121st ASEE Annual Conference & Exposition
Indianapolis, IN
June 15-18, 2014
Paper ID #10323

Testing a Prototype System for Mining of Student Notes and Questions to Create Study Guides

Dr. Perry Samson, University of Michigan

Perry Samson is Professor of Atmospheric, Oceanic and Space Sciences and Professor of Entrepreneurship in the College of Engineering at the University of Michigan. He holds an Arthur F. Thurnau Professorship at the University of Michigan in recognition of "outstanding contributions to undergraduate education" and is the recipient of the 2009 Teaching Innovation Award at the University of Michigan and a past recipient of the College of Engineering Excellence in Teaching Award. In 2010 Perry was named Distinguished Professor of the Year by the President's Council of Universities in the State of Michigan.

Testing a Prototype System for Mining of Student Notes and Questions to Create Study Guides

The Issue

In the foreseeable future it will be technically possible for instructors, advisors and other delegated representatives of a college or university to access student participation and performance data in near-real time. One potential benefit of this increased data flow could include an improved ability to identify students at risk of academic failure or withdrawal. The availability of these data could also lead to creation of new adaptive learning measures that can automatically provide students personalized guidance.

Methods

(Samson, 2010) reported that the availability of mobile tools that deliberately engage students during class dramatically changed the mechanics of course at the University of Michigan with over 80% of students attending lecture voluntarily bringing mobile devices to class. On one hand, surveys showed that students believe the availability of a laptop was more likely to increase their time on tasks unrelated to the conduct of the course. On the other hand, the surveys also ascertained that students felt more attentive with the technology, significantly more engaged, and able to learn more with the technology than in similar classes without it.

The mobile technology led to a dramatic increases in the number of students posing questions during class time, with more than half posing at least one question during class over the course of a semester, a percentage far higher than achieved in semesters prior to the use of this technology. Moreover, while 50% of men and 80% of women in the science course surveyed claimed to be uncomfortable asking questions in a large lecture setting, 66% of all students (men and women) ask questions when questions and subsequent answers are posted anonymously.

The tool employed for this study, LectureTools, allows the students to:

- Type notes synchronized with the lecture slides;
- Answer questions posed by the instructor
- Self-assess understanding and indicate when they are confused
- Pose questions to the instructor and view responses;
- Draw on the instructor's lecture slides; and
- Print lecture slides and notes for off-line review.

LectureTools (http://www.lecturetools.com) enables the instructor to ask a wide range of question types including multiple choice, reorder list, free response, numerical and image-based questions, excellent for testing students understanding of graphs, images and maps. These questions are embedded in the slides the instructor uploads into a tray (see Figure 1). The

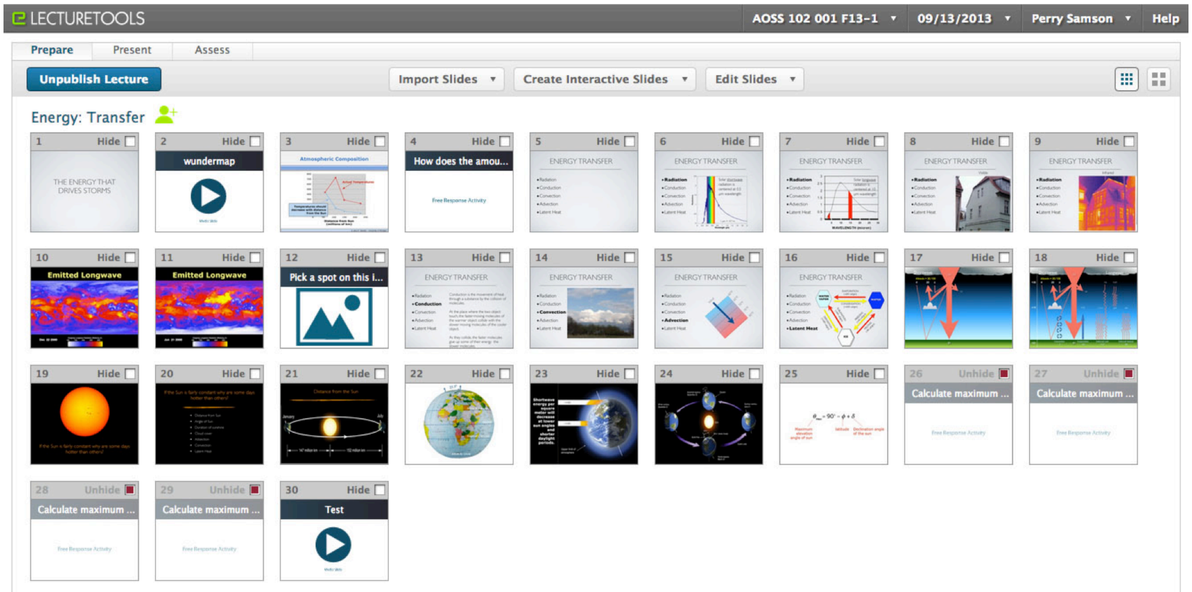

Figure 1. Workspace for instructor in LectureTools. Instructors upload their presentation slides into LectureTools and can add videos and a variety of question types to challenge student understanding. Instructors can also hide slides and reveal them during class.

instructor can "hide" slides so students cannot see them in class until released. The instructor has the additional option that they can add videos to the presentation directly from popular systems such as YouTube, Vimeo and more. An advantage of this is that students will have access to the slides, videos and questions during and after class.

Students report higher levels of engagement using LectureTools than their other classes (Figure 2) largely because the system allows them more opportunities to participate in class. They can take notes synchronized to each slide being presented, they answer questions posed by the instructor, they can pose questions to the instructor and they can even indicate when they are confused during class (see Figure 3).

The instructor also is presented with rich data on student performance that can help identify non-participating students far earlier as well as feedback on which slides and topics caused the most confusion for students.

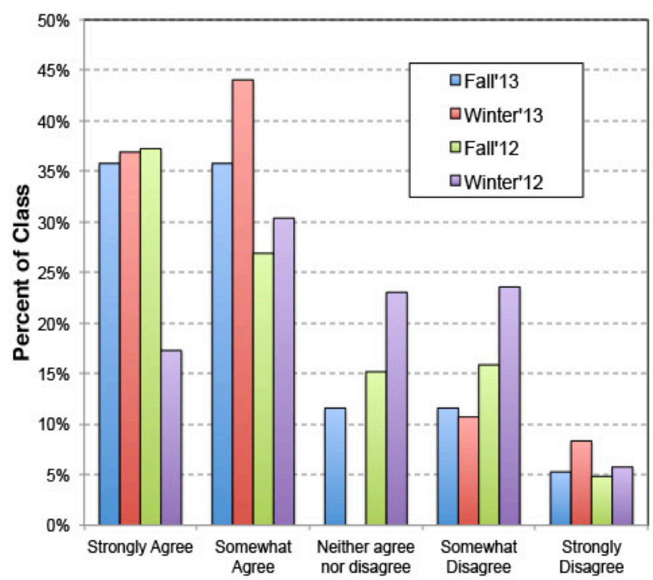

Figure 2. Student responses over four semesters to the end-of-semester question "Using a laptop increased my engagement in this class relative to other classes."

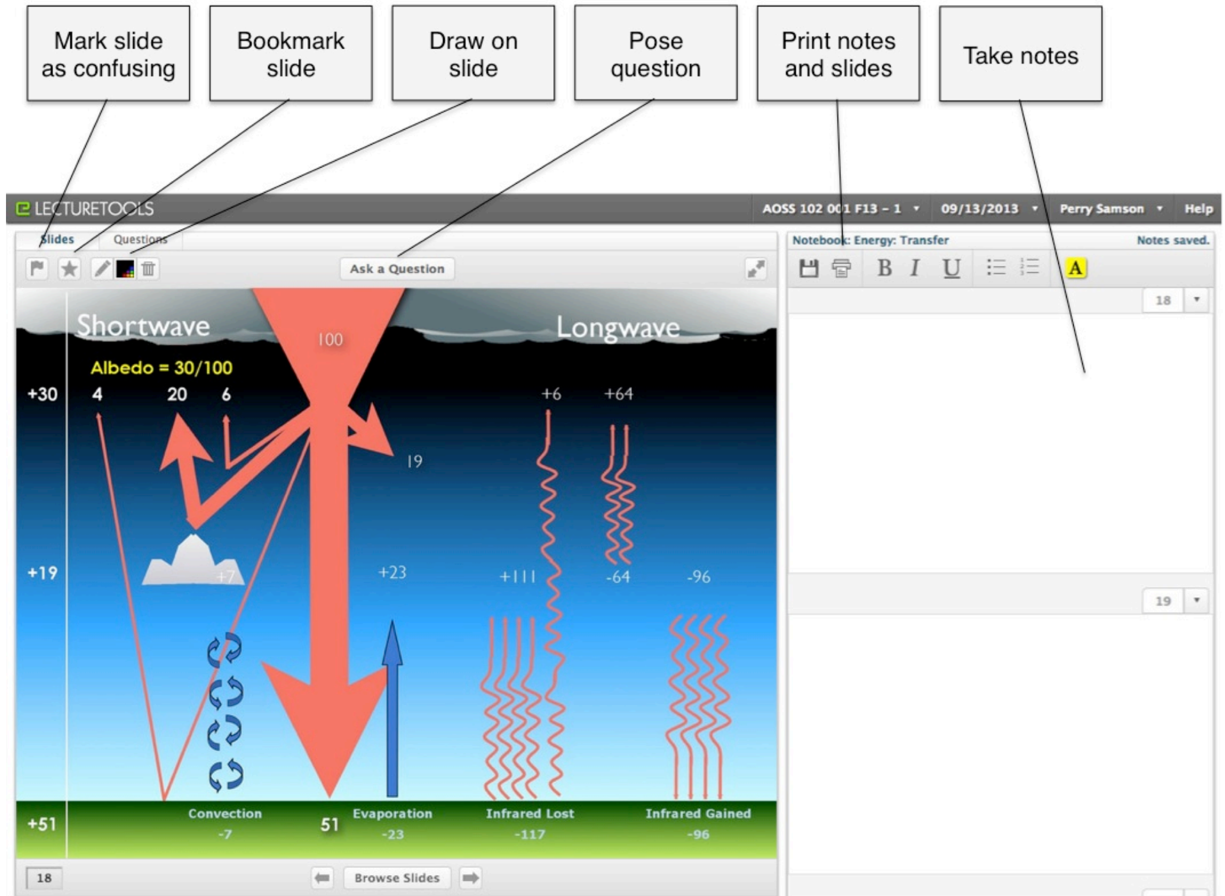

Figure 3. Student view of LectureTools showing various functions students have available to promote participation in class.

Data Mining

Recent national and local reports such as the 2010 report, *A Roadmap for Education Technology* (Woolf, 2010), and the 2012 report, *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief* (Bienkowski et al., 2012), describe the need for increasing the use of educational data mining and learning analytics in order to personalize education and improve teaching and learning. As Technology Enhanced Learning (TEL) tools have become ubiquitous in higher education, a bulk of real-time student behavior data can be captured, broadening opportunities for study and impact of Educational Data Mining (EDM) and Learning Analytics techniques. The *Horizon Report* (Johnson et al., 2013) describes the goal of learning analytics as enabling instructors and institutions to modify educational opportunities and to personalize feedback to each student based on his/her own needs and abilities. Learning analytics models could be used, for example, to predict student-learning performances and to identify student at risk in real time and therefore increase their possibility of success (Arnold, 2010; EDUCAUSE, 2010; Johnson et al. 2011).

Knowledge discovered through educational data mining is used not only to provide feedback to learners, but also to help instructors to manage their classes, understand their students' learning processes, and reflect on their own teaching (Merceron and Yacef, 2005, Romero and Ventura, 2007, Baker and Yacef, 2009, Baker, 2010) Several Educational Data Mining studies of student behavior in online and other educational tools revealed differences between groups of students in terms of such variables as level of participation in discussion boards (Anaya and Boticario, 2009), Questions & Answers boards, completion of assignments, and annotations (Zakrzewska, 2008, Anaya and Boticario, 2009, Macfadyen and Dawson, 2010). Each of these studies has helped to validate these techniques as methods of identifying pedagogically interesting cohorts of students based on their activity with educational technologies.

Figure 4 offers a schematic of the flow in many large survey courses. Before the semester begins the instructor might offer a reading or video that illustrates points to be discussed in class. In class the instructor will present content and optionally ask questions of the students to assess their understanding and/or invite discussion. Following that class the instructor may offer homework, assign readings or video recordings that either review material covered or prepare for the next class session. This cycle continues until a test or quiz is given which often triggers summative review by the students.

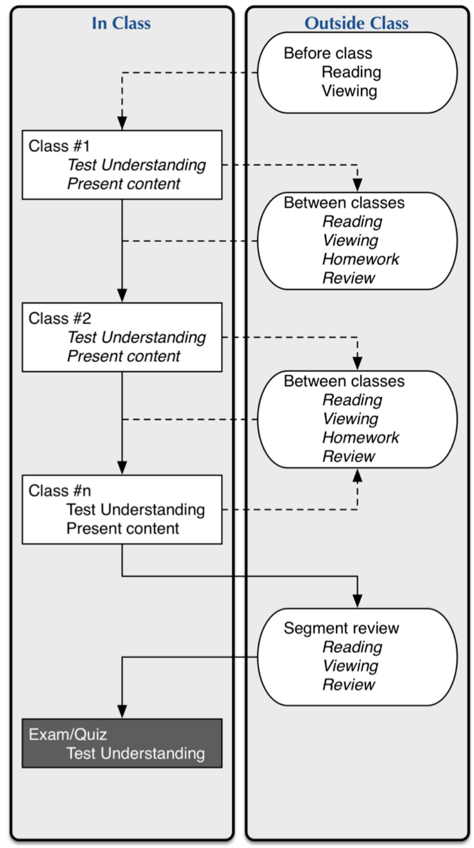

Figure 4. A schematic of the workflow of a course including tasks performed in class and those performed outside of the classroom.

The dual challenge of providing a solid discipline foundation for STEM majors and creating understanding and engagement for non-STEM majors requires a commitment by both groups to participate meaningfully in course activities. Unfortunately few STEM instructors really know how their students behave either in or outside the classroom so offering meaningful guidance about desired study habits is often based on self-reported information from the student who may be reluctant to be totally honest about their effort, especially before they receive their final grades. Moreover, an instructor's advice to students is often informed by their experience as a student and may not represent the best advice for students from a different generation and a different set of background skills and motivations. The end result is that introductory STEM instructors are limited to a post-hoc analysis of student learning challenges, and often advise students without understanding the particular circumstances students are in or goals that they have.

What if, on the other hand, the instructor had an objective and detailed view of each student's behavior with course material as the course was being taught? If the instructor could understand such a mass of data, they could tailor course content, reviews, interactive sessions, assignments, and exams to the needs and desires of the student body. Taking advantage of real-time access to this data, instructors could identify meaningful cohorts based on behavior, researching variability within a cohort to identify factors contributing to poor outcomes, and make actionable teaching activities aimed at strengthening student learning. To make such a task tractable, an instructor would need high fidelity (and pedagogically relevant) student-computer interaction data, a tool or methods by which to summarize this data quickly and effectively, and flexible course delivery that allowed for near real-time adjustment of pedagogical techniques.

LectureTools records a unique and broad spectrum of on-line student activities during and outside classes, including:

1. Notes written on a per student per slide basis as the lecture is delivered (students can opt out if they wish),
2. Student responses to instructor questions on a per student per question basis,
3. Correctness of student answers (when appropriate)
4. Student bookmarking of slides as important or confusing,
5. Student annotations on slides, and
6. Questions posed by students to their instructor.

Together, these technologies cover many of the typical learning tasks described in Figure 4 and offer a database of activities that can be compared with learning outcomes to try to identify relationships.

Additionally, students in the winter 2014 semester were asked to identify their emotional and physical state at the beginning of each lecture. This question was posed with the hypothesis that physical and emotional stresses may influence student performance. Results for one particular day are shown in Figure 5 and illustrate a high degree of collinearity between self-reported emotional and physical conditions.

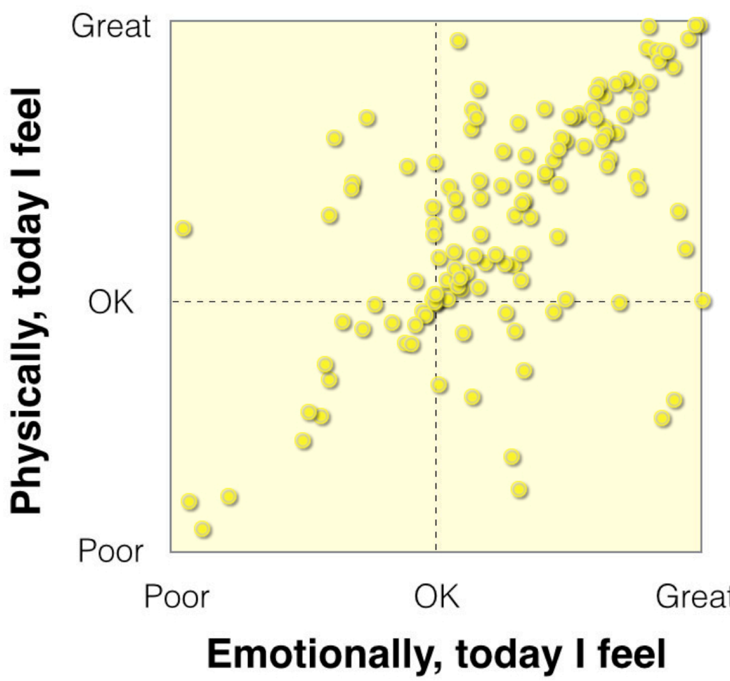

Figure 5. An example of student self reports to daily request "Where on this wellness chart would you put yourself TODAY?" Note collinearity between reported physical and emotional wellness.

Study Guides

One initial outcome of this research has been the generation of student study guides based on the mining of students' notes. Note are "sniffed" in real-time and word clouds (called "Lecture Clouds") are created with greater weight given to a list of keywords defined by the instructor. After class students can view the Lecture Clouds summarized by lecture (Figure 6a) or by slide within a lecture (Figure 6b).

LECTURETOOLS

SESSIONS *Select session for more detail*	FIRST SLIDE	AGGREGATED CLASS NOTES PER SESSION
Wed, Apr 17, 2013 *Weather in Cinema*	Weather in Cinema	tornado wind direction wrong away lightning hail water center cloud like moving pressure rotating tornadoes clear inside storm horizontal
Mon, Apr 15, 2013 *Tornadoes II*	Tornadoes: The Chase	tornado wind rotation divergence pressure instability rain tornadoes cloud convective strong shear moving storm downdraft trigger right center radar winds
Fri, Apr 12, 2013 *Tornadoes*	Tornadoes	tornadoes tornado wind ground shear different rotation kansas rate atmosphere cloud lapse occur coriolis near trigger florida moist convective rise
Wed, Apr 10, 2013 *Derecho and Downbursts*	Hail to the victors	ground rain winds storm microburst damage downburst moving precipitation wind cloud strong gust line center direction heavy thunderstorm derecho
Mon, Apr 8, 2013 *Lightning*	Lightning	lightning cloud ground charge stroke thunder frame metal positive sound electric hear negative return dust strikes potential channel sand hailstone
Fri, Apr 5, 2013 *Weather Forecasting*	Weather Forecasting	range forecasting weather look models wind looking days clouds long short cloud snow atmosphere forecast humidity cold cover temp
Wed, Apr 3, 2013 *Thunderstorms*	Thunderstorms	rain cloud thunderstorm storm updraft ground wind gust cold tornado winds shelf downburst warm line rate severe just anvil green
Mon, Apr 1, 2013 *Heat of the Day*	HEAT WAVE	ground thunderstorms radiation temperature energy heat temp heating rain sunlight time noon cloud losing shortwave uneven florida gaining rate temperatures longwave
Fri, Mar 29, 2013 *Anticyclogenesis and Heat Waves*		heat pressure high sinking night temperature time lower cloud humidity temp water relative ground wave wind cover energy cities urban
Wed, Mar 27, 2013 *Linking the Jet Stream to Cyclogenesis*		divergence pressure wave advection storm atmosphere cold warm winds ground upper short areas stream convergence happens tropical wind east strong temperature

Figure 6a. A "Lecture Cloud" of words typed by students during class. Two categories of words are offered, those included in the list of keyterms provided by the instructor and those words not in the list of keyterms. The words are each automatically linked to external resources (e.g. Wikipedia, YouTube).

LECTURETOOLS

Wednesday, April 3rd, 2013 *Thunderstorms* List of Lectures			Total Words Typed This Class	Aggregated Whole Class Notes	Number Confused	Number Bookmarked	Number Annotated
			4799	rain cloud thunderstorm storm updraft ground wind gust cold tornado winds shelf downburst warm line rate severe just anvil green	0	7	50
Slide #	**View**	**# Students Taking Notes**	**Words Typed per Slide**	**Word Cloud of Notes by Slide**	**Total Number Confused**	**Total Bookmarked**	**Total Annotated**
1		4	24	*Too few words written to analyze*	0	0	0
2		5	45	fronts thunder air mass air mass thunderstorm gust front gust converging cool formation land mass	0	0	0
3		3	21	*Too few words written to analyze*	0	0	0
4		4	44	tornadoes severe severe thunderstorm thunder tornado wind storm thunderstorm makes approaching considered	0	1	0
5		14	334	inversion atmosphere lapse rate unstable need thunderstorms lapse rate rapid rising severe thunderstorm thunder thunderstorm	0	1	0
6		10	252	layer warm cold moist rate saturation ground inversion convective instability cool lapse lapse rate	0	1	0
7		20	1074	updraft thunderstorm rain cloud ground storm severe fall precipitation makes angle flowing rising cold	0	2	20

(Hall et al., 2009)

Figure 6b. The "Lecture Cloud" displayed on a per side level. This view affords a view of which slides produced the most student notes and which slides were most annotated or bookmarked.

Wellness

Student self-reports of emotional and physical state were used to cluster students into similar patterns through the semester. Using Weka (Hall et al., 2009) the emotional and physical states reported prior to the first exam were clustered with an inflection point happening at nine clusters. Figure 7 shows the result that student grades on the first exam were well correlated with both the reported physical and emotions state of the students.

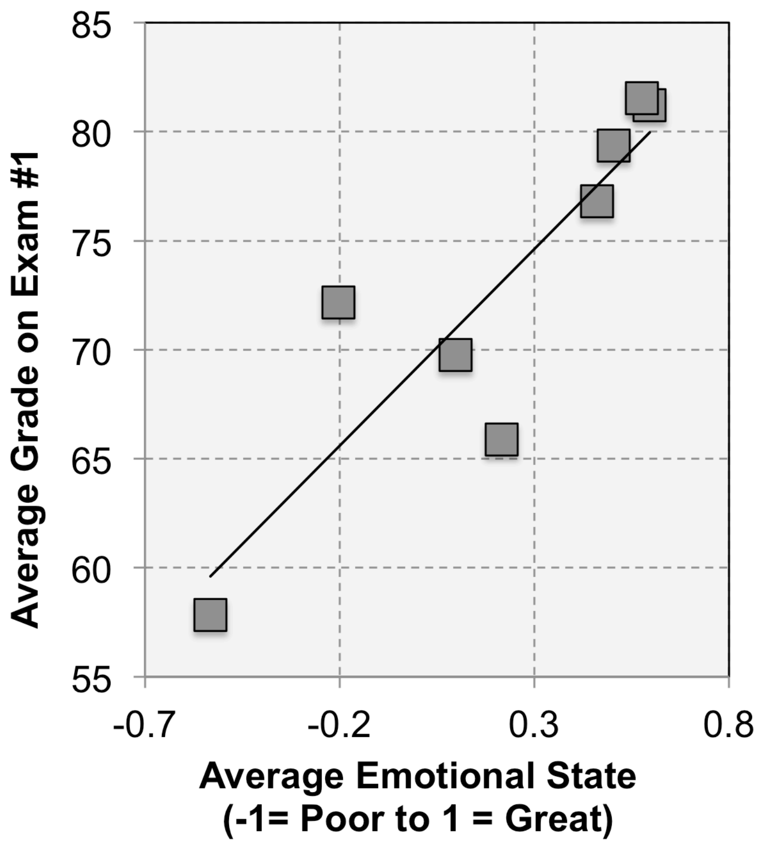

Figure 7. Average exam scores on exam#1 in AOSS 102 averaged by nine groups obtained by clustering their daily self-reported emotional state.

Thoughts

This work illustrates that tools designed to be integral to class conduct can, in fact, increase students' perceptions of engagement positively. When students are given the opportunity tom participate in class, and especially large survey courses, they will. The key here is providing tools that give instructors more opportunities to involve students actively in class through challenging questions and responding to student questions.

The work on mining the data from this system is still in its infancy. Students have anecdotally, warmly received the creation of study guides based on student note taking. They are particularly interested in having the words linked to resources that challenge their understanding on the concept. To this end the system was expanded to link words to the page in their eTextbook that is best matched to the concept.

It remains a challenge to demonstrate whether these interventions have led to deeper student learning. The variation in student outcomes, as measured by grades, are due to many factors that make it difficult to identify the effect of a specific tool. Continued research will cluster students who participate in class in the same way to see if variations within a cohort of "similar" students can allow a firmer understanding of the impact of specific interventions.

One initial clustering effort, based on student self-reports of physical and emotional state demonstrates a strong relationship in outcomes and emotional state. While this is not necessarily surprising this result raises questions about what responsibility do instructors have to identify students having emotional distress? And, once identified, what are the best strategies for dealing with the students who score low in self reported wellness?

References

Anaya, A. R. and J. G. Boticario (2009). A Data Mining Approach to Reveal Representative Collaboration Indicators in Open Collaboration Frameworks. 2nd International Conference On Educational Data Mining. Cordoba, Spain.

Baker, R. S. J. d. (2010). Data Mining. International Encyclopedia of Education (Third Edition). P. Peterson, E. Baker and B. McGaw. Oxford, Elsevier**:** 112-118.

Baker, R. S. J. D. and K. Yacef (2009). "The State of Educational Data Mining in 2009: A Review and Future Visions " Journal of Educational Data Mining **1**(1).

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten (2009). "The WEKA Data Mining Software: An Update." SIGKDD Explorations **11**(1).

Johnson, L., S. Adams-Becker, M. Cummins, V. Estrada, A. Freeman and H. Ludgate (2013). NMC Horizon Report: 2013 Higher Education Edition. Austin, Texas, The New Media Consortium.

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

Merceron, A. and K. Yacef (2005). Educational data mining: A case study. International Conferences on Artificial Intelligence in Education,, Amsterdam, The Netherlands.

Romero, C. and S. Ventura (2007). "Educational data mining: A survey from 1995 to 2005." Expert Systems with Applications **33**(1): 135-146.

Samson, P. (2010). "Deliberate Engagement of Laptops in Large Lecture Classes to Improve Attentiveness and Engagement." Computers in Education **20**(2).

Zakrzewska, D. (2008). Cluster Analysis for Users' Modeling in Intelligent E-Learning Systems. New Frontiers in Applied Artificial Intelligence. N. Nguyen, L. Borzemski, A. Grzech and M. Ali, Springer Berlin Heidelberg. **5027:** 209-214.