2015 Multimodal Learning and Analytics Grand Challenge

Marcelo Worsley
Rossier School of Education
& Institute for Creative Technologies
University of Southern California
Los Angeles, CA, USA
worsley@usc.edu

Katherine Chiluiza
Escuela Superior Politécnica del Litoral
Guayaquil, Ecuador
kchilui@espol.edu.ec

Joseph F. Grafsgaard
Department of Psychology
North Carolina State University
Raleigh, NC, USA
jfgrafsg@ncsu.edu

Xavier Ochoa
Centro de Tecnologías de Información
Escuela Superior Politécnica del Litoral
Guayaquil, Ecuador
xavier@cti.espol.edu.ec

ABSTRACT
Multimodality is an integral part of teaching and learning. Over the past few decades researchers have been designing, creating and analyzing novel environments that enable students to experience and demonstrate learning through a variety of modalities. The recent availability of low cost multimodal sensors, advances in artificial intelligence and improved techniques for large scale data analysis have enabled researchers and practitioners to push the boundaries on multimodal learning and multimodal learning analytics. In an effort to continue these developments, the 2015 Multimodal Learning and Analytics Grand Challenge includes a combined focus on new techniques to capture multimodal learning data, as well as the development of rich, multimodal learning applications.

Categories and Subject Descriptors
H.5.m **[Information Interfaces and Presentation]**: Miscellaneous
J.4 **[Computer Applications]**: Social and Behavioral Sciences

General Terms
Performance, Design, Human Factors

Keywords
Design Challenge, Learning Environments, Human Computer Interaction

1. INTRODUCTION
The 21st century has seen an expansion in the set of tools available for assessing the quality of a given learning environment or experience (i.e., [2,3,16]. A number of the traditional tools, test and quiz performance, speeches and essays, are modes of expression that have been around for centuries and remain the more privileged forms of assessment. For all of their pedagogical shortcomings, these forms of assessment have the benefit of being widely accepted and easy to interpret. However, contemporary learning sciences research is increasingly concerned with additional constructs: motivation, engagement, collaboration, creativity, critical thinking, and problem solving, for example [1,13,17,34]. These are constructs that tend to be much harder to quantify using traditional testing instruments [8] and often necessitate adopting an alternative approach that more closely aligns with the design of said learning environment [22,27]. Fortunately, recent advances in multimodal human perception and low-cost multimodal sensory technology have made it possible to study complex human-human and human-machine interactions [15,18,23,28]. Multimodal learning analytics [5,24,32] works to leverage these advances in multimodal processing in order to address the challenges of studying a variety of complex learning-relevant constructs. In fact, recent reports have identified multimodal learning analytics [12,21], and multimodal evaluations [1], as important emerging areas of research. Multimodality allows researchers and practitioners to triangulate among the various ways that students may evidence learning while also allowing for a more complete and contextualized description of each student's experience and development.

ICMI '15, November 09-13, 2015, Seattle, WA, USA
© 2015 ACM. ISBN 978-1-4503-3912-4/15/11…$15.00
DOI: http://dx.doi.org/10.1145/2818346.2829995

Taking a multimodal perspective becomes increasingly salient in complex learning environments (e.g., Makerspaces, Collaborative Tangible User Interfaces). To this end, recent research in multimodal learning analytics has been particularly useful in studying non-traditional learning activities: collaborative mathematics problem-solving [19,20,33]; makerspaces [30,31]; computer programming [4,9,10], oral presentations [6,7,14] and collaborative tangible user interfaces learning [25,26]. Across these studies, researchers have used multimodal analytic techniques to model student performance, predict student learning, and, more importantly, construct models of student-student/student-artifact interactions. By constructing models of student-student interactions, as well as student-artifact interactions, multimodal learning analytics has been useful for better understanding the intricate learning processes that emerge in complex learning environments.

However, a significant hindrance to advancing multimodal learning analytics is the lack of high quality multimodal data capture tools that researchers and practitioners can easily use to collect meaningful process-oriented data. To address this, the 2015 Multimodal Learning and Analytics Grand Challenge focused on two design challenges that will hopefully advance the field's ability to capture and analyze high quality multimodal data. Both challenges are situated in the domain of multimodal teaching and learning, and should lend to the continuation of data driven challenges in future years.

2. CHALLENGE DESCRIPTIONS

Participation was solicited in two design categories: Multimodal Capture of Learning Environments and Multimodal Learning Applications: Incorporating Human Movement

2.1 Multimodal Capture of Learning Environments

The Multimodal Capture of Learning Environments challenge addresses the need to develop multimodal tools that can be used to effectively gather data from unstructured environments. While multimodal data capture can reasonably be completed in laboratory settings, with a small number of students, undertaking classroom wide multimodal data capture and analysis in authentic, everyday learning settings is quite challenging. However, we believe that the current availability of high quality multimodal sensor technology can significantly improve the state-of-the-art in this area. Furthermore, advances in this specific scenario should have applications across a variety of domains and scenarios outside of education. Submissions in this category were evaluated based on the quality, quantity and diversity of the data captured. Additionally, submissions were evaluated based on their ability to generalize to a variety of classroom learning contexts. Finally, authors were asked to focus on one or more areas for application optimization: data quality, cost, scalability, flexibility and intrusiveness. More specific details concerning submission evaluation guide can be found in the Evaluation and Guidelines section.

2.2 Multimodal Learning Applications: Incorporating Human Movement

The Multimodal Learning Application category sought submissions of software and/or hardware solutions that enable multimodal teaching and learning for one or more users. We were particularly interested in soliciting submissions that include recent low-cost motion sensors (e.g., Microsoft Kinect, Leap Motion, Myo). The low-cost motion sensors can easily be coupled with sensors from other modalities (Eye Tribe, Q-sensor/Empatica E4, etc.) Additionally, we encouraged participants to leverage existing software applications that ease the process of software development (e.g. the Institute of Creative Technology's Virtual Human Toolkit [11] and the LITE Lab's Generalized Intelligent Framework for Tutoring [29]).

For this design challenge, submissions were evaluated based on: how naturalistic the interactions are, the extent to which the affordances of the platform align with the learning goals, the quality of the learning (learning will be considered in the broad sense to include cognitive, socio-emotional, intuitive, etc.) that is taking place and how well the platform is able to record and leverage meaningful multimodal data in real-time. Authors were asked to include quantitative and/or qualitative results from a preliminary user study. A specific set of questions to ask of pilot test participants was provided in the Multimodal Learning Applications Evaluation and Design Guide. Additionally, solutions were rated based on the scalability of the platform (how easily can it be developed across multiple geographies and for multiple students) as well as its ability to offer real-time feedback.

3. EVALUATION CRITERIA

Each submission was evaluated by at least three reviewers that came from a variety of disciplines. Specific evaluation guidelines were provided for each sub-challenge. In the case of the Multimodal Capture of Learning Environments, submissions were evaluated based on the quality and quantity of data captured, and the ease with which that data could be interpreted and analyzed. For the Multimodal Learning Application challenge, submissions were evaluated based on how naturalistic the learning experience was, the quality of the learning gained from the interaction, and the platform's ability to capture and analyze rich multimodal data in real-time. More specific guidelines for the Multimodal Capture of Learning Environments category can be found below.

3.1 Multimodal Capture of Learning Environments Evaluation Guidelines

Capturing multimodal data that can effectively be used for learning analytics research can be challenging. In the submission guide, we provided some suggestions around the types and quality of data needed to conduct meaningful analyses. The suggestions for the modalities and data quality were based on prior work.

3.1.1 Cost

The total cost (both in economic and human resources) is an important variable that determines the feasibility of replicating and applying the capture design. Authors were asked to provide a rough estimate of the total cost of procuring, installing and operating the proposed solution.

3.1.2 Scalability

Learning environments vary widely in terms of size, from small working groups involving just two students to large auditoria with hundreds of students. Authors should report on how feasible it is for their design to adapt to larger or smaller settings.

3.1.3 Flexibility

Learning environments also vary widely in terms of disposition and arrangement, from traditional lecture settings with the instructor in front and students sitting at desks, to more active modes with workgroups of students sitting around tables and a group of instructors roaming among them. The authors should report how feasible it is for their design to adapt to different classroom arrangements.

3.1.4 Intrusiveness

Ideally, multimodal recording solutions for the classroom should be transparent for both the instructor and the students. In reality, different solutions create some level of intrusion (cameras pointing at participants, using special devices to conduct activities, wearing sensor or markers, etc.). As the natural behavior in the classroom could be altered by the recording setup, the authors should provide a small reflection on the intrusiveness of their designs.

3.2 Data Quality

3.2.1 Multimodal Fusion

Across all of the data there is an expectation that each data stream can be resolved to a specific individual (or, in some cases, a pair or triad of students), and to a properly synchronized point in time. Several audio/video, and software based approaches can be used for doing data synchronization, but this information should be clearly reflected in the submission text, as synchronization dramatically impacts data quality. The specific fusion strategy is based on the nature of the tasks and the data streams.

3.2.2 Skeletal Tracking (and Motion)

Skeletal tracking from the Kinect sensor should be frontal. For alternative motion data capture tools (e.g., 2D/3D computer vision algorithms, motion capture with markers) authors were instructed to see the appropriate guidelines on collecting high quality data as described in their documentation.

3.2.3 Audio

Prior work in speech and audio process has established a minimum frequency of 8Hz for speech recognition, and higher frequencies, between 12 Hz and 24 Hz, for conducting prosodic and spectral analysis. Additionally, in order to identify various learning relevant constructs (e.g. collaboration quality, voice quality and individual development) the audio captured should have the capability of being resolved to an individual speaker.

3.2.4 Video

One reason for capturing video data is to do facial expression analysis and head pose estimation. Facial expression analysis using computer-automated techniques typically requires a minimum ear to ear width of 64 pixels. A minimum size of 128 x 96 pixels is recommended, as some facial analysis software requires larger images in order to increase accuracy of detection. Additionally, a frontal view is required for being able to conduct facial expression and head pose analysis. However, the data capture environment is free to determine the appropriate frequency of frontal face capture, provided this is a modality of interest.

4. ACCEPTED PAPERS

4.1 Multimodal Capture of Learning Environments

Submissions in this category ranged from using multiple microphones to capture high quality audio data from classrooms (for the purpose of using ASR and studying dialogic patterns), to leveraging a Multimodal Recording Device that students can use to take multimodal selfies. From these two examples, we begin to see the breadth of approaches that can be used, and some of the technological, as well as ethical considerations that need to be taken into account when proposing the capture and analysis of multimodal data in a learning environment.

A third paper in this category provided an in-depth discussion of the current capabilities of depth sensors, and some of the analytic strategies that can help improve how researchers and designers use depth cameras to support learning.

4.2 Multimodal Learning Applications

Both accepted papers in this category were related to public speaking skills, one of the topics discussed during the 2014 Multimodal Learning Analytics Workshop and Grand Challenge. One paper focused on detecting different body postures (closed versus open) and sought to provide visual feedback to the user when they were exhibited closed poses. In this preliminary study, participants used more open poses when they were provided with the visual feedback about their body pose. The other paper looked at body posture in conjunction with audio-based features (effective use of pauses and presenter volume). It used a combination of visual, haptic and audio channels, in order to provide feedback to the user. These two papers are among of foray of research being conducted in the oral presentation and communication space, and will hopefully contribute to increasingly informative systems that can model and support good public speaking skills.

5. CONCLUSION

This year's challenge offered a noticeable change from previous years. As we look to advance the area of multimodal learning analytics, we need to simultaneously, or perhaps preemptively, push the boundaries in the area of multimodal data capture. The current work in laboratory settings is useful, but there is a need to transition this research to more authentic environments. Furthermore, there is an opportunity to take the work of multimodal learning analytics beyond traditional learning environments, into multimodal learning environments. These multimodal learning environments are not merely spaces where students continue their everyday classroom practices of doing algebra problems, or writing an essay. Multimodal learning environments transform learning by requiring students to employ a variety of modalities in order to effectively complete the task or assignment. It is our hypothesis that as multimodal learning environments are increasingly developed and used, research in multimodal learning and analytics will produce novel benefits for learners and new insights into learning processes above and beyond that of traditional learning.

6. REFERENCES

1. D Atkins, J Bennett, J S Brown, et al. 2010. Transforming American education: Learning powered by technology (National Educational Technology Plan 2010). *Washington, DC: US Department of Education, Office of Educational Technology*.

2. Ryan S J D Baker and Kalina Yacef. 2009. The state of educational data mining in 2009: A review and future visions. *JEDM-Journal of Educational Data Mining* 1, 1, 3–17.

3. Marie Bienkowski, Mingyu Feng, and Barbara Means. 2012. Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Department of Education, Office of Educational Technology*, 1–57.

4. Paulo Blikstein, Marcelo Worsley, Chris Piech, Mehran Sahami, Steve Cooper, and Daphne Koller. 2014. Programming pluralism: Using learning analytics to detect patterns in the learning of computer programming. *Journal of Learning Sciences*.

5. Paulo Blikstein and Marcelo Worsley. (in press) Multimodal Learning Analytics: a methodological framework for research in constructivist learning. *Journal of Learning Analytics*.

6. Lei Chen, Chee Wee Leong, Gary Feng, and Chong Min Lee. 2014. Using Multimodal Cues to Analyze MLA'14 Oral Presentation Quality Corpus: Presentation Delivery and Slides Quality. *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, ACM, 45–52. http://doi.org/10.1145/2666633.2666640

7. Vanessa Echeverría, Allan Avendaño, Katherine Chiluiza, Aníbal Vásquez, and Xavier Ochoa. 2014. Presentation Skills Estimation Based on Video and Kinect Data Analysis. *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, ACM, 53–60. http://doi.org/10.1145/2666633.2666641

8. JS Gorin and RJ Mislevy. 2013. Inherent Measurement Challenges in the Next Generation Science Standards for Both Formative and Summative Assessment. September. Retrieved March 17, 2014 from http://www.k12center.org/rsc/pdf/gorin-mislevy.pdf

9. Joseph F Grafsgaard, Joseph B Wiggins, Kristy Elizabeth Boyer, Eric N Wiebe, and James C Lester. 2014. Predicting Learning and Affect from Multimodal Data Streams in Task-Oriented Tutorial Dialogue. *7th International Conference on Educational Data Mining*, 122–129.

10. Joseph F Grafsgaard. 2014. Multimodal Analysis and Modeling of Nonverbal Behaviors During Tutoring. *Proceedings of the 16th International Conference on Multimodal Interaction*, ACM, 404–408. http://doi.org/10.1145/2663204.2667611

11. Arno Hartholt, David Traum, Stacy C Marsella, et al. 2013. All together now. *Intelligent Virtual Agents*, 368–381.

12. Larry Johnson, Samantha Adams, Malcolm Cummins, Victoria Estrada, Alex Freeman, and Holly Ludgate. 2013. The NMC horizon report: 2013 higher education edition.

13. Linda Katehi, Greg Pearson, and Michael Feder. 2009. The status and nature of K-12 engineering education in the United States. *The Bridge* 39, 3, 5–10.

14. Gonzalo Luzardo, Bruno Guamán, Katherine Chiluiza, Jaime Castells, and Xavier Ochoa. 2014. Estimation of Presentations Skills Based on Slides and Audio Features. *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, ACM, 37–44. http://doi.org/10.1145/2666633.2666639

15. Marwa Mahmoud, Louis-Philippe Morency, and Peter Robinson. 2013. Automatic multimodal descriptors of rhythmic body movement. *Proceedings of the 15th ACM on International conference on multimodal interaction - ICMI '13*, 429–436. http://doi.org/10.1145/2522848.2522895

16. Taylor Martin and Bruce Sherin. 2013. Learning analytics and computational techniques for detecting and evaluating patterns in learning: An introduction to the special issue. *Journal of the Learning Sciences* 22, 4, 511–520.

17. National Research Council. 2012. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press.

18. Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. 2011. Multimodal deep learning. *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 689–696.

19. Xavier Ochoa, Katherine Chiluiza, Gonzalo Méndez, Gonzalo Luzardo, Bruno Guamán, and James Castells. 2013. Expertise estimation based on simple multimodal features. *Proceedings of the 15th ACM on International conference on multimodal interaction - ICMI '13*, 583–590. http://doi.org/10.1145/2522848.2533789

20. S Oviatt and A Cohen. 2013. Written and multimodal representations as predictors of expertise and problem-solving success in mathematics. *Proceedings of the 15th ACM International Conference on Multimodal Interaction – ICMI '13*, ACM, 599-606. http://dx.doi.org/10.1145/2522848.2533793

21. Roy Pea. 2014. THE LEARNING ANALYTICS Workgroup A Report on Building the Field of Learning Analytics.

22. Jean Piaget. 1973. To understand is to invent: the future of education (G. Roberts, Trans.). *NY: Grossman Publishers*.

23. Stefan Scherer, Michael Glodek, Georg Layher, et al. 2012. A generic framework for the inference of user states in human computer interaction. *Journal on Multimodal User Interfaces* 6, 3-4, 117–141.

24. Stefan Scherer, Marcelo Worsley, and Louis-Philippe Morency. 2012. 1st international workshop on multimodal learning analytics. *ICMI*, 609–610.

25. Bertrand Schneider and Paulo Blikstein. Unraveling Students' Interaction Around a Tangible Interface using Multimodal Learning Analytics. *Journal of Educational Data Mining*.

26. Bertrand Schneider, Jenelle Wallace, Paulo Blikstein, and Roy Pea. 2013. Preparing for future learning with a tangible user interface: the case of neuroscience. *Learning Technologies, IEEE Transactions on* 6, 2, 117–129.

27. Daniel L Schwartz. 1992. Constructivism in an age of non-constructivist assessments.

28. Yale Song, Louis-Philippe Morency, and Randall Davis. 2012. Multimodal human behavior analysis: learning correlation and interaction across modalities. *Proceedings of the 14th ACM international conference on Multimodal interaction*, 27–30.

29. Robert A Sottilare, Keith W Brawner, Benjamin S Goldberg, and Heather K Holden. 2012. The generalized intelligent framework for tutoring (GIFT).

30. Marcelo Worsley and Paulo Blikstein. 2013. Towards the Development of Multimodal Action Based Assessment. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, ACM, 94–101. http://doi.org/10.1145/2460296.2460315

31. Marcelo Worsley and Paulo Blikstein. 2015. Leveraging Multimodal Learning Analytics to Differentiate Student Learning Strategies. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, ACM, 360–367. http://doi.org/10.1145/2723576.2723624

32. Marcelo Worsley. 2012. Multimodal learning analytics: enabling the future of learning through multimodal data analysis and interfaces. *Proceedings of the 14th ACM international conference on Multimodal interaction*, 353–356.

33. Jianlong Zhou, Kevin Hang, Sharon Oviatt, Kun Yu, and Fang Chen. 2014. Combining Empirical and Machine Learning Techniques to Predict Math Expertise Using Pen Signal Features. *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, ACM, 29–36. http://doi.org/10.1145/2666633.2666638

34. 2013. Next Generation Science Standards: For States, By States.