(2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213–219.
http://dx.doi.org/10.18608/jla.2016.32.10

Augmenting Learning Analytics with Multimodal Sensory Data

Xavier Ochoa
Escuela Superior Politécnica del Litoral, Ecuador
xavier@cti.espol.edu.ec

Marcelo Worsley
Northwestern University, USA
marcelo.worsley@northwestern.edu

ABSTRACT: The goal of learning analytics (LA) is to understand and improve learning. However, learning does not always occur through or mediated by a technological system that can collect digital traces. To be able to study learning in such environments, several signals, such as video and audio, should be captured, processed, and analyzed to produce traces of the actions and interactions of the actors in the learning process. The use and integration of the different modalities present in those signals is known as multimodal learning analytics (MLA). This editorial presents a brief introduction to this new approach to traditional learning analytics and summarizes four representative articles included in this special section. The editorial closes with a short discussion regarding the current opportunities and challenges in MLA.

Keywords: Multimodal, signal capture, signal integration, education, complex learning environments

1 INTRODUCTION

The main focus of the field of learning analytics has been the study of the actions that students perform while using digital tools. Traditionally these digital tools include learning management systems (LMSs), intelligent tutoring systems (ITSs), massive open online courses (MOOCs), educational video games, or other types of systems that use a computer as an active component in the learning process. On the other hand, comparatively less learning analytics research has been conducted in other learning contexts, such as collaborative face-to-face contexts where computers are not present or have an auxiliary role. This bias towards computer-based learning contexts is well explained by the typical requirement of a learning analytics system to capture a trace of the learning experience.

Computer-based learning systems, even if not initially designed with analytics in mind, provide a means to capture fine-grained human–computer interactions easily. The data describing these interactions is stored in many forms — for example log-files or word-processor documents — that can be processed in real-time or analyzed post-hoc. The relative abundance of readily available data and the low technical barriers to process it make computer-based learning systems the ideal place to conduct LA research. On the contrary, in learning contexts where computers are not used, the actions of the actors in the

(2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213–219.
http://dx.doi.org/10.18608/jla.2016.32.10

learning process are not automatically captured. Even if some learning artifacts exist, such as student-produced physical documents, they need to be converted before they can be processed. Without traces to be analyzed, computational models and tools used in traditional LA are not applicable.

The existence of this bias towards computer-based learning contexts could produce a streetlight effect (Freedman, 2010) in LA, which means looking for solutions where it is easy to search, not where the real solution might be. Within much of LA, investigators are trying to understand and optimize the learning process by looking at what is happening in computer-based contexts, but ignoring real-world environments. Even learner actions that occur in computer-based systems but cannot be logged are usually ignored. For example, a student visibly expressing confusion when presented with a problem in an ITS, or yawning while watching an online lecture, is likely to be overlooked. To avoid the streetlight effect, researchers are now focusing on how to collect fine-grained learning traces from real-world learning contexts automatically, making the analysis of a face-to-face lecture as feasible as the analysis of a MOOC session. This is a main objective of multimodal learning analytics.

In an effort to advance this line of research, a group of researchers organized the 1st International Workshop on Multimodal Learning Analytics at the 2012 International Conference on Multimodal Interaction (Scherer, Worsley, & Morency, 2012). This workshop represented a burgeoning community of learning scientists and computer scientists who shared a mutual interest in understanding and improving learning through recording and measuring the different actions and interactions present in learning contexts and the application of techniques from artificial intelligence and machine learning to make sense of that information. Moreover, the community shared this desire to take the study of learning to new places and explore new paradigms, previously deemed too complex or too naturalistic. This special section of the *Journal of Learning Analytics* will highlight some of the methodological advances that the field has made over the course of five years. Specifically, this special section includes four papers that serve to provide readers with a more in-depth introduction to multimodal learning analytics and reference some of the seminal literature in this field.

2 A BRIEF DESCRIPTION OF MULTIMODAL LEARNING ANALYTICS

Multimodal learning analytics works to leverage advances in multimodal data capture and signal processing to address the challenges of studying a variety of complex learning-relevant constructs as observed in complex learning environments. Some examples of multimodal data include speech, video, electro cardiology, and eye tracking (for a more detailed list with descriptions, see Blikstein & Worsley's paper in this section). Whereas the areas of educational data mining and learning analytics have significantly benefited from the ability to capture trace data from an individual student's work within computer mediated learning environments, a primary goal for multimodal learning analytics is the ability to study collaborative, real-world, non-computer mediated environments. Like many of the previous studies in multimodal learning analytics — e.g., collaborative mathematics problem-solving (Ochoa et al., 2013; Oviatt & Cohen, 2013); makerspaces (Worsley & Blikstein, 2013, 2015); computer programming (Blikstein et al., 2014; Grafsgaard, Wiggins, Boyer, Wiebe, & Lester, 2014); and

(2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213–219.
http://dx.doi.org/10.18608/jla.2016.32.10

collaborative tangible user interfaces learning (Schneider, Wallace, Blikstein, & Pea, 2013) — the papers featured in this special section share elements of that objective. Furthermore, the adoption of multimodal sensory data represents more than an effort to employ the newest gadget; it is instead done to make analysis possible and tractable. In some cases, utilizing multimodal sensory data provides a glimpse into physiological measures that would be nearly impossible for humans to perceive. Furthermore, many of the affordances of multimodal data capture and processing move beyond a simple desire for automation, and are instead motivated by a need to go deeper in one's analysis of intra- and interpersonal interactions. In this way, we see that while we often use such terms as "automatic," "computation," and "machine learning," multimodal learning analytics is not only quantitative in nature (Worsley et al., 2016).

3 SUMMARY OF FEATURED ARTICLES

3.1 Multimodal Learning Analytics and Education Data Mining: Using Computational Technologies to Measure Complex Learning Tasks

The special section opens with an introductory paper by Blikstein and Worsley that aims to further motivate the potential for multimodal learning analytics to provide novel approaches for studying and modelling learner experience in open-ended student-centered learning environments (e.g. Papert, 1980). The paper provides example research from a variety of modalities: text, speech, handwriting, sketch, actions/gestures, affect, neurophysiology, and eye gaze. It also discusses prior research that brings together data from several of the aforementioned modalities. One area that does not receive considerable discussion is computer vision. While the authors do highlight some prior work that utilizes gesture- and action-based analysis, the other three papers in this special section provide more in-depth examples of video-derived analyses.

3.2 Sleepers' Lag: Study on Motion and Attention

In this pioneering work, Raça, Tormey, and Dillenbourg study the level of student attention during a traditional lecture through video recordings and computer vision techniques. Instead of conducting the analysis on individual students, Raça et al. focus their study on the relative differences between learner reaction times, measured automatically by their body movements. Methodologically this paper makes an important contribution by providing an extremely non-invasive way to study student attention across an entire classroom without the need for significant instrumentation. In addition to their methodological contribution, Raça et al. also presents learning-related findings that suggest that less attentive students have a "sleeper's lag" when compared with self-reported attentive students. The setup presented in the article could be used to assess the level of attention in a classroom automatically and provide feedback to the instructor during or after the lecture.

(2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213–219.
http://dx.doi.org/10.18608/jla.2016.32.10

3.3 Designing an Automated Assessment of Public Speaking Skills Using Multimodal Cues

Chen et al., present a well-designed experiment to find useful multimodal features for the automatic estimation of public speaking skills. Due to limitations in existing datasets, they carefully prepare a new set of 56 recordings of public presentations, together with human assessment of the different dimensions of those skills. Apart from being an example for further multimodal experimentation, this article provides evidence of the feasibility of the automatic assessment of public speaking skills through multimodal learning analytics. It is worth noting that while unimodal analysis provides, independently, weak prediction models, the combination of the information from different modalities creates a more precise and robust model.

3.4 Using Multimodal Learning Analytics to Model Student Behaviour: A Systematic Analysis of Epistemological Framing

Finally, Andrade, Delandshere, and Danish demonstrate how traditional education research could benefit from multimodal analysis. Their dataset includes interviews with first and second graders about basic science concepts. They use information from posture, gesture, gaze, language, and speech to predict the different epistemological frames that students adopt during interviews and, based on the clustering of these frames, establish the level of mechanistic reasoning done by the students. This work is an example of a multimodal learning analytics study guided by theory, but that has the potential to inform the construction of new theories based on empirical findings.

4 DISCUSSION: OPPORTUNITIES AND CHALLENGES

Multimodal learning analytics is still a nascent field with a small but very active and open community of researchers. The existence of regular challenges and workshops (Morency, Oviatt, Scherer, Weibel, & Worsley, 2013; Ochoa, Worsley, Weibel, & Oviatt, 2016; Ochoa, Worsley, Chiluiza, & Luz, 2014; Scherer, Worsley, & Morency, 2012; Worsley, Chiluiza, Grafsgaard, & Ochoa, 2015; Worsley et al., 2016), where multimodal datasets are freely shared and jointly analyzed and new designs ideas are openly discussed, create a research environment where new knowledge is generated rapidly.

The availability and affordability of sophisticated multimodal sensors facilitates the collection of high-frequency data about the actions and interactions of learners in a learning context. These new affordances provide the research community with an instrument to observe and analyze learning in ways that were not possible before. As such this research could foster new understandings and theories about learning, and also influence the design of learning interfaces and experiences. The papers presented in this special section are provide a small but excellent sample of the way multimodal learning analytics can improve the learning process. However, several issues still prevent this line of research from entering mainstream practice:

- **Recording:** Capturing media in learning contexts that are not computer-mediated requires the

(2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213–219.
http://dx.doi.org/10.18608/jla.2016.32.10

procurement, installation, and use of multimodal sensors: cameras, microphones, digital pens, etc. These recording devices need to be synchronously enabled and the appropriate measures taken into account for the various forms of noise that can degrade analysis. These measures include accounting for occlusions, interference, and/or noise through advanced data capture configuration, or through post-processing algorithms. Having recording systems that could work as effortlessly and efficiently as computer logging is a primary barrier for making multimodal learning analytics mainstream, and would facilitate data capture at scale. While commercial solutions do exist for conducting multimodal data capture, these are quite limited in their ability to capture data from multiple individuals and from non-laboratory-based environments.

- **Privacy:** The capture of interaction information in digital tools already raises privacy questions among students and instructors. The installation and use of recording systems that amount, and even surpass, "1984" levels of surveillance, is bound to confront strong resistance. The option of signing informed consent forms could work for early research stages, but adoption of multimodal data collection systems in the real-world would require different, more creative, distributed approaches regarding data ownership and control. Prior work has begun to address this question (Domínguez, Chiluiza, Echeverria, & Ochoa, 2015) but there remains a great deal of work to do on learner privacy, especially as we move closer to using multimodal learning analytics for real-time user feedback.

- **Data Fusion:** One of the questions that arises with the availability of large amounts of raw learning traces is how to combine the data to produce useful information to understand and optimize the learning process. Traces extracted from different modes and with different extraction processes are bound to have very different characteristics. For example, the time granularity of the traces extracted from different modes can vary widely or the level of certainty of the extracted traces can be different. As has been a central question in multimodal signal processing, determining the appropriateness of different fusion strategies (e.g., decision fusion, or feature fusion) is an important consideration when looking at different learning-related outcomes. A closely related challenge is determining effective ways for representing and segmenting multimodal data streams. Some work has begun to bring these questions more to the forefront of the multimodal learning analytics community (Worsley, 2014; Worsley, Scherer, Morency, & Blikstein, 2015), but considerably more research is needed to address these questions.

- **Impact on Learning:** While the end-user tools and interventions based on multimodal learning analytics are similar to those based on unimodal analyses, the required usefulness of multimodal analysis should be higher to justify the additional complexity of data acquisition. The increased impact or understanding of learning should be in proportion to the increased complexity. The requirement of using multiple real-world signals to analyze learning should also come with the promise to provide more useful insights and more measurable impacts on learners. That said, it is also important for the multimodal analyses to incorporate some of the

(2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213–219. http://dx.doi.org/10.18608/jla.2016.32.10

measures traditionally used within mainstream education research studies. Without a concrete and valid connection to existing or prior research in education and the learning sciences, it will be difficult for multimodal learning analytics to have true impact.

These issues, far from being insurmountable, are the source of active research lines that constantly provide innovative solutions to improve the affordability and feasibility of LA and MLA practices.

Finally, the editors of this special section invite LA researchers and practitioners to explore the use of multiple modalities in their own studies and tools. The multimodal LA community will openly share its knowledge, data, code, and frameworks. Only by embracing these different modalities will LA have an impact across the diversity of contexts in which learning takes place.

REFERENCES

Blikstein, P., Worsley, M., Piech, C., Gibbons, A., Sahami, M., Cooper, S., & Koller, D. (2014). Programming pluralism: Using learning analytics to detect patterns in novices' learning of computer programming. *International Journal of the Learning Sciences, 4*(23), 561–599. http://dx.doi.org/10.1080/10508406.2014.954750

Domínguez, F., Chiluiza, K., Echeverria, V., & Ochoa, X. (2015). Multimodal selfies: Designing a multimodal recording device for students in traditional classrooms. *Proceedings of the 17th ACM International Conference on Multimodal Interaction* (ICMI '15), 567–574. http://dx.doi.org/10.1145/2818346.2830606

Freedman, D. H. (2010). Why scientific studies are so often wrong: The streetlight effect. *Discover Magazine*, *26*. Retrieved from http://discovermagazine.com/2010/jul-aug/29-why-scientific-studies-often-wrong-streetlight-effect

Grafsgaard, J., Wiggins, J., Boyer, K. E., Wiebe, E., & Lester, J. (2014). Predicting learning and affect from multimodal data streams in task-oriented tutorial dialogue. In J. Stamper, S. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM 2014), (pp. 122-130). Retrieved from http://www.educationaldatamining.org/EDM2014/

Morency, L., Oviatt, S., Scherer, S., Weibel, N., & Worsley, M. (2013). ICMI 2013 Grand Challenge Workshop on Multimodal Learning Analytics. *Proceedings of the 15th ACM International Conference on Multimodal Interaction* (ICMI '13), 373–378. http://dx.doi.org/10.1145/2522848.2534669

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). Expertise estimation based on simple multimodal features. *Proceedings of the 15th ACM International Conference on Multimodal Interaction* (ICMI '13), 583–590. http://dx.doi.org/10.1145/2522848.2533789

Ochoa, X., Worsley, M., Weibel, N., & Oviatt, S. (2016). Multimodal learning analytics data challenges. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 498–499. http://dx.doi.org/10.1145/2883851.2883913

Ochoa, X., Worsley, M., Chiluiza, K., & Luz, S. (2014). *Proceedings of the 2014 ACM workshop on*

(2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213–219. http://dx.doi.org/10.18608/jla.2016.32.10

Multimodal Learning Analytics Workshop and Grand Challenge (MLA '14), 531–532. http://dx.doi.org/10.1145/2663204.2668318

Oviatt, S., & Cohen, A. (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), 599–606. http://dx.doi.org/10.1145/2522848.2533793

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.

Scherer, S., Worsley, M., & Morency, L. P. (2012). 1st International Workshop on Multimodal Learning Analytics: Extended Abstract. *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (ICMI '12), 609–610. http://dx.doi.org/10.1145/2388676.2388803

Schneider, B., Wallace, J., Blikstein, P., & Pea, R. (2013). Preparing for future learning with a tangible user interface: The case of neuroscience. *IEEE Transactions on Learning Technologies, 6*(2), 117–129. http://dx.doi.org/10.1109/TLT.2013.15

Worsley, M. (2014). Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors. *Proceedings of the 16th ACM International Conference on Multimodal Interaction* (ICMI '14), 1–4. http://dx.doi.org/10.1145/2666633.2666634

Worsley, M., Abrahamson, D., Blikstein, P., Grover, S., Schneider, B., & Tissenbaum, M. (2016). Situating multimodal learning analytics. In C.-K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Proceedings of the International Conference of the Learning Sciences: Transforming Learning, Empowering Learners* (ICLS '16), (Vol. 2, pp. 1346–1349). Singapore: International Society of the Learning Sciences.

Worsley, M., & Blikstein, P. (2013). Towards the development of multimodal action based assessment. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 94–101. http://dx.doi.org/10.1145/2460296.2460315

Worsley, M., & Blikstein, P. (2015). Leveraging multimodal learning analytics to differentiate student learning strategies. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 360–367. http://dx.doi.org/10.1145/2723576.2723624

Worsley, M., Chiluiza, K., Grafsgaard, J., & Ochoa, X. (2015). 2015 Multimodal Learning and Analytics Grand Challenge. *Proceedings of the 17th ACM International Conference on Multimodal Interaction* (ICMI '15), 525–529. http://dx.doi.org/10.1145/2818346.2829995

Worsley, M., Scherer, S., Morency, L. P., & Blikstein, P. (2015). Exploring behavior representation for learning analytics. *Proceedings of the 17th ACM International Conference on Multimodal Interaction* (ICMI '15), 251–258. http://dx.doi.org/10.1145/2818346.2820737