Data for free: Using LMS activity logs to measure community in online courses

[note: figures and table at end of document]

1. Introduction

The United States has experienced unprecedented growth in both the availability of online education programs and participation in these programs. At present, there are over 3.2 million online students at the college and university level (Allen & Seaman, 2007); over 96 percent of the very largest higher-educational institutions have online course offerings (Allen & Seaman, 2006). Over 700,000 students also participate in K-12 online education (Smith, Clark & Blomeyer, 2005). Clearly, online learning is experiencing phenomenal growth. However, methods and tools for researching experiences within online communities have not kept pace; education lags behind industry and government in the use of comprehensively-gathered and carefully-analyzed data to support decision making (Author A, in press). The growing use of Learning Management Systems (LMS), many of which automatically keep logs of student activity, presents an exciting means of narrowing this gap. Lately, many researchers have worked to exploit this potential, both in academic research and the design of practical online learning applications. The present study continues this work. We explore whether students' perceptions of community can be measured via logs of student activity within graduate level online courses. Since feelings of community are known to significantly affect online learning performance, such a simple and immediately accessible measurement of this affective variable would be useful for e-learning instructors and researchers alike. Further, the measurement would provide a non-invasive alternative to currently-employed survey methodologies; this is a growing need as students increasingly develop "survey fatigue," an apathy toward completing surveys. We begin with a discussion of the importance of community in e-learning.

2. Literature Review

2.1 Community in online learning

Throughout the last 10-15 years, online learning researchers and instructional professionals have promoted the significance of community in online learning environments (Wallace, 2003).This importance is likely only to grow as online students increasingly come to see community as a fundamental part of online life (Weller, 2007). Collaboration between both students and online teachers is necessary to effectively cultivate a thriving online community (Berge & Collins, 1995; Palloff & Pratt, 1999). According to Wallace (2003), community in online environments arises at the intersection of three contemporary components in educational research: social learning theories, the affordances of computers as communication devices and increased utilization of theory in online course development.

Alfred Rovai (2002c) defines community in online learning environments as:

> *...consisting of two components: feelings of connectedness among community members and commonality of learning expectations and goals….Classroom community is strong when learners (a) feel connected to each other and to the instructor, (b) manifest the immediate communication behaviors that reduce social and psychological distance between people, (c) share common interests and values, (d) trust and help each other, (e) actively engage in two-way communications, and (f) pursue common learning objectives. (p. 322)*

Hung and Chen's (2001, p.10) dimensions of principles of learning support Rovai's definition of online community:

1. ***Situatedness****: fostered by contextualized activities, e.g. tasks and projects based on demand and needs.*
2. ***Commonality****: fostered by shared interests, e.g. in books; and shared problems.*
3. ***Interdependency****: fostered by varying expertise levels; varying perspectives or opinions;varying needs, mutual benefits; and complementary motives.*

This further grounds the concept of online community within the work of Vygotsky and Spiro.

It is clear that community is an essential part of successful online education. Limited face-to-face communication can lead to feelings of isolation which, in turn, can lead to dissatisfaction, poor performance and course non-completion (Cereijo, Young & Wilhelm, 2001; Curry, 2000; Rovai & Whighting, 2005). Research by Haythornthwaite, Kazmer, Robins and Shoemaker (2000) relates feelings of isolation to a low sense of community. Findings by Eastmond (1995) indicate that isolation can be alleviated when learners support one another. Additionally, Rovai (2002b) has demonstrated that encouraging a sense of community will effect student satisfaction, learning and retention.

Given this well-established importance of community in e-learning, instructors and administrators are typically keen to foster a sense of community in e-learning students (Mazzolini & Maddison, 2007). However, the nature of e-learning often makes this troublesome. Specifically, the lack of face-to-face interactions in the online environment makes it very difficult to appraise online classroom community (Vrasidas, 2004; Mazza & Milani, 2005; Mazzolini & Maddison, 2007). Lacking access to the same breadth of social indicators as their classroom counterparts, E-learning instructors must assess community through a diminished interaction "bandwidth" (Van Lehn, 1988). It is little surprise, then, that instructors are often mistaken in their assessments of online social situations such as class discussions (Mazzolini & Maddison, 2007).

Researchers, administrators, and instructors have turned to survey data to answer questions relating to classroom community (Rovai & Whiting, 2005). However, there are significant limitations to this approach. First, today's online students are over-surveyed (Dillman, 2002), subjected to increasing numbers of surveys and assessments seeking to

understand their motivations, concerns and mind-set. Students see little relevance in many of these surveys, increasing student apathy and non-response (Kalton, 2000; La Bruna & Rathod, 2005). Some universities, recognizing that "...student cooperation with surveys [is] a scarce and valuable resource that should be used wisely," have begun to institute policies guiding and limiting survey access to students (Porter, 2005). This one-two punch of decreasing reliability and availability of survey data will no doubt impact the usefulness of this methodology. Second, assessment tools necessary for the measurement and evaluation of key factors that equate to online learning success have not kept pace with online education's explosive growth. A limited range of assessments are available for use within online education programs. Few of these assessments have proven valid and reliable (Author, in press).

2.2 Non-invasive measures in online environments

In order to satisfy the need for valid and reliable assessment tools in today's environment of survey-saturated students, many have advocated adopting new approaches to data-gathering (Sinickas, 2007; Grofton, 1999). Until recently, educators seemed reticent to embrace data mining and statistical techniques to analyze data recorded by computing media themselves (Lowes, Lin & Wang, 2007; Lopes & David, 2006; Klassen & Smith, 2004); however, such methods are now rapidly gaining popularity (Romero & Ventura, 2007). A common theme to these approaches is that they are less intrusive and subjective, though typically requiring more processing than survey methods (Pahl, 2004). Within this general paradigm of non-invasive assessment several different approaches have emerged, each with its own advantages and weaknesses. Researchers have made use of data from three main sources: (1) recorded text, (2) web server log files, and (3) learning software log files. Several such studies are listed in Table 1.

2.3 Recorded Text

Several authors (Lowes, Lin & Wang., 2007; Dringus & Ellis, 2005; Mazzolini & Maddison, 2007) have employed data mining of text communications in learning management systems (LMS) and computer supported collaborative learning (CSCL). This is a particularly rich source of data which has yielded significant findings. Unfortunately, while automated text mining using artificial intelligence algorithms has shown considerable promise in educational applications (Mochizuki, Kato, Yaegashi, Nagata, Nishimori, Fujitani et al., 2005; Tane, Schmidt & Stumme, 2004), mining for relatively subtle social indicators remains impractical (Dringus & Ellis, 2005). Consequently, this methodology is limited by the need to perform relatively labor-intensive hand-coding.

2.4 Web server log files

Another source of automatically collected data is web server logs; these are vast collections of data relating the accessing of specific web pages (Hanna, 2004). E-learning researchers (Zaïane 2001; Klassen & Smith 2004; Monk, 2005; Zorrilla, Menasalvas, Marin, Mora & Segovia, 2005; Lopes & David 2006) have employed data mining techniques to gain useful insight from these data. Though, the low-level information collected in server logs would seem to be ill-suited for observing high level, social phenomena; the above-mentioned studies make little or no mention of e-learning

community, presumably for this reason. In addition, server logs are plagued with a low signal-to-noise ratio; simply preparing the data for modeling can consume 80% to 95% of a projects resources (Edelstein, 2001), making this—like text mining—somewhat labor-intensive.

2.5 LMS log files

Perhaps the most promising source of automatically gathered e-learning data is the learning software itself, particularly the learning management system. Since students typically log in to such systems, keeping track of users and sessions—a major hurdle in examining server logs (Zorilla, 2005)—is done automatically. In addition, many such systems gather a range of relatively high-level student data such as quiz grades and forum posts (Mazza & Milani, 2005). These data are both more focused than raw server logs, and more convenient than hand-categorizing text communication.Drawing in part upon work analyzing community in online threaded discussions (such as Donath, Karahalios & Viegas, 1999), several researchers have mined data from LMS to examine community in e-learning. Many, including Shen, Nuankhieo, Huang, Amelung & Laffey (2007) and Reffay & Chanier (2002), have employed social network analysis (SNA) to examine the nature of e-learning communities. Although not examining LMS data as such, Nurmela, Lentinen & Palonen (1999) used a similar source in applying data gathered from CSCL log files to SNA techniques, as well. Our study extends this research by exploring whether students' perceptions of community, an affective variable known to be important to online learning success, may be measured using data from LMS log files.

Table 1:Alternate sources of e-learning data

3. Methodology

3.1 Research Question

Does a student's perception of community relate to the number of LMS data log events that student generates? Specifically, this method focuses on Alfred Rovai's CCS: Classroom Community Scale (Rovai, 2002a) and its relationship to data logs. The CCS questionnaire is included as Appendix A.

3.2 Participants

Data for this study were collected from 67 individuals, 22 male and 45 female. The sample had an average age of 37.4 (SD = 11.6). All were enrolled in graduate level online educational technology courses at a large university in the Southeast United States. All participants had previous experience with online courses; on average, participants had taken 2.1 (SD = 1.6) online courses prior to participation in the study.

Table 2: Students were recruited from the following courses

Given that all courses were graduate level and taught within the educational technology program, similarity was assumed. It should be noted there was variation in the teaching experiences of the instructors for the courses; the course content and course materials were also a source of differentiation. Further, requirements for interaction with both the instructor and other members of the course varied based on the content within the course.

3.3 Data Collection

Survey data were collected during the final week of each eight week course; students logged onto a website and completed the Classroom Community Scale (CCS) (Rovai, 2002a). The CCS is a 20 item questionnaire consisting of three constructs: learning, connectedness and classroom community. The CCS is a reliable measure of community ($\alpha = .93$); both the connectedness and learning subscales are also considered reliable measures with α of .92 and .87 respectively. In addition, the CCS exhibits both validity (KMO: .94; Barlett's test of sphericity: $X^2 = 3883.95$, $P<.01$) and usability (Fleisch-Kincaid grade level score: 6.6) (Rovai, 2002a). Once the courses had been completed, the cumulative number of data logs for each student was downloaded utilizing the Moodle learning management system's data reporting feature. The reporting feature records and compiles student clicks within the course environment. Although our study was interested only in the cumulative number of data logs, Moodle sorts logs by type as well, including the number of pages within the LMS visited, messages read in discussions, posts in discussion and the utilization of intra-LMS communication (email, chat). A teacher or administrator can access the reporting feature and view or download (in Excel or text format) multiple reports based on their preferences. For example, reports can be generated for individual students, multiple students, and specific assignments or for a specified time period.

Once downloaded, the log data were sorted with Microsoft Excel, creating a single spreadsheet containing the complete set of data logs for the entire semester. This spreadsheet was then imported into SPSS v. 13. A frequency analysis was run on the data set, grouping by student names. Through this method a cumulative number of logs for each student was obtained; also, a summary of the students' activities in the LMS was obtained allowing for future analysis of specific types of events (e.g. content views, new postings, replies to existing postings). Cumulative logs were combined with CCS scores for each participant; linear regression procedures were then performed upon the data set to discern the relationships between the independent variables (sense of community, connectedness and learning) and the dependent variable (data log events). Pearson product moment correlations were conducted to understand the relationship between connectedness and learning. Finally, analysis of variance procedures were utilized to explore differences between the studied courses.

4. Results

Data logs were determined to be a predictor of sense of community (Adj $R^2 = .086$, $F(2, 64) = 4.090$, $p = .021$). Course was non-significant in this regression equation. Results from this regression procedure are summarized in Table 3.

Table 3: Regression results, dependent variable = community, independent variables = data logs, course

Course and data logs have a relationship to the connectedness construct (Adj $R^2 = .066$ $F(2, 64) = 3.341$, $p = .042$). This result is significant, though a minimal amount of the variance within the dependent variable, connectedness, can be accounted for by data logs and course. Results from this regression procedure are summarized in Table 4.

Table 4: Regression results, dependent variable = connectedness, independent variables = data logs, course

Neither course nor log files were related to the learning construct (Adj $R^2 = .045$, $F(2,64) = 2.54$, $p = .087$). An analysis of Pearson product moment correlations (see table 5) between the three constructs produced by the CCS instrument reveals a significant strong positive correlation between connectedness and community ($r = .774$, $p < .01$) and a significant strong positive correlation between learning and community ($r = .597$, $p < .01$).

Table 5: Pearson product moment correlations: Connectedness, Learning and Community

5. Discussion

The current study examines, what is to our knowledge, a novel measurement of feelings of community: the total number of LMS log events. Results indicate that data logs generated during an online graduate level course have a relationship to both classroom community and feelings of connectedness. Specifically, log files were determined to be a predictor of sense of community (Adj $R^2 = .086$, $F(2, 64) = 4.090$, $p = .021$), and course and log files have a relationship to the connectedness construct (Adj $R^2 = .066$ $F(2, 64) = 3.341$, $p = .042$).

The total number of log entries, discussion posts, or similar simple events is bound to have a complicated relationship, to say the least, with students' feelings of community (Mazzolini & Maddison, 2007). Logs can record many different actions, each of which can and does have a variety of different causes. Partly for this reason, Lowes et al. (2007) found inspection of LMS log data to be of little use in the study of e-learning community. However, that study did not attempt to correlate log data with survey data;

indeed, very few have (for an exception see Shen, Piyanan, Xinxin, Amelung & Laffey, 2007). Our data indicate that such a correlation exists, atop a web of more complicated and (to date) less clear relationships between students' mental states and LMS data. This study suggests that the simple measure of log events reveals a forest from these trees.

We see two uses for this measurement: first, to support real time data analysis for constructing visualizations and modeling students, and second, to augment survey data for informing long and medium-term decision-making.
Unlike complicated data mining techniques, counting total log entries would be easy to automate for use in real time. This automated tool could be applied to a LMS as a module; this would be a relatively cheap and easy way to facilitate a move towards intelligent learning management systems. Such a system could be used until the adoption of more sophisticated modules, or could evolve more sophistication itself In either case, it represents a very practical step toward a system to comprehensively broaden an online instructor's perception of her students. Additionally, the use of data logs to predict affective information about students in online courses can decrease the need for surveying and provide the opportunity to measure an affective variable without impacting the student. Given recent concerns over survey fatigue (Dillman, 2002), an alternative method of data collection that doesn't inconvenience students may be an important resource for decision-makers and researchers.

Data analysis is used in both the private and public sectors to collect information about individuals for a more comprehensive understanding of their behaviors, needs and concerns (Davenport & Harris, 2007; Ayers, 2007). Strategic decisions made at many of the largest, highly respected and most successful corporations in the United States, including Google, Wal-Mart and CapitalOne, are guided in a large part by real-time and post-hoc data-driven processes (Davenport & Harris, 2007). Unfortunately, the use and adoption of data-analytic practices has not been at the forefront of the education movement; few have fully embraced a strategic focus on the collection for the improvement of learning outcomes.

According to Guthrie (2007),

> …there are few 21st century operations as outmoded as educational data systems….Wal-Mart managers routinely know more regarding the location…of a toy bear manufactured in China, from the original point of purchase manufacturing specifications to the vendor's ocean shipping arrangement, to local store delivery and shelving and time of final placement into a customer's shopping basket than school district administrators know regarding the day-to-day status and school progress of their enrolled students. (p. 1)

Online education providers, while using web-based technologies, are no further ahead of their traditional counterparts with regards to the data collection and analysis capabilities

of their educational data systems (Pahl, 2004; Zaiane, 2001). However, this mistake is beginning to be rectified, as researchers find ways to apply automatically collected web data to the study of e-learning.

Large stores of automatically-collected personal data carry with them the risk of abuse, making the construction of data and policy structures to safeguard student privacy a priority for future researchers. Future research should also work to apply new methods of data analysis, such as geospatial data analysis, to web data. Additionally, researchers should examine path analysis procedures based on click stream data and increased application of regressive techniques to predict student behavior and learning in online environments. The analysis of particular log types (discussion posting, forum posting, etc.), while more complex, also carries great potential for delivering more complete, nuanced information about students; researchers should continue to empirically examine the complex interrelationships between student states and these log types. All of these represent immense opportunities for distance education providers given the diversity of e-learning students.

6. Limitations

There are several limitations to this study worth noting. First, the narrow sample size greatly hinders the generalizability of the results. The content of the courses was not taken into consideration, and further, course instructors had varying degrees of experience in online instruction; it would be expected that these particulars would shape student perceptions and outcomes. Second, there were differences in the students who participated in the courses; while all students had prior experience as online learners, there were differing levels of comfort and technological sophistication amongst the learners. Third, it is possible that the CCS instrument, which was intended to analyze the outcomes of a single course, became a moratorium on the online learning program in general. Many students who participated in the study were enrolled in multiple online courses, allowing for the possibility that student responses bridged multiple course experiences. Finally, greater variation in the values assigned to different types of logged events may have a significant effect upon predictability. This study did not take into account the nature of the activity engaged by the student; for example, it would be natural to assume that the process of responding to a peer's discussion post would have greater influence on community then the process of reading the syllabus. Research by Lowes, Lin and Wang (2007) finds that specific typologies of responses in discussion forums have the ability to facilitate a more robust discourse. Of course, finding the best indicators and writing an algorithm to represent them is easier said than done, and is the topic of considerable active research (Tzoumakas & Tzoumakas, 2005; Dringus & Ellis, 2005).

7. Conclusion

Use of computers in education has often been criticized for doing old things new ways—failing to take advantage of the revolutionary possibilities of new technology (Jonassen, Davidson, Collins, Campbell & Haag, 1995; Foshay & Bergeron, 2000). Using a computer to disseminate and collect a survey, for example, is certainly more

convenient—but not revolutionary (indeed, it may accelerate the affects of survey fatigue).Data from LMS, on the other hand, with its inherent capability to collect and report student actions, interactions and testing data, have potential to qualitatively change teaching and learning. Instead of relying on often-fallible intuitions based on an impoverished data stream, future e-learning instructors may well take advantage of what computers are good at—gathering and sorting data—to build representations of online students that are in many ways richer and more accurate than they'd have had in the classroom. At higher levels, administrators and course designers will be able to embed features and dynamic content that encourage a deeper exploration of content. Current research attempting to identify indicators of student attitudes like sense of community brings us closer to such a reality.

References

Author (in press). An overview of evaluative instrumentation for virtual high schools. Paper to appear in *The American Journal of Distance Education*.

Allen, I.E. & Seaman, J. (2006). *Making the Grade: Online Education in the United States.* Needham, MA: Sloan Consortium.

Ayers, I. (2007). *Super Crunchers: Why thinking by the numbers is the new way to be smart.* New York: Bantam Dell.

Berge, Z.L. & Collins, M.P. (1995). *Computer Mediated Communication and the Online Classroom.* Cresskill, NJ: Hampton Press.

Cereijo, M. V. P., Young, J., & Wilhelm, R. W. (2001). Factors Facilitating Student Participation in Asynchronous Web-Based Courses. *Journal of Computing in Teacher Education*, *18*.

Curry, D. B. (2000). Collaborative, Connected and Experiential Learning: Reflections of an Online Learner, Retrieved 7/15/2007, from http://www.mtsu.edu/~itconf/proceed01/2.html.

Davenport, T. & Harris, J. (2007). *Competing on analytics: The new science of winning*. Boston: Harvard Business School Press.

Dillman, D. (2002). Navigating the rapids of change: Some observations on survey methodology in the early twenty-first century. *Public Opinion Quarterly, 66(3),* 473-494.

Donath, J., Karahalios, K., & Viegas, F. (1999). Visualizing conversation, *System Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on System Sciences-Volume 2*.

Dringus, L. P., & Ellis, T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education*, *45*, 141-160.

Eastmond, D.V. (1995*). Alone but together: Adult distance study through computer conferencing*. Cresskill, NJ: Hampton Press.

Foshay, R., & Bergeron, C. (2000). Web-based education: A reality check. *TechTrends*, *44*, 16-19.

Hanna, M. (2004). Data mining in the e-learning domain. *Campus-Wide Information Systems*, *21*, 29-34.

Haythornthwaite, C., Kazmer, M., Robins, J., and Shoemaker, S. (2000). Making connections: Community among computer-supported distance learners. Paper presented at the *Association for Library and Information Science Education 2000 Conference*. San Antonio, Texas. Retrieved July 15, 2007 from: http://www.alise.org/conferences/conf00_Haythornthwaite_Making.htm

Hung, D.W.L. & Chen, D.T. (2001) Situated cognition, vygotskian thought and learning from communities of practice perspective: implications for the design of web-based e-learning, *Education Media International*, *38*(1), 3–12.

Jonassen, D., Davidson, M., Collins, M., Campbell, J., & Haag, B. B. (1995). Constructivism and computer-mediated communication in distance education. *The American Journal of Distance Education*, *9*, 7-26.

Kalton, G. (2000). Developments in survey research in the past 25 years. *Survey Methodology, 26*(1).

Klassen, K.J. & Smith, W. (2004). Web log analysis: A study of instructor evaluations done online. *Journal of Information Technology Education 3,* 291-312.

LaBruna, A. & Rathod, S. (2005). Questionnaire length and fatigue effects. Bloomerce White Paper #5, accessed 7/15/2007 http://www.websm.org/uploadi/editor/1143807053whitepaper5.pdf

Lopes, C. T., & David, G. (2006). Higher education web information system usage analysis with a data warehouse. Computational Science and its Applications – ICCSA 2006. Berlin: Springer.

Lowes, S., Lin, P. & Wang, Y. (2007). Studying the effectiveness of the discussion forum in online professional development courses. *Journal of Interactive Online Learning, 6*(3), 181-210.

Mazza R., Dimitrova V. (2004). Visualising Student Tracking Data to Support Instructors in Web-Based Distance Education. 13th International World Wide Web Conference (WWW 2004) Alternative Education Track.

Mazza, R., Milani, C. (2005). Exploring Usage Analysis in Learning Systems: Gaining Insights From Visualisations. 12th International Conference on Artificial Intelligence in Education (AIED 2005). Amsterdam, The Netherlands.

Mazzolini, M., & Maddison, S. (2007). When to jump in: The role of the instructor in online discussion forums. *Computers & Education*, *49*, 193-213.

Mochizuki, T., Kato, H., Yaegashi, K., Nagata, T., Nishimori, T., Hisamatsu, S., et al. (2005). Promotion of self-assessment for learners in online discussion using the visualization software In , *Proceedings of the 2005 Conference on Computer Support for Collaborative Learning: Learning 2005:The next 10 years!* (440-449). Taipei, Taiwan: International Society of the Learning Sciences.

Monk, D. (2005). Using Data Mining for e-Learning Decision Making. *The Electronic Journal of. e-Learning*, 3(1), 41-54.

Nurmela, K., Lehtinen, E., & Palonen, T. (1999). Evaluating CSCL log files by social network analysis. *Proceedings of the 1999 Conference on Computer Support for Collaborative Learning*.

Pahl, C. (2004). Data Mining Technology for the Evaluation of Learning Content Interaction. *International Journal on E-Learning*, *3*, 47-55.

Palloff, R.M. & Pratt, K. (1999) *Building Learning Communities in Cyberspace: Effective strategies for the online classroom.* San Francisco, CA: Jossey–Bass.

Porter, S.R. (2005). Survey research policies: An emerging issue for higher education. *New Directions in Institutional Research, 2005*(127), 5-15.

Reffay, C., Chanier, T. (2002). Social network analysis used for modeling collaboration in distance learning groups. In proceedings of Intelligent Tutoring Systems 6th Conference, Biarritz, France.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, *33*, 135-146.

Rovai, A. P. (2002a). Building sense of community at a distance. *International review of research in open and distance learning, 3*(1). Available: http://www.irrodl.org/index.php/irrodl/article/view/79/152

Rovai, A. P. (2002b). Development of an instrument to measure classroom community. *Internet and Higher Education*, *5*(3), 197-211.

Rovai, A. P. (2002c). Sense of community, perceived cognitive learning, and persistence in asynchronous learning networks. *Internet and Higher Education 5,* (4), 319-332.

Rovai, A.P. & Wighting M. J. (2005). Feelings of alienation and community among higher education students in a virtual classroom. *The Internet and higher education, 8*(2), 97-110

Shen, D., Nuankhieo, P., Huang, X., Amelung, C. & Laffey, J. (2007). Social Network Analysis to Understand Sense of Community in an Online Learning Environment, AERA, Chicago, IL.

Smith, R., Clark T. & Blomeyer, R.L. (2005). *A synthesis of new research on K–12 online learning.* Naperville, IL: Learning Point Associates.

Tane, J., Schmitz, C., & Stumme, G. (2004). Semantic resource management for the web: an e-learning application. *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, 1-10.

Tzoumakas, V., & Tzoumakas, B. (2005). Force Based Visualizations for Instructor Support. *Advanced Learning Technologies, 2005. ICALT 2005. Fifth IEEE International Conference on*, 452-456.

Van Lehn, K. (1988). Student modeling In M. Polson & J. Richardson (Eds.), *Foundations of Intelligent Tutoring Systems*. New Jersey: Lawrence Erlbaum Associates.

Wallace, R. (2003). Online learning in higher education: A review of research on interactions among teachers and students. *Education Communication, and Information, 3*, 241–280.

Weller, M. (2007). The distance from isolation: Why communities are the logical conclusion in e-learning. *Computers & Education*, *49*, 148-159.

Zaiane, O. R. (2001). Web usage mining for a better web-based learning environment. *Proceedings of Conference on Advanced Technology for Education*, 60-–64.

Zorrilla, M. E., Menasalvas, E., Marin, D., Mora, E., & Segovia, J. (2005). Web Usage Mining Project for Improving Web-Based Learning Sites. *Computer Aided Systems Theory Eurocast 2005: 10th International Conference on Computer Aided Systems Theory, Las Palmas de Gran Canaria, Spain, Revised Selected Papers*.

Figure captions, tables, figures, schemes.

Table 1:Alternate sources of e-learning data

Data source	Method of analysis	Applied to community?
Text communication records: rich, high-level data; time intensive coding		
Lowes, Lin & Wang (2007)	DM, SNA	Yes
Dringus and Ellis (2005)	DM	Yes
Mochizuki et al. (2005)	Real-time visualization, keyword recognition	Yes
Server log files: low-level data, high noise, difficult to organize		
Lopes & David (2006)	OLAP	No
Monk (2005)	Basic statistical	No
Zorilla, Menasalvas, Marin, Mora, and Segovia (2005)	DM, OLAP	No
Klassen and Smith (2004)	spreadsheet	No
Zaïane (2001)	DM	No
LMS log files: high-level data, more organized but still needs sorting		
Not real time		
Lowes, Lin, and Wang (2007, found of little use)	Basic statistical	Yes
Nurmela et al. (1999) (CSCL system log files)	SNA	Yes
Reffay and Chanier (2002)	SNA	Yes
Laffey, Shen, Nuankhieo, Huang, Amelung (2007)	SNA	Yes
Silva and Vieira (2002) (platform-agnostic)	DM	Somewhat
Real time		
Moodie and Kunz (2003, proposed iLMS)	AI	Yes
Santos, Rodriguez, Gaudioso, and Boticario (2003, proposed CSCL system)	AI	Yes
Kosba, 'TADV' iLMS (2004)	AI	Somewhat
Mazza 'CourseVis' LMS tool (2004)	Visualization	Yes
Ueno 'Samurai' iLMS (2004)	DM, AI	Yes
Mazza and Milani, 'GISMO' Moodle module (2005)	Visualization	Yes

DM=data mining **SNA**=social network analysis **AI**=artificial intelligence
OLAP=Online Analytical Processing (an analytic method similar to data mining)

Table 2: Students were recruited from the following courses

Course	Course Name	n
A	Instructional Computing 1	9
B	Instructional Computing 2	9
C	Internet and K-12	23
D	Designing and Delivering Online Content	6
E	Digital Photography and Visual Literacy	5
F	Instructional Design	15

Table 3: Regression results, dependent variable = community, independent variables = data logs, course

Variable	***B***	***SE B***	***β***	***Sig***
Logs	.008	.003	.341	.006*
Course	.538	.631	.102	.398
*$p < 05$.				

Table 4: Regression results, dependent variable = connectedness, independent variables = data logs, course

Variable	***B***	***SE B***	***β***	***Sig***
Logs	.005	.002	.288	.021*
Course	-.278	.513	-.066	.589
*$p < 05$.				

Table 5: Pearson product moment correlations: Connectedness, Learning and Community

		Connectedness	Learning	Community
Connectedness	Pearson Correlation	1	-.046	.774*
	Sig. (2-tailed)		.713	.000
Learning	Pearson Correlation	-.046	1	.597*
	Sig. (2-tailed)	.713		.000
Community	Pearson Correlation	.774*	.597*	1
	Sig. (2-tailed)	.000	.000	
*p < 05.				

Figure 1: Classroom Sense of Community Index (Rovai, 2002a): Observed variables (connectedness and learning) and their relationship to the latent variable (classroom community).

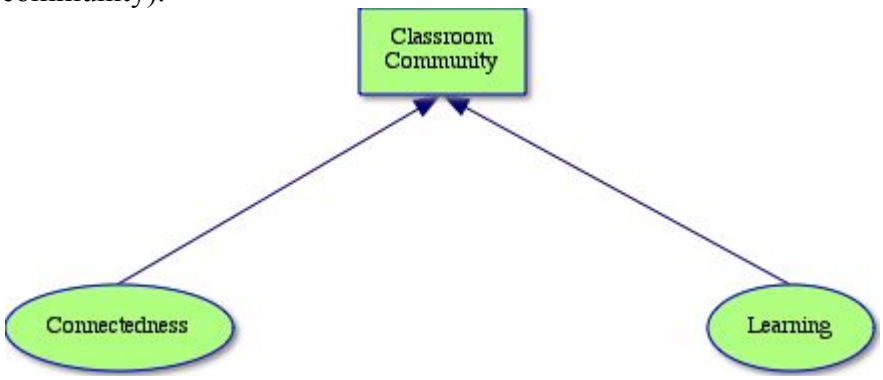

Figure 2: Screenshot, the Moodle LMS (Reporting Link Highlighted)

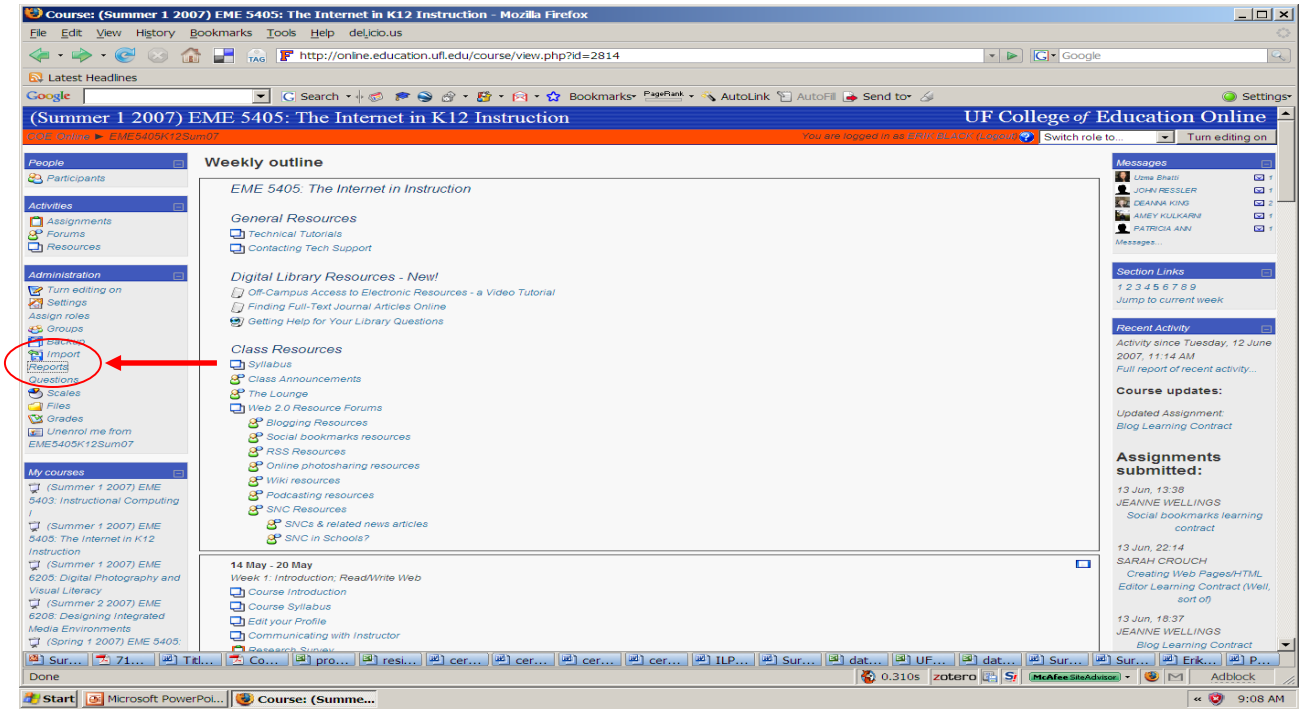

Figure 3: Screenshot, the Reporting Page in the Moodle LMS

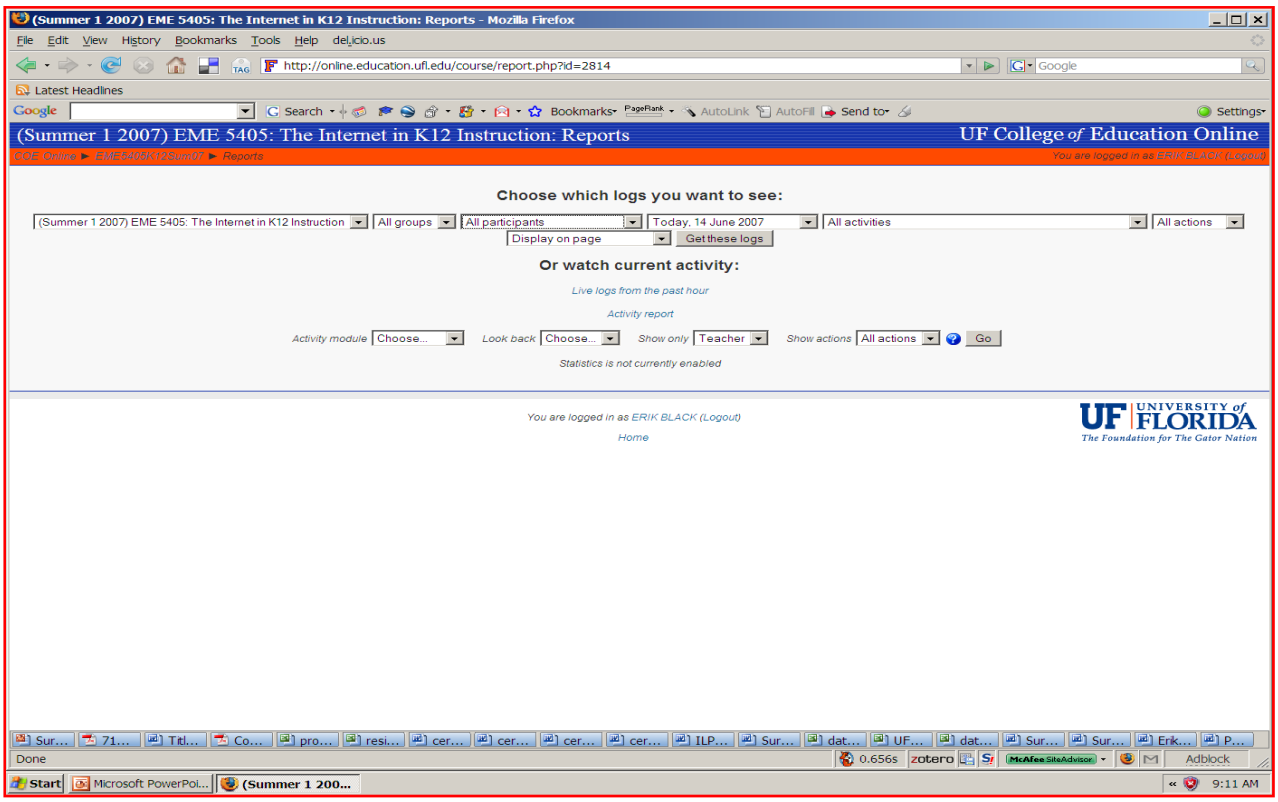

Figure 4: Scatter plot and linear regression line of data logs and community construct (RawScore)