SOLAR
SOCIETY for LEARNING ANALYTICS RESEARCH

The 4th International Conference on

# Learning Analytics and Knowledge

March 24-28, 2014

INDIANAPOLIS
INDIANA, USA

## Conference Theme

### *Intersection of learning analytics research, theory and practice*

The International Learning Analytics and Knowledge conference is now in its fourth year! LAK 14 will keep up the momentum generated in the earlier meetings as we continue to define who we are as a community and how our research can inform the important decisions facing education today. We are experiencing conditions that create a "perfect storm" for making analytics vitally important for supporting learning in our information age. Not only do we know more about the basic processes involved in learning, the explosion of online technology platforms and services provide new opportunities for learning and generate new forms of data about those processes. Coupled with increasing societal demands to make learning more accessible and effective for all, we need scholarly research to ensure that advances in learning technologies and practices benefit society as a whole. Advances in analytics, data mining and data visualization promise a way forward for making sense of the massive amounts of data available so that we can thoughtfully inform learners, educators, and policy makers.

The location for the conference this year, Indianapolis, is called "The City at the Crossroads of America." Accordingly, the theme for LAK 14 focuses on the crossroads, or intersection, of basic research on learning, learning theory, and educational practice. We invite papers that address at least one of these foci, but especially encourage submissions that demonstrate how they fit together in ways that make our work relevant to improving learning for people of all ages, leverage the technologies available to today's learners, and scalable within the many places where learning occurs. To accomplish these goals, we seek theoretically motivated papers that utilize analytics to provide direction for understanding and supporting learning.

## Sponsors

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

PURDUE UNIVERSITY

bluecanary

Intellify LEARNING

Desire2Learn

## LAK 2014 Program Chairs' Welcome

It is our pleasure to welcome you to LAK 2014. This is the fourth meeting of Learning Analytics and Knowledge, and this conference has quickly become an established meeting place for top researchers from learning sciences, data mining, learning analytics, computer sciences and other fields. Since the first conference held in 2011, the Learning Analytics community has grown and matured. The annual meetings of LAK are convened by an international professional society, the Society for Learning Analytics Research (SoLAR), which has also just this year launched a new scholarly journal, the Journal of Learning Analytics. The rapid growth of this field and the public attention drawn to the explosion of online technology platforms and services has convinced many of us that we are engaged in an exciting area that holds great promise for innovating teaching and learning.

The theme of the conference this year, the "Intersection of Learning Analytics Research, Theory and Practice", represents the location of the meeting, Indianapolis, called "The City at the Crossroads of America", as well as our intent to select papers that represent the intersection of basic research on learning, learning theory, and educational practice. While the term analytics generally describes the discovery and communication of meaningful patterns in data, when we talk about Learning Analytics we aim for discovery and communication that is not only meaningful but also actionable. What we do is in service of the belief that all learners can be successful learners, and it is our responsibility to understand and support the conditions that make this true. In addition, our work needs to be able to thoughtfully inform educators, administrators, and policy makers so we can help learning to be more accessible and effective for everyone. The papers included in these proceedings responded to this call.

The submissions to the LAK 14 conference continued the strong trajectory of the first three meetings. This year we received a total of 44 full paper submissions from which 13 were accepted (27% acceptance rate). The conference program also includes 25 short papers, 3 panel sessions, and 1 poster session with 11 posters representing authors from 26 countries. Our final program was selected with the help of the program committee and reviewers who produced 250 reviews. We thank these colleagues, as well as all of the submitting authors, for their assistance in making a strong conference program. We are also grateful to have a Doctoral Consortium this year with 9 students participating who come from Ph.D. programs in institutions in the US, Canada, and Europe. We especially want to thank our colleagues who held various chair positions for the conference without whom our job would have been measurably more difficult and certainly less enjoyable.

What ultimately makes a conference program successful is its impact after the conference itself is over. We hope that the papers presented here combined with the conference experience in Indianapolis spark lively debate about how we can effectively and ethically utilize "big data" about learning and help us collectively expand the impact of our field. As we said in our original call for papers, we are in the middle of "perfect storm" for making analytics vitally important for supporting learning in our information age. Now more than ever we need scholarly research to ensure that advances in learning technologies and practices benefit society as a whole. We hope you enjoy the LAK conference and being part of this exciting effort.

**Abelardo Pardo**
*The University of Sydney, Australia*

**Stephanie D. Teasley**
*University of Michigan, MI, USA*

# Schedule at A Glance

Registration will be open 8:00-4:30 March 24 and March 25, and 7:00-4:30 March 26, 27, and 28 in Grand Foyer 1-3 near the escalator.

**PRE-CONFERENCE WORKSHOPS AND TUTORIALS**

| Time | March 24 | March 25 |
|---|---|---|
| 9:00 AM – 4:00 PM | Computational Approaches to Connecting Levels of Analysis in Networked Learning (Senate 1)<br><br>The Second International Workshop on Discourse-Centric Learning Analytics (Senate 2&3) | Learning Analytics and Machine Learning (House)<br><br>Tutorial on Epistemic Network Analysis (Senate 2)<br><br>Doctoral consortium (Senate 3) |
| 9:00 AM – 1:00 PM | Tutorial on Introduction to Data Mining (House) | |
| 12:00 PM – 4:00 PM | | The LAK Data Challenge 2014 (Senate 1) |

**CONFERENCE SESSIONS**

| Time | March 26 | March 27 | March 28 |
|---|---|---|---|
| 7:00 AM – 8:00 AM | Breakfast (Grand 5) | Breakfast (Grand 5) | Breakfast (Grand 5) |
| 8:00 AM – 9:15 AM | **Keynote Session (Grand 5)**<br>Art Graeser<br>*Scaling Up Learning by Communicating with AutoTutor, Trialogs, and Pedagogical Agents* | **Keynote Session (Grand 5)**<br>Nancy Law<br>*Is Learning Analytics a Disruptive Innovation?* | **Keynote Session (Grand 5)**<br>Scott Klemmer<br>*Design at Large* |
| 9:15 AM – 9:30 AM | Break (Grand Foyer) | Break (Grand Foyer) | Break (Grand Foyer) |
| 9:30 AM – 10:45 AM | 1A. Discussion with keynote speaker (Grand 2)<br><br>1B. Process Mining (Grand 3) | 5A. Discussion with keynote speaker (Grand 2)<br><br>5B. MOOCs (Grand 3) | 9A. Discussion with keynote speaker (Grand 2)<br><br>9B. Discourse and Argumentation (Grand 3) |
| 10:45 AM – 12:00 PM | 2A. Panel. Setting Learning Analytics in Context: Overcoming the Barriers to Large-Scale Adoption (Grand 2)<br><br>2B. Predictive models and recommendations (Grand 3) | 6A. Panel. Learning Analytics for the Social Media Age (Grand 2)<br><br>6B. Learning Analytics for "at risk" students (Grand 3) | 10A. Panel: Building Institutional Capacities and Competencies for Systemic Learning Analytics Initiatives (Grand 2)<br><br>10B. Who we are and who we want to be (Grand 3) |
| 12:00 PM – 1:00 PM | Lunch (Grand 5) | Lunch and SoLAR General Meeting (Grand 5) | Closing Session (Grand 5) |
| 1:00 PM – 1:30 PM | | | Lunch on your own |
| 1:30 PM – 3:00 PM | 3A. Alternative Analytics (Grand 2)<br><br>3B. Learning Mathematics (Grand 3) | 7A. Text analytics and collaborative environments (Grand 2)<br><br>7B. Institutional Perspectives (Grand 3) | |
| 3:00 PM – 3:15 PM | Break (Grand Foyer) | Break (Grand Foyer) | |
| 3:15 PM – 4:30 PM | 4. Posters (Grand 5) | 8A. Analysis of Resource use in LMS (Grand 2)<br><br>8B. Learning Analytics and Learning Design (Grand 3) | |
| 4:30 PM – 5:30 PM | Reception (Grand 4) | | |
| 6:00 PM | | Conference Dinner at the Eiteljorg Museum of American Indians and Western Art | |

**Map of the venue is found on page 22**

■ Plenary Sessions



# Keynote Speaker – Dr. Art Graesser

## *"Scaling Up Learning by Communicating with AutoTutor, Trialogs, and Pedagogical Agents"*

Dr. Art Graesser is a professor in the Department of Psychology and the Institute for Intelligent Systems at the University of Memphis, as well as an Honorary Research Fellow at University of Oxford. He received his Ph.D. in psychology from the University of California at San Diego. His primary research interests are in cognitive science, discourse processing, computational linguistics, and the learning sciences. More specific interests include knowledge representation, question asking and answering, tutoring, text comprehension, inference generation, conversation, reading, education, memory, artificial intelligence, and human-computer interaction. He served as editor of the journal Discourse Processes (1996–2005) and is the current editor of Journal of Educational Psychology. His service in professional societies includes president of the Empirical Studies of Literature, Art, and Media (1989-1992), the Society for Text and Discourse (2007-2010), the International Society for Artificial Intelligence in Education (2007-2009), and the Federation of Associations in the Behavioral and Brain Sciences Foundation (2012-13). In addition to publishing over 500 articles in journals, books, and conference proceedings, he has written two books and edited 12 books. He and his colleagues have designed, developed, and tested software in learning, language, and discourse technologies, including AutoTutor, AutoTutor-lite, MetaTutor, GuruTutor, DeepTutor, HURA Advisor, SEEK Web Tutor, Operation ARIES!, iSTART, Writing-Pal, Point & Query, Question Understanding Aid (QUAID), QUEST, & Coh-Metrix.

# Keynote Speaker – Dr. Nancy Law

## *"Is Learning Analytics a Disruptive Innovation?"*

Professor Nancy Law is currently Deputy Director of the Centre for Information Technology in Education (CITE) in the Faculty of Education of the University of Hong Kong, after serving as its Founding Director for 15 years from 1998. She serves/has served on a number of policy advisory boards/working groups related to ICT in education for the University of Hong Kong, the Hong Kong government and other community groups. She was a core member of the International Study Centre for the SITES 2006 (Second Information Technology in Education Study 2006), served on the Editorial and Publication Committee of the IEA and the Technology Working Group of the Cisco-Intel-Microsoft Project on Assessment and Teaching of 21st Century Skills, the Board of Directors of the International Society of the Learning Sciences, the Editorial Boards of the International Journal of Computer-Supported Collaborative Learning and the International Journal of Web Based Communities. She has also been invited to provide expert input/consultancy to the European Commission, UNESCO and OECD on various aspects of technology-enhanced learning. Her research interests include international comparative studies of pedagogical innovations and information technology, models of ICT integration in schools and change leadership, computer supported collaborative learning and the use of expressive and exploratory computer-based learning environments.

# Keynote Speaker – Dr. Scott Klemmer

## *"Design at Large"*

Scott is an Associate Professor of Cognitive Science and Computer Science & Engineering at UC San Diego, and a Visiting Associate Professor of Computer Science at Stanford. He previously served as an Associate Professor of Computer Science at Stanford, where he co-directed the Human-Computer Interaction Group and held the Bredt Faculty Scholar chair. He has a dual BA in Art-Semiotics and Computer Science from Brown (with studio work at RISD), and a PhD in Computer Science from Berkeley. The open-source design tools and curricula created in his lab have been adopted by organizations around the world. He has served on advisory boards for design programs (like CCA), research labs (like DoCoMo), and startups. His former graduate students are leading professors (including Berkeley, CMU, and UIUC), researchers (at Adobe), founders (including Instagram and Pulse), social entrepreneurs, and engineers. He helped introduce peer assessment to open online education, and taught the first peer-assessed online course. He has been awarded the Katayanagi Emerging Leadership Prize, Sloan Fellowship, NSF CAREER award, and Microsoft Research New Faculty Fellowship. He has authored and co-authored more than 40 peer-reviewed articles; eight were awarded best paper or honorable mention at the premier HCI conferences (CHI/UIST/CSCW).

# Pre-Conference Schedule

## MONDAY, MARCH 24

**9:00 AM – 1:00 PM | House**

***Introduction to Data Mining for Educational Researchers***
Christopher Brooks and Zachary Pardos

**9:00 AM – 4:00 PM | Senate 1**

***Computational Approaches to Connecting Levels of Analysis in Networked Learning Communities***
H. Ulrich Hoppe and Daniel D. Suthers

**9:00 AM – 4:00 PM | Senate 2 & 3**

***The Second International Workshop on Discourse-Centric Learning Analytics***
Rebecca Ferguson, Anna De Liddo, Denise Whitelock, and Maarten de Laat

## TUESDAY, MARCH 25

**9:00 AM – 4:00 PM | House**

***Learning Analytics and Machine Learning***
George Siemens, Carolyn Rose, Dragan Gasevic, Annika Wolff, and Zdenek Zdrahal

**9:00 AM – 4:00 PM | Senate 2**

***Tutorial on Epistemic Network Analysis***
Wesley Collier, David Shaffer, and Chandra Orrill

**9:00 AM – 4:00 PM | Senate 3**

***Doctoral Consortium***
Chris Quintana, Alyssa Wise, Katherine Maillet, Stephanie Teasley

**12:00 PM – 4:00 PM | Senate 1**

***LAK Data Challenge 2014***
Stefan Dietze, Mathieu D'Aquin, Taibi Davide, Eelco Herder and Hendrik Drachsler





# Program Schedule

## WEDNESDAY, MARCH 26

**8:00 AM – 9:15 AM | Keynote | Grand 5**
*Sponsored by: Blue Canary, Session Chair: Stephanie Teasley*

***Scaling Up Learning by Communicating with AutoTutor, Trialogs, and Pedagogical Agents***
Art Graesser

Learning has occurred over a few millennia by the learner communicating with the teacher, tutor, master, or mentor in natural language. Apprenticeship learning has always occurred one-on-one or in small groups with an expert. Researchers in the discourse and learning sciences have documented the conversation patterns that occur in these interactions. Researchers in computational linguistics, artificial intelligence, and intelligent tutoring systems have developed computer agents that simulate many of these conversation patterns and help people learn. This is the moment in history when these systems are being launched on the internet.

This presentation will present recent systems on the internet that help students learn by holding a conversation in natural language. AutoTutor engages in dialogue with the student on a variety of subject matters in Science, Technology, Engineering, and Mathematics. Trialogs are conversations between the human students and two computer agents, typically a student agent and a tutor agent. Students can either observe two agents interact vicariously, interact with a tutor agent as a student agent periodically chimes in, or teach a student agent while a tutor rescues a problematic interaction. Agents can argue with each other over issues and ask what the human students think about the argument. Trialogs are being developed for the Internet in serious games with Pearson Education (Operation ARA), in assessments with Educational Testing Service, and in a new Center for the Study of Adult Literacy for struggling adult readers. AutoMentor is being developed for computer mediated communication between a mentor agent and small groups of students in a simulation game on urban planning. Tests of these systems have shown very encouraging learning gains. We are currently conducting data mining analyses to explore the features of conversation and learning experiences that predict learning, motivation, and student emotions.

**9:15 AM – 9:30 AM | Morning Break | Grand Foyer**

**9:30 AM – 10:45 AM | Discussion with Art Graesser | Grand 2**
*Session Chair: Leah MacFayden*

**9:30 AM – 10:45 AM | Process Mining | Grand 3**
*Session Chair: Dragan Gašević*

***Formative Assessment Method of Real-world Learning by Integrating Heterogeneous Elements of Behavior, Knowledge, and the Environment (Full)***
Masaya Okada, Masahiro Tada

Real-world learning in a field is an important educational area for experience-based activities. Formative assessment by constant monitoring of the intellectual achievement of real-world learners is essential for adaptive learning support, but no assessment methodology has yet been developed. We consider a method to systematically integrate heterogeneous factors of real-world learning: learners' internal situations, their external situations, and their learning field. Then, we propose a method for formatively assessing the situation of real-world learning. The method enables us to recognize the sequence of characteristic stay behavior and the associated body posture of a learner, and to estimate the 3D location of his/her interest. The method enables the estimation of not only the learning topic that a learner is currently examining in a field but also the prospective topics that he/she should learn. Our assessment method is the basis for context-aware support to promote the emergence of new knowledge from intellectual collaboration in the world.

***Clustering for improving Educational Process Mining (Short)***
Alejandro Bogarín, Cristóbal Romero, Rebeca Cerezo, Miguel Sánchez-Santillán

In this paper, we propose to use clustering to improve educational process mining. We want to improve both the performance and comprehensibility of the models obtained. We have used data from 84 undergraduate students who followed an online course using Moodle 2.0. We propose to group students firstly starting from data about Moodle's usage summary and/or the students' final marks in the course. Then, we propose to use data from Moodle's logs about each cluster/group of students separately in order to be able to obtain more specific and accurate models of students' behaviour. The results show that the fitness of the specific models is greater than the general model obtained using all the data, and the comprehensibility of the models can be also improved in some cases.

**10:45 AM – 12:00 PM | Panel | Grand 2**
*Session Chair: Lori Lockyer*

**Rebecca Ferguson, Doug Clow, Leah Macfadyen, Shane Dawson, Alfred Essa, Shirley Alexander**
*Setting Learning Analytics in Context: Overcoming the Barriers to Large-Scale Adoption*
Once learning analytics have been successfully developed and tested, the next step is to implement them at a larger scale – across a faculty, an institution or an educational system. This introduces a new set of challenges, because education is a stable system, resistant to change. Implementing learning analytics at scale involves working with the entire technological complex that exists around technology-enhanced learning (TEL). This includes the different groups of people involved – learners, educators, administrators and support staff – the practices of those groups, their understandings of how teaching and learning take place, the technologies they use and the specific environments within which they operate. Each element of the TEL Complex requires explicit and careful consideration during the process of implementation, in order to avoid failure and maximise the chances of success. In order for learning analytics to be implemented successfully at scale, it is crucial to provide not only the analytics and their associated tools but also appropriate forms of support, training and community building.

**10:45 AM – 12:00 PM | Predictive Models and Recommendations (Full) | Grand 3**
*Session Chair: Hendrik Drachsler*

***Customized Course Advising: Investigating Engineering Student Success with Incoming Profiles and Patterns of Concurrent Course Enrollment (Full)***
SungJin Nam, Steven Lonn, Thomas Brown, Cinda-Sue Davis, Darryl Koch
Every college student registers for courses from a catalog of numerous offerings each term. Selecting which courses in which to enroll, and in what combinations, can dramatically impact each student's chances for academic success. Taking inspiration from the STEM Academy, we wanted to identify the characteristics of engineering students who graduate with 3.0 or above grade point average. The overall goal of the Customized Course Advising project is to determine the optimal term-by-term course selections for all engineering students based on their incoming characteristics and previous course history and performance, paying particular attention to concurrent enrollment. We found that ACT Math, SAT Math, and Advanced Placement exam can be effective measures to measure the students' academic preparation level. Also, we found that some concurrent course-enrollment patterns are highly predictive of first-term and overall academic success.

***Explaining Predictive Models to Learning Specialists using Personas (Short)***
Christopher Brooks, Jim Greer
This paper describes a method we have developed to convert statistical predictive models into visual narratives which explain student classifications. Building off of the work done within the user experience community, we apply the concept of personas to predictive models. These personas provide familiar and memorable descriptions of the learners identified by data mining activities, and bridge the gap between the data scientist and the learning specialist.

***Temporal Learning Analytics for Computer Based Testing (Short)***
Zacharoula K. Papamitsiou, Vasileios Terzis, Anastasios A. Economides
Predicting student's performance is a challenging, yet complicated task for institutions, instructors and learners. Accurate predictions of performance could lead to improved learning outcomes and increased goal achievement. For that reason, prediction of performance is acknowledged as one of the major objectives of Learning Analytics and Educational Data Mining research. In this paper we explore the predictive capabilities of student's time-spent on answering (in-)correctly each question of a multiple-choice assessment quiz, along with student's final quiz-score, in the context of computer-based testing. We also explore the correlation between the time-spent factor (as defined here) and goal-expectancy. We present a case study and investigate the value of using this parameter as a learning analytics factor for improving prediction of performance during computer-based testing. Our initial results are encouraging and indicate that the temporal dimension of learning analytics should be further explored.

**12:00 PM – 1:30 PM | Lunch | Grand 5**

**1:30 PM – 3:00 PM | Alternative Analytics (Full) | Grand 2**
*Session Chair: Xavier Ochoa*

***Sleepers' Lag - Study on Motion and Attention (Full)***
Mirko Raca, Roland Tormey, Pierre Dillenbourg
Human body-language is one of the richest and most obscure sources of information in inter-personal communication which we aim to re-introduce into the classroom's ecosystem. In this paper we present our observations of student-to-student influence and measurements. We show parallels with previous theories and formulate a new concept for measuring the level of attention based on synchronization of student actions. We observed that the students with lower levels of attention are slower to react then focused students, a phenomenon we named "sleepers' lag".

***Clustering of Design Decisions in Classroom Visual Displays (Short)***
Ma. Victoria Almeda, Peter Scupelli, Ryan Baker, Mimi Weber, Anna Fisher
In this paper, we investigate the patterns of design choices made by classroom teachers for decorating their classroom walls, using cluster analysis to see which design decisions go together. Classroom visual design has been previously studied, but not in terms of the systematic patterns adopted by teachers in selecting what materials to place on classroom walls, or in terms of the actual semantic content of what is placed on walls. This is potentially important, as classroom walls are continuously seen by students, and form a continual off-task behavior option, available to students at all times. Using the k-means clustering algorithm, we find four types of visual classroom environments (one of them an outlier within our data set), representing teachers' strategies in classroom decoration. Our results indicate that the degree to which teachers place content-related decorations on the walls, is a feature of particular importance for distinguishing which approach teachers are using. Similarly, the type of school (e.g. whether private or charter) appeared to be another significant factor in determining teachers' design choices for classroom walls. The present findings begin the groundwork to better understand the impact of teacher decisions and choices in classroom design that lead to better outcomes in terms of engagement and learning, and finally towards developing classroom designs that are more effective and engaging for learners.

***Data Wranglers: Human Interpreters to Help Close the Feedback Loop (Short)***
Doug Clow
Closing the feedback loop to improve learning is at the heart of good learning analytics practice. However, the quantity of data, and the range of different data sources, can make it difficult to take systematic action on that data. Previous work in the literature has emphasised the need for and value of human meaning-making in the process of interpretation of data to transform it in to actionable intelligence.

***Toward Unobtrusive Measurement of Reading Comprehension Using Low-Cost EEG (Short)***
Yueran Yuan, Kai-Min Chang, Jessica Nelson Taylor, Jack Mostow
Assessment of reading comprehension can be costly and obtrusive. In this paper, we use inexpensive EEGs to detect reading comprehension of readers in a school environment. We use EEG signals to produce above-chance predictors of student performance on end-of-sentence cloze questions. We also attempt (unsuccessfully) to distinguish among student mental states evoked by distracters that violate either syntactic, semantic, or contextual constraints. In total, this work investigates the practicality of classroom use of inexpensive EEG devices as an unobtrusive measure of reading comprehension.

**1:30 PM – 3:00 PM | Learning Mathematics | Grand 3**
*Session Chair: Agathe Merceron*

***Learning Analytics in CSCL with a Focus on Assessment: An Exploratory Study of Activity Theory-Informed Cluster Analysis (Full)***
Wanli Xing, Bob Wadholm, Sean Goggins
In this paper we propose an automated strategy assessing participation in a multi-mode math discourse environment called Virtual Math Teams with Geogrebra (VMTwG). A holistic participation clustering algorithm is presented based on activity theory. Our activity theory-informed algorithm is a first step in accelerating heuristic approaches to assessing collaborative work in technology mediated environments like VMTwG. Exploratory findings provide an example of a novel, time-efficient, valid, and reliable participatory learning assessment tool for teachers in computer mediated learning environments.

WEDNESDAY, MARCH 26

***On Using Markov Chain to Evidence the Learning Structures and Difficulty Levels of One Digit Multiplication (Short)***

Behnam Taraghi, Martin Ebner, Anna Saranti, Martin Schön

Understanding the behavior of learners within learning applications and analyzing the factors that may influence the learn process play a crucial role in designing and optimizing learning applications. In this work we focus on a specific application named "1x1 trainer" that has been designed for primary school children to learn one digit multiplications. We investigate the database of learners' answers to the asked questions (N > 440000) by applying the Markov chains. Considering the user's sequence of asked questions, the answer types to each question represent the states and the probabilities to the answer type of the subsequent question in the sequence as transition links between states. We want to understand whether the learners' answers to the already asked questions can affect the way they will answer to the subsequent asked questions and if yes to what extent. Through our analysis we identify at first the most difficult and easiest multiplications for the target learners by observing the probabilities of the different answer types. Next we try to identify influential structures in the history of learner' answers considering Markov chain of different orders. Based on this analysis we abstract a way from global answer type transitions to transitions between answer types within each difficulty level of questions (difficult, intermediate and easy) individually and show that the structures do not hold for the set of easiest questions. The results are used to identify pupils who have difficulties with multiplications very soon (after couple of steps) and to optimize the way questions are asked for each pupil individually.

***Context Personalization, Preferences, and Performance in an Intelligent Tutoring System for Middle School Mathematics (Short)***

Stephen Fancsali, Steven Ritter

Learners often think math is unrelated to their own interests. Instructional software has the potential to provide personalized instruction that responds to individuals' interests. Carnegie Learning's MATHia software for middle school mathematics asks learners to specify domains of their interest (e.g., sports & fitness, arts & music), as well as names of friends/classmates, and uses this information to both choose and personalize word problems for individual learners. Our analysis of MATHia's relatively coarse-grained personalization contrasts with more fine-grained analysis in previous research on word problems in the Cognitive Tutor (e.g., finding effects on performance in parts of problems that depend on more difficult skills), and we explore associations of aggregate preference "honoring" with learner performance. To do so, we define a notion of "strong" learner interest area preferences and find that honoring such preferences has a small negative association with performance. However, learners that both merely express preferences (either interest area preferences or setting names of friends/classmates), and those that express strong preferences, tend to perform in ways that are associated with better learning compared to learners that do not express such preferences. We consider several explanations of these findings and suggest important topics for future research.

***Interaction Design for Improved Analytics (Short)***

Maria Mendiburo, Brian Sulcer, Ted Hasselbring

In this paper, we explain the design research process that we used to develop the learning analytics for the a fractions intervention program called HALF. In particular, we highlight a set of qualitative interviews that we conducted with individual students after a short study in which students in three classes at the same school learned to use the virtual manipulatives embedded in the technology-enhanced learning environment (TELE) to compare pairs of proper fractions and order groups of 3 proper fractions. During the intervention, the analytics we collected for each of these 7 students indicated that they failed to master the content taught during the intervention, but they provided little insight into why and how the students were struggling. In contrast, the qualitative interviews provided us with considerable information that helped us diagnose and address misconceptions and incomplete understandings. These insights led to design changes for the lessons students who use the most up-to-date version of HALF experience via the TELE, which in turn improved the learning analytics for our system.

**3:00 PM – 3:15 PM | Afternoon Break | Grand Foyer**

**3:15 PM – 4:30 PM | Posters and Demos | Grand 5**

***eGraph tool: graphing the learning process in LMSs***

Rebeca Cerezo, Miguel Sánchez-Santillán, Natalia Suarez, J. Carlos Núnez

eGraph is a virtual tool developed with the aim of make easier to track the students´ learning process in Learning Management Systems like Moodle. It is based in the log files that the learning platform records when the students are interacting with and allows teachers, students, and researchers to track the learning route that learners have followed during a particular time span.

***Visualizing semantic space of online discourse: The Knowledge Forum case***

Bodong Chen

This poster presents an early experimentation of applying topic modeling and visualization techniques to analyze knowledge building discourse. Latent Dirichlet Allocation was used to convert discourse into a high-dimensional space, which was then reduced to two-dimensional by using Locally Linear Embedding for visualization. Time series analysis was applied to track development of different themes over time. Future directions are also discussed.

***National Differences in an International Classroom***

Jennifer DeBoer, Glenda S. Stump

The virtual classrooms of open online courses include students from a vast array of individual, social, economic, and educational contexts. Detailed data were collected for the first course MIT ran on the edX platform, including student behavior, performance, and background information. In this paper, we demonstrate the systematic differences in average performance, distribution of performance, and performance conditional on behaviors for countries with different characteristics (e.g., language, income).

***Effects of Image-based and Text-based Activities on Student Learning Outcomes***

Anne Greenberg, Melissa Gross, Mary Wright

While current research does not support the hypothesis that some people are visual learners while others are verbal learners, research has shown that visual learning formats may be particularly beneficial in course contexts requiring the interpretation of complex visual material [1], such as anatomy or art history. However, this research has relied primarily on lecture-based pedagogy, not accounting for the processing time students need to make sense of both visual and verbal material [6]. In this study, we investigate the potential differential effects of text-based and image-based student learning activities on student learning outcomes in a functional anatomy course. Students are presented with both text- and image-based active learning activities during class lectures, with different concepts being reinforced through each modality. Participation in and performance on these in-class activities are tracked, and the correlation between Lecture- Tools analytics and exam question scores on corresponding material are tested. Preliminary results show that both participation in and performance on in-class image-based activities are correlated with performance on associated exam questions. However, participation in and performance on in-class text-based activities are not correlated with performance on exam questions. Additionally, performance on exam questions associated with in-class text-based activities is correlated with both participation in and performance on image-based in-class activities. While these exam questions were not directly associated with the in-class image- based learning activities, it is possible that image-based activities allowed for additional comprehension that increased student learning even in those subjects taught using text- based activities. Implications for teaching and learning include greater use of active learning activities when images are used in the classroom.

***Open Academic Early Alert System: Technical Demonstration***

Sandeep Markondiah Jayaprakash

The Open Academic Analytics Initiative (OAAI) is a collaborative, multi-year grant program aimed at creating an open source academic alert system to improve student performances and to research into scaling up of learning analytics technologies and solutions across higher education. This paper synthesizes some of the technical decisions, design strategies & concepts explaining the prototype demonstration of the open-source academic early alert system. The paper organizes these details in several dimensions of data processing & analysis: data integration, predictive modelling and scoring with reporting. The paper should be relevant to practitioners and academicians who want to better understand the implementation of an academic early-alert system.

***HanZi Handwriting Acquisition with Automatic Feedback***

Chin Hwa Kuo, Jian Wen Peng, Wen Chen Chang

One of the most crucial distinctions between Chinese and western languages is that one is ideogram and the other is phonogram. Foreigners make character stroke sequence error and/or stroke direction error. However, that is unusual to native Chinese speaker. In this paper, we designed a HanZi writing environment with automatic feedback to wrestle with the above obstacle. Before massive handwriting collection, we also performed a pilot study to collect Chinese handwritings from the college students in USA. The corresponding experiment and experience that we gained enable us to refine the learning environment. As a result, we have the potential to design suitable HanZi learning and teaching strategies for learners and teachers.

***Analyzing Student Notes and Questions to Create Personalized Study Guides***

Perry Samson

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.

WEDNESDAY, MARCH 26

***Visual Analytics of Academic Writing***

Duygu Simsek, Simon Buckingham Shum, Anna De Liddo, Rebecca Ferguson, Agnes Sandor

This paper describes a novel analytics dashboard which visualises the key features of scholarly documents. The Dashboard aggregates the salient sentences of scholarly papers, their rhetorical types and the key concepts mentioned within these sentences. These features are extracted from papers through a Natural Language Processing (NLP) technology, called Xerox Incremental Parser (XIP). The XIP Dashboard is a set of visual analytics modules based on the XIP output. In this paper, we briefly introduce the XIP technology and demonstrate an example visualisation of the XIP Dashboard.

***Peer Evaluation of Student Generated Content***

Jared Tritz, Ginger Shultz, Nicole Michelotti, Tim McKay, Barsaa Mohapatra

We present a study that examines tournament style peer evaluation of student-generated explanations for a missed exam problem in introductory physics. We will include two studies from separate courses and semesters in our poster; both were performed using a tournament module as part of the interactive online coaching system called E2Coach at the University of Michigan. In the first study students created YouTube videos and in the second study students created written Google documents.

***Patterns of Persistence: What Engages Students in a Remedial English Writing MOOC?***

John Whitmer, Eva Schiorring, Pat James

This poster will present the initial findings of research on a MOOC in Pre-College English writing that provides a context for how MOOCs can help to address these critical issues. Specifically, the poster will present research results on the "Crafting an Effective Writer: Tools of the Trade" course offered through the Coursera platform in Summer 2013. Sources that inform the research include MOOC logfile data documenting student participation with course resources and student responses to two surveys conducted at the beginning and end of the course.

***Educational Technology Approach toward Learning Analytics: Relationship between Student Online Behavior and Learning Performance in Higher Education***

Taeho Yu, Il-Hyun Jo

The aim of this study is to suggest more meaningful components for learning analytics in order to help learners improving their learning achievement continuously through an educational technology approach. Multiple linear regression analysis is conducted to determine which factors influence student's academic achievement. 84 undergraduate students in a women's university in South Korea participated in this study. The six-predictor model was able to account for 33.5% of the variance in final grade, $F(6, 77) = 6.457$, $p < .001$, $R2 = .335$. Total logon time in LMS and interactions among peer students had a significant ($p < .05$) correlation with final grade. In addition, (ir)regularity of learning interval in LMS and number of downloads had significant ($p <. 10$) partial effects in the full model. However, total login frequency in LMS (ß=-.131, t=-.980, $p > .05$) and interaction with instructor (ß=-.035, t=-.344, $p > .05$) did not predict final grade. Total studying time in LMS, interaction with peers, (ir)regularity of learning interval in LMS, and number of downloads were determined to be significant factors for students' academic achievement in online learning environment. These four controllable variables not only predict learning outcomes significantly but also are changeable if learners put more effort to improve their academic performance. The results provide a rationale for the treatment for student time management effort.

**3:15 PM – 4:30 PM | Doctoral Consortium Posters | Grand 5**

***Detecting Usage-based Similarities of Data Objects to Enhance Recommender Systems***

Katja Niemann

This paper presents a PhD project that tries to discover similarities between data objects by analyzing their usage in online portals in order to enhance recommender systems. This approach is inspired by linguistic methods that are applied to discover semantic relations between words by analyzing their usage in speech and texts. By transferring these approaches to the usage of data objects, information that are implicitly given in the users' behaviour can be made explicit and usable, e.g. to compensate the scarcity of data sets recommender systems often need to deal with.

***Learning Analytics for Academic Writing through Automatic Identification of Metadiscourse***

Duygu Simsek

This paper describes PhD research into how the academic writing process can be supported in higher education through learning analytics. This research investigates: 1) whether computational techniques can automatically identify the attributes of good academic writing, as correlated with grades and as identified in the literature; 2) if this proves possible, how best to feedback actionable analytics to support students and educators; and 3) whether this feedback has any demonstrable benefits. To date, this work: 1) clarifies the potential contribution of language technology able to parse metadiscourse signaling academic rhetorical moves, and 2) shows how the output of such a parser might be rendered as visual analytics. The next phase will evaluate the accuracy of such technology, the degree to which its output correlates with grades, and user reactions. Do educators recognize value in such analytics, and do learners value it as formative feedback on their writing?



***From Texts to Actions: Reflection Support for Teachers and Students by Extracting Key Actions from Usage Data***

Maren Scheffel

The goal of this PhD project is to provide teacher as well as learner support by applying learning analytics to usage data with methodologies originally stemming from linguistics and used for text analysis. For example, key actions are extracted to identify action sequences and patterns of individual users or groups. Key actions are those actions that are significant for an underlying set of actions and that give an overview of the essential activities. First tests have shown that key action extraction is indeed a simple, but useful method of data distillation and reduction. Further tests, however, will have to be done to find out whether such analyses can be done "on the fly" to produce meaningful results and how key actions can be provided to teachers and learners in a meaningful way in order to support them during their courses.

***Data and Design in Online Education***

Michael Marcinkowski

The rise of big data and online education present a tremendous opportunity for learning analytics and education in general. With massive open online courses (MOOCs) offering educational opportunities for students around the globe, an enormous amount of data is being generated. The possibilities that this data offers education also present challenges for educators and designers who are now forced to confront new paradigms for how to conduct and understand the aims of education in which the terms and measures of success are not so well defined. Focusing on in-depth interviews with instructors and designers involved with the design of MOOCs and a qualitative examination of course data, this research examines how the interpretation of student data affects the design of online courses. This work looks at how analytic data can play a role in shaping our understanding of new approaches to education and provide guidance in future design work.

***Learning Analytics for Group Knowledge Building***

Bodong Chen

My program of research in learning analytics aims to tackle the significant challenge of assessing group knowledge advancement in knowledge building. It will focus on using semantic analysis to assess and model knowledge progress in knowledge building discourse from different aspects. Insights gained from this analysis will be fed to design of new analytic tools in Knowledge Forum to facilitate students' metadiscourse. This work will contribute to the learning analytics community for its unique approaches of assessing group learning processes and to the knowledge building community as well for its practical implications in classrooms.

***E3: Emotions, Engagement in Educational Games***

Ani Aghababyan

This study investigates ways to collect student engagement and gameplay data from an educational game called Quantum Spectrum in order to answer questions regarding student engagement in these environments and the effect of certain affective states on their performance. In addition, this study attempts to compare self-report and observational methods of data collection in order to provide insight on their influence on student concentration. Study participants are elementary students who play the game over the course of multiple class sessions. Preliminary findings suggest that there is an interesting inter-relation between frustration and confusion that is influence by the amount of external support provided. These findings will help guide my dissertation study that will concentrate on significant patterns of frustration and confusion emotional states along with their automatic detection.

***Interpretable Student Performance Prediction Model to Enable Personalized Intervention in a CSCL Environment: An Exploratory Study of Activity Theory based Genetic Programming Method***

Wanli Xing

Building a practical student performance prediction model is a difficult task not only because the data or factors involved are always overwhelming but also because there is a lack of semantic background for teachers to interpret the model developed to provide personalized help for struggling students. In this study, an activity theory informed genetic programming (GP) method is proposed to solve those problems using real data from a CSCL learning environment. First, in order to reduce data dimensionality and allow for teachers to provide personalized help, this work focuses on students' participation measures based on a theory proposed by Hrastinski (2009): online learning as online participation. Second, activity theory is used to systematically quantify students' participation and interaction with the environment. As a result, 6 variables are created based on activity theory. This variable development method not only holistically frames students' participation in a social-technical construct but also provides a semantic base for the prediction model. Third, this work uses a genetic programming method to develop a comprehensible model to help teachers comprehend and provide personalized interventions. The preliminary results show that the proposed methodology builds a reliable model that can identify academically at-risk students who might fail the course and this model can be easily implemented in a real environment. Further, teachers can depend on the model to offer personalized suggestions to struggling students.

***Identifying at-risk students in a First-Year Engineering Course***
Farshid Marbouti
The first step in helping students who may fail a course is identifying them early in the semester. One way to achieve this goal is to use predictive modeling techniques. One common problem with existing models is that they typically maximize the overall accuracy of the model at the cost of decreasing the accuracy of predicting at-risk students. Furthermore, they typically employ a general model that cannot address the complexity of all courses, especially those with active learning strategies. In my dissertation research, I will build models for important weeks of the semester (e.g. withdrawal deadline) to identify at-risk students in a course using academic factors, create an accuracy diagram based on the models and information available each week, and finally identify the optimal time to start using the models. I also compare three different methods (regression model, neural networks, and support vector machines) to find the most accurate predictive method.

***A Social Media Analytics Framework for Understanding Students' Learning Experiences***
Xin Chen
College students are among the most active users on social media sites, such as Twitter, Google+, and Facebook. Students' informal conversations on social media could shed light into their educational experiences—opinions, feelings, and concerns about the learning process. Data from such uninstrumented environments can provide valuable knowledge to inform student learning and improvement of education quality. Researchers in the learning analytics community usually focus on data produced from various course management systems or virtual learning environments. However, the open data on the web created by students have rarely been studied. This dissertation aims at understanding college students' educational experiences by analyzing student-generated data on Twitter. Among the vast amount of information that can be extracted from student generated data, issues related to students' community engagement, psychological well-being, and cultural diversity are the topics of special interests.

**4:30 PM – 5:30 PM | Reception | Grand 4**
*Sponsored By: Desire2Learn*

## THURSDAY, MARCH 27

**8:00 AM – 9:15 AM | Keynote | Grand 5**
*Sponsored By: Blue Canary, Session Chair: Stephanie Teasley*

***Is Learning Analytics a Disruptive Innovation?***
Nancy Law
Disruptive technologies have the potential to fundamentally change the ecological terrain of a market, not because they have outstanding performance or impact at its debut, but because they have features that enable them to address new needs and bring about totally different ways of (imagining) operations, new relationships and new services (Christensen, 1997). Learning analytics (LA) has all the features of a disruptive innovation. It is potentially empowering learners to have a much better understanding of their own progress and characteristics as a learner in relation to their own aspirations and their peers (in many different senses of the word), and on that basis to receive recommendations on further steps to take on their learning pathways. LA systems may provide for teachers real-time information about learner progress as well as feedback on effectiveness of different pedagogical approaches and learning designs. It may serve as an alternative methodology for developing theories about human learning. A lot has also been written about how institutions and whole school districts may make use of LA to track progress and make use of such information in multiple ways that could have significant implications at many levels. The LA research community is well aware of the many technological challenges yet to be addressed before these possibilities can be realized. However, given the very impressive developments in the short history of LA as a research field, and in particular the intentional efforts of those in the related disciplines to forge interdisciplinary partnership and to engage in multivocal dialogues, the technological futures for LA is a rosy one. On the other hand, if we review the last three decades of development in technology enhanced learning (TEL), a closely related field of research and innovation, it is reasonable to conclude that the promises of currently available TEL solutions are far from having gained a significant foothold in mainstream educational settings, despite the wide acceptance and support at the policy level. It is only in recent years that the disruptive nature of TEL as a socio-technical innovation and its implications for TEL implementation strategies is recognized. This paper discusses the challenges to and possible strategies for scalable development and adoption of LA, drawing on the literature as well as the author's own research in international comparative studies on sustaining ICT enabled pedagogical innovations and design-based research with K-12 teachers on integrating LA with e-learning design.

**9:15 AM – 9:30 AM | Morning Break | Grand Foyer**

**9:30 AM – 10:45 AM | Discussion with Nancy Law | Grand 2**
*Session Chair: Simon Buckingham Shum*



**9:30 AM – 10:45 AM | MOOCs | Grand 3**
*Session Chair: Negin Mirriahi*

***Visualizing patterns of student engagement and performance in MOOCs (Full)***
Carleton Coffrin, Linda Corrin, Paula de Barba, Gregor Kennedy
In the last five years, the world has seen a remarkable level of interest in Massive Open Online Courses or MOOCs. A consistent message from universities participating in MOOC delivery is their eagerness to understand students' online learning processes. This paper reports on an exploratory investigation of students' learning processes in two MOOCs which have dierent curriculum and assessment designs. The paper illustrates that when viewed through the lens of common MOOC learning analytics, these two courses appear very similar, to each other and MOOCs in general with students showing a high degree of initial interest, but the courses ultimately experience high attrition. With the goal of developing a greater understanding of students' patterns of learning behaviour in these courses, we investigate alternative learning analytic approaches and visual representations of the output of these analyses. Using these approaches we were able to meaningfully classify student types and visualize patterns of student engagement which were previously unclear. The ndings from this research contribute to the educational community's understanding of students' engagement and performance in MOOCs, and also provide the broader learning analytics community with suggestions of new ways to approach learning analytic data analysis and visualization.

***Small to Big Before Massive: Scaling up Participatory Learning Analytics (Short)***
Daniel Hickey, Tara Kelly, Xinyi Shen
This paper describes the scaling up of online course features and learning analytics in support of "participatory" forms of personalized learning. These features initially emerged within refinements of two small online university courses. This relatively intimate context allowed for informal and intuitive learning analytics that were well suited to the challenges of fostering participatory learning in formal educational contexts. One of the courses and the learning features and analytics were then scaled up within a "big open online course" offered to hundreds of students. This paper describes how thirteen course features were streamlined to function more systematically in this "BOOC." The features and analytics are now being scaled up further for massive courses that required automated algorithmic analytics. The paper suggests that the scaling of "personalized learning analytics" should (a) be done incrementally, (b) employ design-based research methods, (c) focus on productive forms of disciplinary engagement, and (d) draw from emerging situative and participatory approaches to assessment.

***Success, activity and drop-outs in MOOCs. An explorative study on the UNED COMA courses (Short)***
Jose Luis Santos, Joris Klerkx, Erik Duval, David Gago, Luis Rodriguez
This paper presents an exploratory study about the UNED COMA courses. In order to apply proper learning analytics strategies, we need beforehand the existing relations between user success and the platform itself beforehand. Supported by previous literature of more conventional courses, we propose five research questions that are explored throughout the paper in order to understand how success, dropouts and the platform OpenMooc relate to each other. Throughout the paper, we discuss whether the structure of the course, the activity within the platform, and the use of the forum in language courses relate to success. We conclude that activity within the forum as well as the access to the activities have a correlation with the success in the course. However it is not clear how forum helps the students.

**10:45 AM – 12:00 PM | Panel | Grand 2**
*Session Chair: George Siemens*

***Learning Analytics for the Social Media Age***
Anatoliy Gruzd, Caroline Haythornthwaite, Christopher Andrew Paulin, Rafa Absar, Michael Huggett

**10:45 AM – 12:00 PM | Learning Analytics for "at risk" Students | Grand 3**
*Social Chair: Ryan Baker*

***Engagement vs Performance: Using Electronic Portfolios to Predict First Semester Engineering Student Retention (Full)***
**Best Paper Award Candidate**
Everaldo Aguiar, Nitesh Chawla, Jay Brockman, George Alex Ambrose, Victoria Goodrich
As providers of higher education begin to harness the power of big data analytics, one very fitting application for these new techniques is that of predicting student attrition. The ability to pinpoint students who might soon decide to drop out of a given academic program allows those in charge to not only understand the causes for this undesired outcome, but it also provides room for the development of early intervention systems. While making such inferences based on academic performance data alone is certainly possible, we claim that in many cases there is no substantial correlation between how well a student performs and his or her decision to withdraw. This is specially true when the overall set of students has a relatively similar academic performance. To address this issue, we derive measurements of engagement from students' electronic portfolios and show how these features can be effectively used to augment the quality of predictions.

***Perceptions and Use of an Early Warning System During a Higher Education Transition Program (Short)***
Stephen Aguilar, Steven Lonn, Stephanie Teasley
This paper reports findings from the implementation of a learning analytics-powered Early Warning System (EWS) by academic advisors who were novice users of data-driven learning analytics tools. The information collected from these users sheds new light on how student analytic data might be incorporated into the work practices of advisors working with university students. Our results indicate that advisors predominantly used the EWS during their meetings with students, despite it being designed as a tool to provide information to prepare for meetings and identify students who are struggling academically. This introduction of an unintended audience brings significant design implications to bear that are relevant for learning analytics innovations.

***Modest analytics: using the index method to identify students at risk of failure (Short)***
Tim Rogers, Cassandra Colvin, Belinda Chiera
Regression is the tool of choice for developing predictive models of student risk of failure. However, the forecasting literature has demonstrated the predictive equivalence of much simpler methods. We directly compare one simple tabulation technique, the index method, to a linear multiple regression approach for identifying students at risk. The broader purpose is to explore the plausibility of a flexible method that is conducive to adoption and diffusion. In this respect this paper fits within the ambit of the modest computing agenda, and suggests the possibility of a modest analytics. We built both regression and index method models on 2011 student data and applied these to 2012 student data. The index method was comparable in terms of predictive accuracy of student risk. We suggest that the context specificity of learning environments makes the index method a promising tool for educators who want a situated risk algorithm that is flexible and adaptable.

**12:00 PM – 1:30 PM | Lunch and SoLAR General Meeting | Grand 5**

**1:30 PM – 3:00 PM | Text Analytics and Collaborative Environments | Grand 2**
*Session Chair: Alyssa Wise*

***Analytics of the Effects of Video Use and Instruction to Support Reflective Learning (Full)***
Dragan Gasevic, Negin Mirriahi, Shane Dawson
Although video annotation software is no longer considered as a new innovation, its application in promoting student self-regulated learning and reflection skills has only begun to emerge in the research literature. Advances in text and video analytics provide the capability of investigating students' use of the tool and the psychometrics and linguistic processes evident in their written annotations. This paper reports on a study exploring students' use of a video annotation tool when two different instructional approaches were deployed – graded and non-graded self-reflection annotations within two courses in the performing arts. In addition to counts and temporal locations of self-reflections, the Linguistic Inquiry and Word Counts (LIWC) framework was used for the extraction of variables indicative of the linguistic and psychological processes associated with self-reflection annotations of videos. The results indicate that students in the course with graded self-reflections adopted more linguistic and psychological related processes in comparison to the course with non-graded self-reflections. In general, the effect size of the graded reflections was lower for students who took both courses in parallel. Consistent with prior research, the study identified that students tend to make the majority of their self-reflection annotations early in the video time line. The paper also provides several suggestions for future research to better understand the application of video annotations in facilitating student learning.

***Peer Assessment Based on Ratings in a Social Media Course (Short)***
Andrii Vozniuk, Adrian Holzer, Denis Gillet
Peer assessment is seen as a powerful supporting tool to achieve scalability in the evaluation of complex assignments in large courses, possibly virtual ones, as in the context of massive open online courses (MOOCs). However, the adoption of peer assessment is slow due in part to the lack of ready-to-use systems. Furthermore, the validity of peer assessment is still under discussion. In this paper, in order to tackle some of these issues, we present as a proof-of-concept of a novel extension to Graasp, a social media platform, to setup a peer assessment activity. We then report a case study of peer assessment using Graasp in a Social Media course with 60 master students and analyze the level of agreement between students and instructors in the evaluation of short individual reports. Finally, to see if both instructor and student evaluations were based on appearance of project reports rather than on content, we conducted a study with 40 kids who rated reports solely on their look. Our results convey the fact that unlike the kid evaluation, which shows a low level of agreement with instructors, student assessment is reliable since the level of agreement between instructors and students was high.



***Collaborative Spatial Classification (Short)***
Eric Coopey, R. Benjamin Shapiro, Ethan Danahy
Interactive technologies have become an important part of teaching and learning. However, the data and information that these systems generate is increasingly unstructured, complex, and therefore difficult to make sense of. Current computationally driven methods (e.g., latent semantic analysis or learning based image classifiers) for classifying student contributions don't include the ability to function on multi-modal artifacts (e.g., sketches, videos, or annotated images) that new technologies enable. We have developed and implemented an approach based on learners' interactions with the artifacts they create. This new form of semi-automated concept classification, Collaborative Spatial Classification, leverages the spatial arrangement of artifacts to provide a visualization that generates summary level data about about idea distribution. This approach has two benefits. First, students learn to identify and articulate patterns and connections among classmates ideas. Second, the teacher receives a high-level view of the distribution of ideas, enabling them to decide how to shift their instructional practices in real-time.

***Assessing Elementary Students' Science Competency with Text Analytics (Short)***
Samuel Leeman-Munk, James Lester, Eric Wiebe
Real-time formative assessment of student learning has become the subject of increasing attention. Students' textual responses to short answer questions offer a rich source of data for formative assessment. However, automatically analyzing textual constructed responses poses significant computational challenges, and the difficulty of generating accurate assessments is exacerbated by the disfluencies that occur prominently in elementary students' writing. With robust text analytics, there is the potential to accurately analyze students' text responses and predict students' future success. In this paper, we present WRITEEVAL, a hybrid text analytics method for analyzing student-composed text written in response to constructed response questions. Based on a model integrating a text similarity technique with a semantic analysis technique, WRITEEVAL performs well on responses written by fourth graders in response to short-text science questions. Further, it was found that WRITEEVAL's assessments correlate with summative analyses of student performance.

**1:30 PM – 3:00 PM | Institutional Perspectives | Grand 3**
*Session Chair: Grace Lynch*

***Techniques for Data-Driven Curriculum Analysis (Full)***
**Best Paper Award Candidate**
Gonzalo Mendez, Xavier Ochoa, Katherine Chiluiza
One of the key promises of Learning Analytics research is to create tools that could help educational institutions to gain a better insight of the inner workings of their programs, in order to tune or correct them. This work presents a set of simple techniques that applied to readily available historical academic data could provide such insights. The techniques described are real course difficulty estimation, dependance estimation, curriculum coherence, drop-out paths and load/performance graph. The description of these techniques is accompanied by its application to real academic data from a Computer Science program. The results of the analysis are used to obtain recommendations for curriculum re-design.

***The Impact of Learning Analytics on the Dutch Education System (Short)***
Hendrik Drachsler, Slavi Stoyanov, Marcus Specht
This article reports about a Group Concept Mapping study that was conducted within the framework of the Learning Analytics Summer Institute (LASI) in Amsterdam from July until August 2013 by the SURF SIG on Learning Analytics. The conducted Group Concept Mapping study shows some interesting results towards the impact of Learning Analytics on the Dutch Education system. It provides evidence for the convergence of Dutch Learning Analytics community and contributes towards a national research agenda for Learning Analytics in the Netherlands.

***An Exercise in Institutional Reflection: The Learning Analytics Readiness Instrument (LARI) (Short)***
Kimberly E. Arnold, Steven Lonn, Matthew D. Pistilli
While the landscape of learning analytics is relatively well defined, the extent to which institutions are ready to embark on an analytics implementation is less known. Further, while work has been done on measuring the maturity of an institution's implementation, this work fails to investigate how an institution that hasn't implemented analytics to date might become mature over time. To that end, the authors developed and piloted a survey, the Learning Analytics Readiness Instrument (LARI), in an attempt to help institutions successfully prepare themselves for a successfully analytics implementation. The LARI is comprised of 90 items encompassing five factors related to a learning analytics implementation: (1) Ability, (2) Data, (3) Culture and Process, (4) Governance and Infrastructure, and, (5) Overall Readiness Perception. Each of the five factors has a high internal consistency, as does the overall tool. This paper discusses the need for a survey such as the LARI, the tool's psychometric properties, the authors' broad interpretations of the findings, and next steps for the LARI and the research in this field.

***Competency Map: Visualizing Student Learning to Promote Student Success (Short)***
Jeff Grann, Deborah Bushway
Adult students often struggle to appreciate the relevance of their higher educational experience to their careers. Capella University's competency map is a dashboard that visually indicates each student's status relative to specific assessed competencies. Students who utilize the competency map are more satisfied with the design of their courses and demonstrate competencies at higher levels. These students also persist in their program at greater rates, even after statistically controlling for powerful covariates, such as engagement. Implementation suggestions for engaging faculty are shared and implications discussed.

**3:00 PM – 3:15 PM | Afternoon Break | Grand Foyer**

**3:15 PM – 4:30 PM | Analysis of Resource Use in LMS | Grand 2**
*Session Chair: Katherine Maillet*

***Analysis of Dynamic Resource Access Patterns in a Blended Learning Course (Full)***
Tobias Hecking, Sabrina Ziebarth, H. Ulrich Hoppe
This paper presents an analysis of resource access patterns in a recently conducted master level university course. The specialty of the course was that it followed a new teaching approach by providing additional learning resources such as wikis, self-tests and videos. To gain deeper insights into the usage of the provided learning material we have built dynamic bipartite student – resource networks based on event logs of resource access. These networks are analysed using methods adapted from social network analysis. In particular we uncover bipartite clusters of students and resources in those networks and propose a method to identify patterns and traces of their evolution over time.

***Analyzing the Log Patterns of Adult Learners in LMS Using Learning Analytics (Short)***
Il-Hyun Jo, Dongho Kim, Meehyun Yoon
In this paper, we describe a process of constructing a proxy variable from recorded log data within Learning Management System (LMS) that represent adult learners' time management strategies in an online course. Based upon previous research, three values were selected from data set along with a guideline for manipulating log data. According to the result of multiple regression analysis conducted to determine whether the suggested values predict learning performance, an (ir)regularity of the learning interval was proven to be correlative with and predict learning performance. As indicated in previous research, regularity of learning is a powerful indicator to explain learners' consistent endeavors. This study demonstrates the possibility of using learning analytics to address a learner's specific competence on the basis of a theoretical background. Implications for the learning analytics field seeking a pedagogical theory-driven approach are discussed.

***Practice Exams Make Perfect: Incorporating Course Resource Use into an Early Warning System (Short)***
Richard Waddington, Sungjin Nam
Early Warning Systems (EWSs) are being developed and used more frequently across college campuses to aggregate multiple sources of data and provide timely information to various stakeholders about students in need of academic support. As these systems grow more complex in their ability to provide knowledge about a student's developing grade trajectory, there is an increasing need to incorporate relevant and real-time course-related information that could be precursors to a student's success or failure. This paper presents the initial investigation of how to incorporate additional information from students' use of course resources from a Learning Management System (LMS) into an existing EWS. Specifically, we focus our efforts understanding the relationship between course resource use and a student's grade. Using ten semesters of LMS data from a required Chemistry course, we categorized course resources into four distinct categories and used a multinomial logistic regression model with semester fixed-effects to estimate the impact of course resource use on the likelihood that a student receives an "A" or "B" in the course as compared to a "C or Lower." Results suggest that students who use resources related to Exam Preparation or Lectures to a greater degree than their peers are more likely to receive an "A" or "B" in the course as compared to a "C" or Lower". We discuss the implications of our results for the further development of the EWS and EWSs in general.



**3:15 PM – 4:30 PM | Learning Analytics and Learning Design | Grand 3**
*Session Chair: Hendrik Drachsler*

***Educational Data Sciences – Framing Emergent Practices for Analytics of Learning, Organizations, and Systems (Full)***
Philip Piety, Daniel Hickey, Mj Bishop
In this paper, we develop a conceptual framework for organizing emerging analytic activities involving educational data that can fall under broad and often loosely-defined categories, including Academic/Institutional Analytics, Learning Analytics/ Educational Data Mining, Learner Analytics/Personalization, and Systemic Instructional Improvement. While our framework is substantially informed by both higher education and K-12 settings, this framework is developed to apply across all educational contexts where digital data are used to inform the management of learning processes. Indeed, while we can identify movements that are relatively independent of each other today, we believe they will in all cases expand from their current margins to encompass larger domains and increasingly overlap. The growth in these analytic activities leads to a need for ways to synthesize across these different areas, to find common language and frames of reference as these movements develop into a field.

***Designing Pedagogical Interventions to Support Student Use of Learning Analytics (Full)***
Alyssa Wise
This article addresses a relatively unexplored area in the emerging field of learning analytics, the design of learning analytics interventions. A learning analytics intervention is defined as the surrounding frame of activity through which analytic tools, data, and reports are taken up and used. It is a soft technology that involves the orchestration of the human process of engaging with the analytics as part of the larger teaching and learning activity. This paper first makes the case for the overall importance of intervention design, situating it within the larger landscape of the learning analytics field, and then considers the specific issues of intervention design for student use of learning analytics. Four principles of pedagogical learning analytics intervention design that can be used by teachers and course developers to support the productive use of learning analytics by students are introduced: Integration, Agency, Reference Frame and Dialogue. These principles are united in a preliminary model of pedagogical learning analytics intervention design for students, presented as a starting point for further inquiry.

***A Cognitive Processing Framework for Learning Analytics (Short)***
Andrew Gibson, Kirsty Kitto, Jill Willis
Incorporating a learner's level of cognitive processing into Learning Analytics presents opportunities for obtaining rich data on the learning process. However, obtaining this data without significant changes to existing curriculum can be a complex undertaking. We propose a framework called COPA that provides a basis for mapping levels of cognitive operation into a learning analytics system. We utilise Bloom's taxonomy, a theoretically respected conceptualisation of cognitive processing, and apply it in a flexible structure that can be implemented incrementally and with varying degree of complexity within an educational organisation. We identify some of the potential benefits that the framework can bring to Learning Analytics, and identify some possible approaches to its use. We also apply the framework to a new University unit of study, offered for the first time in 2013, and its power in identifying key missing elements in the structure of the course is demonstrated. This in turn provides suggestions as to how the unit should be restructured in its next offering in 2014, and current work in this direction will be briefly covered.

**6:00 PM – 9:00 PM | Conference Dinner | Eitlejorg Museum of American Indians and Western Art**
Pre-registration required

## FRIDAY, MARCH 28

**8:00 AM – 9:15 AM | Keynote | Grand 5**
*Sponored By: Blue Canary, Session Chair: Abelardo Pardo*

**Design at Large**
Scott Klemmer
This is the most exciting time of my career so far. Learning, design, and technology are on the front page of the news and the topic of Hollywood films. Enrollments in our classes have skyrocketed, both in person and online. For me, the most powerful dynamic of the current moment is the large number of people who are excited about making stuff.

This current wave of enthusiasm offers a tremendous opportunity for durable impact, but also a challenge. I've spent a lot of time with design teachers over the past two decades. While the design fields have had enormous impact, I think nearly everyone who has taught design has wished for more and better theory (vocabulary, principles, predictive power) to help coach students towards expertise. Why the gap? First, many see design as a mystical endeavor because creative work is clearly complex and multifarious. Consequently, many people give up on creating theory — or seeking to apply/adapt existing theory — without trying on the belief that it's prima facie impossible. Second, much of the theory I'm talking about here is really social science. At many universities, the social sciences are across campus from the arts and engineering programs that house design, and even further away intellectually.

In this talk, I'll share our adventures in facilitating peer interactions in online education. In person, communities of peers, shoulder-to-shoulder, can create powerful learning experiences. For example, hallmark traits of design education are studio critique and working in a collocated physical space. With online learning, are students "alone together" — viewing the same materials with no peer contact?

In 2012, my research group collaborated with Coursera to launch the first massive-scale class with self and peer assessment. Since then, more than 80 other massive online classes have featured peer assessment — from world music to nutrition to mathematical thinking. On the whole, it has worked surprisingly well. I'll also share challenges and failures, and talk about integrating peer and automated assessment.

This led us to experiments with small group video discussion, social networks, and global meet-ups — also deploying our systems in collaboration with teachers from diverse fields. My goal is to offer techniques that attendees can adapt for their own teaching and avoid our missteps. And I look forward to hearing more about the adventures of others.

More broadly, because online learning platforms embed pedagogy into software, they provide a powerful setting for using and building theory through experimentation. These online learning interventions illustrate how design at large — at scale, embedded in real-world activities, and occasionally subversive — can be a potent and relevant research strategy.

**9:15 AM – 9:30 AM | Morning Break | Grand Foyer**

**9:30 AM – 10:45 AM | Discussion with Scott Klemmer | Grand 2**
*Session Chair: Erik Duval*

**9:30 AM – 10:45 AM | Discourse and Argumentation | Grand 3**
*Session Chair: Ming Ming Chiu*

***Statistical Discourse Analysis of Online Discussions: Informal Cognition, Social Metacognition and Knowledge Creation (Full)***
Ming Chiu, Nobuko Fujita
To statistically model large data sets of knowledge processes during asynchronous, online forums, we must address analytic difficulties involving the whole data set (missing data, nested data and the tree structure of online messages), dependent variables (multiple, infrequent, discrete outcomes and similar adjacent messages), and explanatory variables (sequences, indirect effects, false positives, and robustness). Statistical discourse analysis (SDA) addresses all of these issues, as shown in an analysis of 1,330 asynchronous messages written by 17 students during a 13-week online educational technology course. The results showed how attributes at multiple levels (individual and message) affected knowledge creation processes. Men were more likely than women to theorize. Asynchronous messages created a micro-time context; opinions and asking about purpose preceded new information; and anecdotes, opinions, different opinions, elaborating ideas, and asking about purpose or information preceded theorizing. These results show how informal thinking precedes formal thinking and how social metacognition affects knowledge creation.



***Uncovering What Matters: Analyzing Sequential Relations Among Contribution Types in Knowledge-building Discourse (Short)***
Bodong Chen, Monica Resendes
Temporality matters for analysis of collaborative learning, which is a dynamic, interactive process that evolves over time. This study highlights the significance of analyzing the temporal dimension of knowledge-building discourse. In particular, the authors contrasted sequential relations among different contribution types in "productive" inquiry threads with those in "improvable" threads, with a goal to uncover distinguishing temporal patterns. In this analysis, productivity of threads was assessed through content analysis of knowledge advancement and contribution types were coded using a "ways of contributing to explanation-seeking discourse" scheme. Sequential relations among different ways of contributing were uncovered by applying Lag-sequential Analysis. Results indicated that productive discourse involved significantly more transitions among questioning, theorizing, obtaining information, and working with information; in contrast, responding to questioning and theorizing by merely giving opinions is not sufficient to achieve knowledge progress. This study highlights the importance of investigating temporality in collaborative learning and calls for attention to developing and testing temporal analysis methods in learning analytics research.

**10:45 AM – 12:00 PM | Panel | Grand 2**
*Session Chair: John Whittmer*

The last five years has brought an explosion of research in the learning analytics field. However, much of what has emerged has been small scale or tool centric. While these efforts are vitally important to the development of the field, in order to truly transform educational, learning analytics must scale and become institutionalized at multiple levels throughout an educational system. The University of Wisconsin System (UWS) is currently undertaking the challenge; however, there is little literature guiding institutions on how to scale-up a comprehensive learning analytics program. As a system with 26 campuses and an extension, UWS serves over 180,000 students annually. In order to be successful in a system-wide learning analytics initiative, the needs of each student, faculty member or instructor, administrator, and campus must be systematically addressed. This panel presentation will use the following question to guide discussion: What organizational capacities are necessary for successful systematic adoption of learning analytics? Pugliese's five stages of student success analytics will be used to frame the discussion: 1) technology infrastructure, analytics tools, and applications; 2) policies, processes, practices, and workflows; 3) values and skills; 4) culture and behavior; and 5) leadership.

***Building Institutional Capacities and Competencies for Systemic Learning Analytics Initiatives***
Kimberly Arnold, Grace Lynch, Daniel Huston, Christopher Olsen, Linda Jorn, Lorna Wong, Clare Huhn, Daniel Voeks, Andrew Taylor

**10:45 AM – 12:00 PM | Who We Are and Who We Want to Be | Grand 3**
*Session Chair: James Willis, III*

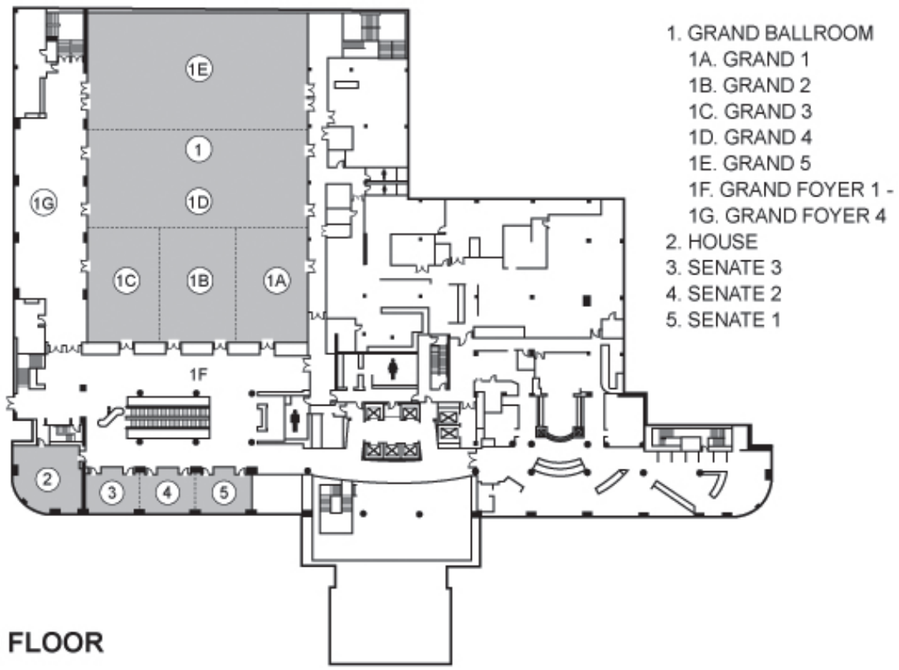
***Current State and Future Trends: a Citation Network Analysis of the Learning Analytics Field (Full)***
**Best Paper Award Candidate**
Shane Dawson, Dragan Gasevic, George Siemens, Srecko Joksimovic
This paper provides an evaluation of the current state of the field of learning analytics through analysis of articles and citations occurring in the LAK conferences and identified special issue journals. The emerging field of learning analytics is at the intersection of numerous academic disciplines, and therefore draws on a diversity of methodologies, theories and underpinning scientific assumptions. Through citation analysis and structured mapping we aimed to identify the emergence of trends and disciplinary hierarchies that are influencing the development of the field to date. The results suggest that there is some fragmentation in the major disciplines (computer science and education) regarding conference and journal representation. The analyses also indicate that the commonly cited papers are of a more conceptual nature than empirical research reflecting the need for authors to define the learning analytics space. An evaluation of the current state of learning analytics provides numerous benefits for the development of the field, such as a guide for under-represented areas of research and to identify the disciplines that may require more strategic and targeted support and funding opportunities.

***Teaching the Unteachable: On The Compatibility of Learning Analytics and Humane Education (Short)***
Timothy D. Harfield
This paper is an exploratory effort to find a place for learning analytics in the humanities. After distinguishing humane education from training on the basis of the Aristotelian model of intellectual capabilities, and arguing that what marks humane as distinct is its interest in cultivating prudence, which is unteachable, an account of three key characteristics of humane education is provided. Appealing to thinkers of the Italian Renaissance, it is argued that ingenium, eloquence, and self-knowledge constitute the what, how, and why of humane education. Lastly, with special attention to a piece by Wise, Zhao, and Hausknecht, it is demonstrated that learning analytics are not only helpful as aids to ensuring success in scientific and technical disciplines, but in the humanities as well. In order to function effectively as an aid to humane education, however, learning analytics must be embedded within a context that encourages continuous reflection, responsiveness, and personal responsibility for learning.

**Establishing an Ethical Literacy for Learning Analytics (Short)**
Jenni Swenson
This paper cites seminal work in the emerging discipline of academic analytics and situates the discipline within social analysis research through word clouds and content analysis of LAK Conference abstracts. A mapping metaphor is used to define the current strengths of academic analytics (pedagogy and research), and identify future direction for academic analytics (theory and ethics as "social change"). Numerous frameworks are presented to provide direction for the ethical production and presentation of academic analytics "artifacts," as well as to begin establishing an ethical literacy for the discipline.

**12:00 PM – 1:00 PM | Closing Session | Grand 5**

## The Westin Indianapolis Hotel Map




SECOND FLOOR



# LAK 2014 Conference Organization

**ORGANIZING INSTITUTIONS**

SOLAR SOCIETY for LEARNING ANALYTICS RESEARCH

PURDUE UNIVERSITY

John N. Gardner Institute FOR EXCELLENCE IN UNDERGRADUATE EDUCATION

WISCONSIN UNIVERSITY OF WISCONSIN–MADISON

**CONFERENCE CO-CHAIRS**

Matthew Pistilli
*Purdue University, IN, USA*

James Willis
*Purdue University, IN, USA*

Drew Koch
*John N. Gardner Institute for Excellence in Undergraduate Education, NC, USA*

Kimberly Arnold
*University of Wisconsin-Madison, WI, USA*

**PROGRAM CHAIRS**

Stephanie Teasley
*University of Michigan, MI, USA*

Abelardo Pardo
*The University of Sydney, Australia*

**WORKSHOP AND TUTORIAL CHAIRS**

Kalina Yacef
*The University of Sydney, Australia*

Hendrik Drachsler
*Center for Learning Sciences and Technologies, Open University Netherlands*

**DEMONSTRATIONS AND POSTER CHAIRS**

Zach Pardos
*University of California at Berkeley, USA*

Anna de Liddo
*The Open University, UK*

**VOLUNTEER CHAIR**

Vic Borden
*Indiana University, IN, USA*

**DOCTORAL CONSORTIUM CHAIRS**

Chris Quintana
*University of Michigan, MI, USA*

Alysa Wise
*Simon Fraser University, Canada*

Katherine Maillet
*Institut Mines- Télécom, Télécom Ecole de Management, France*

Stephanie Teasley
*University of Michigan, MI, USA*

**SOCIAL MEDIA CHAIR**

Sean Goggins
*Drexel University, PA, USA*

**WEBMASTER CHAIR**

Krista Galyen
*University of Missouri, MO, USA*

**PUBLICITY CHAIRS**

Jenni Swenson
*Lake Superior College, MN, USA*

Tricia Leggett
*Zane State College, OH, USA*

**LOCAL CHAIR**

James Willis
*Purdue University, IN, USA*

**PROGRAM DESIGN**

Jenny Meek
*Purdue University, IN, USA*