resource. Burr (2004) maintains that where active collaboration among learners is occurring, an increase in the ratio of replies to postings will be observed; conversely, where the forum is being used as a simple question-and-answer resource, the ratio of replies to posts is anticipated to be approximately a 1:1 ratio. Furthermore, the data mined also tags the contribution to discussion forums according to whether the author is a learner (student) or designer (staff). In this way, it is possible for an overview of the types of interactions on the forum to indicate how much staff input is driving the discussion.

Enterprise-wide Discussion Forum Findings

Despite a dramatic increase in overall access to OLT sites over the same time period (Figure 2), it appears that there is very limited consequential increase in the number of additional discussion forums implemented within the same time frame. The average number of forums added to an OLT unit has remained consistent over the 2003 -2004 time frame with small increases observed (increasing from 1.5 – 1.83), while units have an average of 28 contributions to the unit discussion fora. The establishment of institutional benchmarks in such areas as average number of contributions to discussion forums (28) and the ratio of posts to replies (which has approximately a 1:1 ratio in the beginning of 2005), permits faculties and units to compare their results against an institutional average.

Faculty-wide Discussion Forum Findings

Faculty results for an analysis of average forum contributions per unit website show a high degree of variation in the numbers of posts on discussion forums across faculties at QUT (figure 4). Some of this variation is attributable to different forum software being used (for example in the Faculty of Science), while specific faculty strategies also account for the observed variance in usage.

The Faculty of Education’s strategy to encourage greater exploitation of the online medium with a staff development program has seen the number of average contributions jump 33% from 32.44 in January to 43.38 in March this year (compare Edu1 and Edu2 in Figure 4). Comparison of the faculty’s result against changes in contributions across all faculties allows a closer inspection of the learner and teacher behaviour change in the context of the overall trend for increased usage of ICTs across the institution. A closer look at the percentage posts versus the percentage of replies across Faculties assists in providing a faculty-level snapshot of the degree of discussion occurring within the increased usage of the discussion forum. Clearly, the posts for the Faculty here outnumber the replies, and could be indicative of

- little genuine interaction and isolated comments on the forum without response;
- usage of the discussion forum for a different purpose than primarily communication or collaboration (e.g. for students to record their progress or store their draft assignments);
- user misunderstanding of how the discussion forum works with posts and replies; or
- an increase of usage of the discussion forum in early stages before interactions and discussions formulate during the course of the semester.

Investigation into some units which have particularly high postings is discussed below to provide some insight into possible reasons for the results. Faculty averages essentially provide a reference point for further investigation into individual units, and potentially help frame the types of learning and teaching activities within the faculty context as well as identify units that may serve as good examples of successful learner collaboration. Within the replies for each faculty, data mined also summarized the percentage of learner-to-learner interactions (Table 2), a lead indicator for potential learner collaboration and support and potentially greater depth to the discussion.

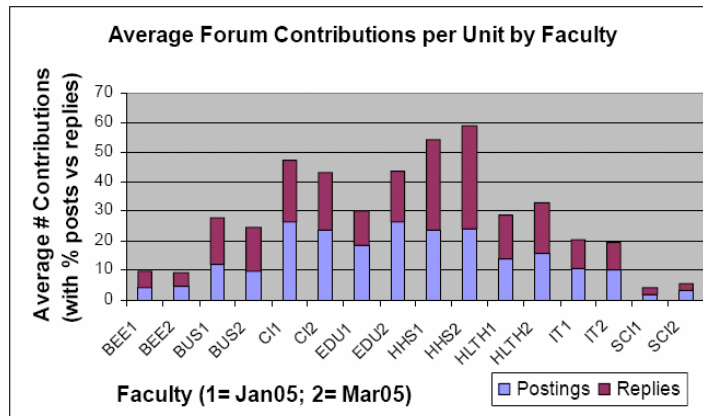


Figure 4: Average contributions, percentage of posts and replies: a comparison across faculties (1) early January 2005 and (2) late March 2005 (NB Science has parallel forum software that is not included in the data mining)

BEE	BUS	CI	EDU	HHS	HLTH	IT	LAW	SCI
38%	20%	36%	23%	44%	43%	28%	37%	44%

Table 2: Faculty averages for student to student replies as a total of discussion forum contributions (Jan 05).

Unit level

findings will assist unit coordinators in the integration of their different teaching delivery modes (lectures, tutorials, online, etc.) and aid in the reflecting on implemented learning and teaching activities within the overall benchmarks set by the institutional and Faculty averages, as well as specific unit-level objectives. For example, in a postgraduate course in the Faculty of Education, the unit coordinators implemented a discussion forum for each student with the dual intention of

- requiring students to keep a reflective journal of their research progress, while
- promoting an online community of practice among the student cohort, through students being able to read and respond to their peer's journal entries.

Thus, the reflective journal was intended to lead to valuable discussions and student sharing of information via the online discussion forum. However, mined data on learner to learner interactions indicates a relatively low percentage of the total contributions (9.5% in the period Jan-Mar 2005), despite forum contributions being above faculty average. Discussions with teaching support staff led to the examination of the forum data in relation to the stated aim of generating a community of practice, and has resulted in the redesign of the site in an attempt to foster greater student to student communication. Further monitoring of the quantitative data on student-to-student interactions of the site will assist unit designers to evaluate the results of this design change.

Conclusion

This study has outlined some of the quantitative evaluation usages of LMS data on systems usage, using the case study of Queensland University of Technology. The findings indicate a general 400% growth in usage of the OLT system within the two years 2001-2003. Usage is literally 24 hours, peaking on Mondays and Tuesdays, however low usage times are restricted to between 2am and 7am, and generally lower on weekends. This quantitative usage information helps assess university goals such as flexibility of time and place of study for students, as well as plan systems downtimes. This study mined usage data of discussion forums to examine the way this tool supported communication and collaboration within individual units and across the university. Data on discussion forums usage was mined to look at a) the overall averages of contributions to discussion forums, b) the amount of postings vs replies (where replies may indicate an

ongoing conversation) and c) the amount of learner to learner interactions vs learner to teacher or teacher to learner interactions (which may indicate how much peer support and interaction is occurring).

Further investigation into some comparative data mining is needed to realize the full potential of this method for evaluating learning and effectiveness of learning designs. For example, access trends to types of material - (number of times, how long for each time and timing in relation to the semester), and correlation between various learning designs and assessment results, would provide valuable additional information on learning and teaching behaviours.

Bibliography

- Burr, L. (2003). *The nature of interaction within an online computer mediated environment*. Published thesis. Charles Sturt University 2003.
- Burr, L. & Spennemann, D. (2004). Patterns of user behaviour in university online forums. *International Journal of Instructional Technology & Distance Learning*, 1(10) [Online] http://www.itdl.org/Journal/Oct_04/article01.htm [Accessed 20 March 2005]
- Dawson, S.P., Winslett, G., Tchakov, N.(2004), *Evaluation Report 2004*, Unpublished Report, Queensland University of Technology.
- Ehrmann, S. C., (1995). Asking the right questions: What does research tell us about technology and higher learning? In *Change. The Magazine of Higher Learning*, XXVII (2) (March/April), pp. 20-27
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996), From data mining to knowledge discovery in databases, *AI Magazine*, pp. 37-53.
- Heathcote, L., Prakash, S. (2003), What Your Learning Management System Is Telling You About Supporting Your Teachers: Monitoring System Information To Improve Support For Teachers Using Educational Technologies, In *Proceedings of the International Conference on Information Communication Technologies in Education* [Accessed 27/04/2005] http://www.ineag.gr/iciete/repository/ICICTE_Program.zip
- Huang, H.M. & Wu, C.H. (2004). Applying Data Mining in E-learning Environment. *Learning Technology newsletter*, 6 (2):31-32
- Kitto, S. & Higgin, V. (2003). Online university education: Liberating the student? *Science as Culture*, 12(1):23-58
- Ladyshewsky, R. (2004). E-learning compared with face to face: Differences in the academic achievement of postgraduate business students. *Australasian Journal of Educational Technology*, 20(3):316-336
- Moen, W. E. and McClure, C. R. (1997) "An Evaluation of the Federal Government's Implementation of the Government Information Locator Service (GILS)" Final Report Appendix E-4 *Web Server Transaction Log Analysis Results* submitted to General Services Administration Office of Information Technology Integration, Washington, DC. <http://www.unt.edu/wmoen/publications/gilseval/e4.htm>
- Oliver, M. (2000). An introduction to the evaluation of learning technology. *Educational Technology & Society*, 3(4):20-30
- Palloff, R., & Pratt, K. (1999). *Building Learning Communities in Cyberspace*. San Francisco: Josey-Bass Publishers
- Tolle, J. E (1984), Monitoring and evaluation of information systems via transaction log analysis, *Annual ACM Conference on Research and Development in Information Retrieval* (British Computer Society, Swinton, UK) <http://portal.acm.org/citation.cfm?id=636821&jmp=cit&dl=GUIDE&dl=ACM>

Acknowledgements

The authors wish to thank Greg Winslett and Nikita Tchakov for their contributions in the design of this project.

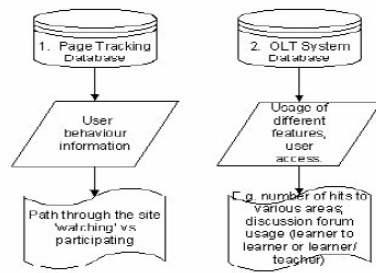


Figure 1: Data mining from usage tracking databases on OLT system (from Heathcote & Prakash: 2004)