Multimodal Learning Analytics - Enabling the Future of Learning through Multimodal Data Analysis and Interfaces

Marcelo Worsley
Stanford University
520 Galvez Mall
CERAS 102
Stanford, CA 94305
mworsley@stanford.edu

ABSTRACT
Project-based learning has found its way into a range of formal and informal learning environments. However, systematically assessing these environments remains a significant challenge. Traditional assessments, which focus on learning outcomes, seem incongruent with the process-oriented goals of project-based learning. Multimodal interfaces and multimodal learning analytics hold significant promise for assessing learning in open-ended learning environments. With its rich integration of a multitude of data streams and naturalistic interfaces, this area of research may help usher in a new wave of education reform by supporting alternative modes of learning.

Categories and Subject Descriptors
K.3.1 [**Computers and Education**]: Computer Uses in Education – *collaborative learning, computer assisted instruction, computer managed instruction;* I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems – *natural language interface;*. I.2.7 [**Artificial Intelligence**]: Natural Language Processing – *discourse, language parsing and understanding, text analysis;* I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding – *motion, video analysis;*

Keywords
Constructionism; Probabilistic Modeling; Learning; Data Mining

1. INTRODUCTION
Since the work of Dewey [1][2], Vygotsky [3] and Papert [4] there has been an increased integration of student-designed project-based learning (SPBL) – constructionism [4], inquiry based learning and constructivism, for example - in classrooms and in informal contexts. However, systematically assessing learning in these environments remains a significant challenge and an area of significant importance [5]. Traditional assessments, which are easy to distribute and analyze, may necessarily be incongruent with the objectives of SPBL because most assessments tend to focus on outcomes, whereas SPBL is largely focused on the process. Furthermore, the forms of assessment that may be well suited to the objectives SPBL environments: ethnographies, micro-genetic analysis and portfolio based assessments; are unable to scale to the level that is needed for educators to consistently and reliably turn to SPBL. This creates a conflict that can often times be difficult to resolve, and leave individuals dissatisfied with the lack of measurable changes in learning outcomes in project based learning studies.

In order to develop a more appropriate measure and means for assessing SPBL I want to construct a complete picture of how learning takes place in SPBL environments through the use of multimodal learning interfaces and techniques. To this end, the overarching research question that I have is, 'How do students learn in SPBL environments?' with primary sub-question: 'Can the learning processes in SPBL environments be characterized using a system of continuous, multi-modal monitoring based on a composition of artificial intelligence technologies?' It is my hypothesis that leveraging multimodal data can better allow educators and education researchers to capture a holistic picture of their students. This complete picture will lend to developing a rich and intricate representation of each student's learning. Furthermore, as we are able to better understand and automatically process this data, we will be able usher in the future of learning interfaces.

ICMI '12, October 22 -26, Santa Monica, California, USA.

2. PREVIOUS LITERATURE
My work builds on literature from the learning sciences, cognitive science and learning analytics (or educational data mining). More specifically, previous works describe the importance of student-designed project based learning experiences for promoting learning of STEM (Fortus, Dershimer, Krajcik, Marx, & Mamlok-Naaman 2004; Roth, 1996; 1997; 1998 in [6]); techniques for observing behavioral and developmental changes among students through rich ethnographic studies [7][8][9][10][11]; and using automated techniques for identifying salient markers of learning in a range of modalities [12][13][14]. Furthermore, this research builds on the work that I have done during the past three years which has found that there are meaningful cues in student speech [15][16][17], gaze [18], programming state [19]; epistemological beliefs and identity [20]; and in a combination of modalities [21][22]. It also builds on a number of published and unpublished research tools that enable: object tracking [23]; user localization [24]; and multi-modal data capture of collaborative work [25][26][27]. To this end, my previous research has looked at a variety of modalities in isolation, but as I move forward into my dissertation work I want to begin to better leverage the integration of different data streams in extended analyses of student learning.

3. INITIAL STUDY DESIGN

In order to realize the above considerations and answer my research questions, I will be studying high school and undergraduate students as they participate in engineering design workshops. During these workshops students do bifocal modeling, computational modeling, digital fabrication, robotics and introductory electronics, computer programming, wood working, polymer casting and more. As students interact within this space, I will be using previously developed tools, in conjunction with in process multimodal interfaces, in order to synchronously capture data from a range of modalities. The specific modes of data capture that I am currently considering include:

Digital Design Drawing Data - Using digital pens and paper, I will capture continuous inking streams of student drawings. This will enable me to have the full evolution of their designs in a high precision fashion. These drawings will be aligned with speech data.

Student Gaze Data, Explanations and NetLogo logs – students will be asked to participate in a pair of studies in which they explore STEM phenomena in Netlogo, an agent based modeling environment. I will log their gaze, verbal explanations and Netlogo actions through synchronous data capture.

Student Wifi-based Localization – Using a custom Android application, I will capture student's relative locations at half-second increments as they move around the lab. This information will be useful for studying relative student collaboration, as well as for bootstrapping other data streams.

Student Motivation and Sentiment – This data will come from a mobile phone survey platform. Students will be asked multiple choice, likert scale and free response questions through periodic polling. This information is useful for grounding some of the interpretations that I make from looking at students speech and actions near the time of polling.

Student Dialogue Capture – An array of microphones arrays, lapel and head-mounted microphones will be used to capture student dialogue throughout the lab. I am still trying to find the best solution to do this synchronously, in a relatively large space, and for a large number of users.

Student Location Capture – Using Kinect Sensors and a Teachscape Panoramic Camera, I will capture student location and actions.

4. DATA ANALYSIS

As one can imagine, this dataset represents a large and difficult technological undertaking. As mentioned above, there are significant challenges in building a system that can synchronously capture data from a large number of participants (n=20), for hours at a time, and in a relatively large space (1000 sq ft.). At the moment I am considering extending existing tools. For example, one possibility is to extend the OpenGesture [24][25] platform: an easy to author platform that individuals can use to make speech and gesture-based applications from HTML and Javascript. Furthermore the platform can synchronously capture speech, gestures, audio and the number of people and faces in the vicinity of the user. I am also working on a couple of applications that make use of arrays of Kinect sensors. Beyond the technological tasks associated with synchronously capturing the data, aligning data from different modalities also presents a significant challenge. From previous observation a user's actions and words are often misaligned, with one noticeably lagging behind the other. This will be a challenge when trying to integrate elements from natural language understanding, gesture recognition and action recognition, recalling that the goal is to use these channels to make an accurate prediction about a student's learning.

However, despite the technological challenges, I believe that my previous work has properly positioned me to successfully complete this work. As mentioned in the previous literature section, I have done a number of projects that involve the use of the various modalities. Moreover, some of my published work includes: 1) demonstrated techniques for using linguistic, semantic and syntactic cue; 2) creation of a multimodal interface that use multimodal integration to address the challenges of an open microphone environment; and 3) involved doing rich data capture and analysis of prosodic, spectral and vision based features. In addition to this, I have a number of unpublished works that look at areas of gesture recognition, natural language understanding, domain adaptation for language modeling, object tracking and more.

5. TENTATIVE TIMELINE

This summer I will be refining and testing some of the research tools that I have developed, so that I can iterate on them again during the fall before launching into a focused data capture phase. I will then instrument the Learning Fabrication Lab at Stanford , where I will collect data from the high school and undergraduate students that use the lab. Data collection will likely occur over the course of six months, and will then move me into data preprocessing at the beginning of 2013. Once the data has been preprocessed I will move into a data annotation phase in which I begin to explore ways for automating data analysis. During data annotation I will also be working with human annotators in order to provide validity for the automated techniques that I use.

6. RESEARCH IMPACT

In this short document I have attempted to summarize my initial ideas for dissertation work, which is to study project-based learning techniques through multimodal technologies. The larger goal is to transform the nature and design of the technological tools used for promoting and supporting student learning. Furthermore, this work will help in advancing the use of multimodal interfaces and multimodal learning analytics in both formal and informal learning environments by elucidating techniques that can be used to make the bridge between multimodal interfaces and learning. In addition to this, the work is intended to uncover novel techniques and inferences that can generally improve multimodal interface development. For example, the research has the potential to assist in the areas of scalability, both in terms of number of participants and in terms of the physical space being analyzed. Additionally, I am also interested in advancing the use of open-microphone environments, through multimodal cues, and the community's

ability to do speech recognition, and speaker identification in learning environments, which tend to be noisy and unstructured. Lastly, this research should help the future design and implementation of multimodal learning interfaces. While it is nice to know more about how students learn, the ultimate objective is to find useful ways to improve each student's learning by making learning assessments more naturalistic, and create interfaces that enable learners to better understand their own learning.

7. ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Paulo Blikstein for his on-going support and guidance in my research. Additionally, I would like to recognize AT&T Labs and the Stanford Vice Provost of Graduate Education DARE Fellowship for supporting this research.

8. REFERENCES

[1] Dewey, J. (1897). My Pedagogic Creed. School Journal 54, 77-80.

[2] Dewey, J. (1913). Interest and effort in education. Cambridge, MA: The Riverside Press.

[3] Vygotsky, L.S. (1978). Mind and society: The development of higher psychological processes. Cambridge, MA: Harvard University Press

[4] Papert, S. (1980). Mindstorms : children, computers, and powerful ideas. New York: Basic Books.

[5] U.S. Department of Education, Office of Educational Technology (2010). Transforming American Education: Learning Powered by Technology. National Education Technology Plan 2010. Washington, DC.

[6] Kelly, T and Copabianco, B. (2012) Think-aloud Protocol Analysis as a Measure of Students' Science Learning through Design Assessment. Paper Presented at National Association for Research in Science Teaching Annual Meeting, March 25-28, 2012, Indianapolis, IN.

[7] Shulman, L. S. (2006). From hermeneutic to homelitic: The professional formation of clergy. Change. March/April, 28-31.

[8] Abrahamson, D. (2009). Orchestrating Semiotic Leaps from Tacit to Cultural Quantitative Reasoning—The Case of Anticipating Experimental Outcomes of a Quasi-Binomial Random Generator. Cognition and Instruction, 27(3), 175-224.

[9] Barron. B. (2006). Interest and self-sustained learning as catalysts of development: A learning ecologies perspective. Human Development, 49, 193-224.

[10] Barron, B. (2004). Learning ecologies for technological fluency: Gender and experience differences. Journal of Educational Computing Research, 31(1), 1-36.

[11] diSessa, A.A. 2002. Why "conceptual ecology" is a good idea. In M.Limón & L.Mason (Eds.), Reconsidering conceptual change: Issues in theory and practice (pp. 29–60). Dordrecht: Kluwer.

[12] Litman, D., Moore, J., Dzikovksa, M. and Farrow, E. (2009). Using Natural Language Processing to Analyze Tutorial Dialogue Corpora Across Domains and Modalities. Proceedings 14th International Conference on Artificial Intelligence in Education (AIED), Brighton, UK, July.

[13] Forbes-Riley, K., Rotaru, M. and Litman, J. (2009). The Relative Impact of Student Affect on Performance Models in a Spoken Dialogue Tutoring System. User Modeling and User-Adapted Interaction (Special Issue on Affective Modeling and Adaptation), 18(1-2), February, 11-43.

[14] Forbes-Riley, K. and Litman, D. 2010. Metacognition and Learning in Spoken Dialogue Computer Tutoring. Proceedings 10th International Conference on Intelligent Tutoring Systems (ITS), Pittsburgh, PA.

[15] Worsley, M. and Blikstein P. (2011). What's an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis. In Proceedings for the 4th Annual Conference on Educational Data Mining.

[16] Worsley, M. and Blikstein P. (2010) Towards the Development of Learning Analytics: Student Speech as an Automatic and Natural Form of Assessment. Paper Presented at the Annual Meeting of the American Education Research Association (AERA).

[17] Worsley, M. (2011). Qualifying Paper – What's an Expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis.

[18] Worsley, M. and Blikstein, P. (2012). An Eye For Detail: Techniques For Using Eye Tracker Data to Explore Learning in Computer-Mediated Environments. In the Proceedings of the 2012 International Conference of the Learning Sciences.

[19] Piech, C., Sahami, M., Koller, D., Cooper, S., & Blikstein, P. (2012). Modeling how students learn to program. Paper presented at the 43rd ACM technical symposium on Computer Science Education (SIGCSE '12).

[20] Worsley, M. and Blikstein, P. (2012). A Framework for Characterizing Changes in Student Identity during Constructionist Learning Activities. In the Proceedings of Constructionism 2012.

[21] Worsley, M. and Huang, Z. (2012). Semi-Supervised Object Tracking. (unpublished)

[22] Worsley, M. and Huang, Z (2012). Mining through Mobile: Using Smart Phones to Monitor Student Action, Affect and Interaction In Open-Ended Learning Environments. (unpublished)

[23] Worsley, M. and Johnston, M. (2010). Multimodal Interactive Spaces: MagicTV and MagicMAP. IEEE Workshop on Spoken Language Technology (SLT) Demonstration.

[24] Worsley, M., Johnston, M. and Blikstein P. (2011) OpenGesture: a low cost authoring framework for gesture and speech based application development and learning analytics. In Proceedings for the 10th Annual Conference on Interaction Design and Children.

[25] Worsley, M. and Blikstein P. (2012). OpenGesture: A Low-Cost, Easy-to-Author Application Framework for Collaborative, Gesture-, and Speech-Based Learning Applications. Paper Presented at the Annual Meeting of the American Education Research Association (AERA)

[26] Li, M. Ruiz-Primo, M.A. & Shaveloson, R.J. (2006). Towards a science achievement framework: The case of TIMSS 1999. In S. Howie & T. Plomp (Eds.), Contexts of

learning mathematics and science: Lessons learned from TIMSS (pp. 291–311). London: Routledge

[27] Litman, D., Moore, J., Dzikovksa, M. and Farrow, E. (2009). Using Natural Language Processing to Analyze Tutorial Dialogue Corpora Across Domains and Modalities. Proceedings 14th International Conference on Artificial Intelligence in Education (AIED), Brighton, UK, July.

[28] Worsley, M. and Blikstein P. (2010) Towards the Development of Learning Analytics: Student Speech as an Automatic and Natural Form of Assessment. Paper Presented at the Annual Meeting of the American Education Research Association (AERA).