

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

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

Paulo Blikstein

Graduate School of Education and (by courtesy) Computer Science Department
Stanford University, USA
paulob@stanford.edu

Marcelo Worsley

Learning Sciences & Computer Science
Northwestern University, USA
marcelo.worsley@northwestern.edu

ABSTRACT: New high-frequency multimodal data collection technologies and machine learning analysis techniques could offer new insights into learning, especially when students have the opportunity to generate unique, personalized artifacts, such as computer programs, robots, and solutions engineering challenges. To date most of the work on learning analytics and educational data mining has been focused on online courses and cognitive tutors, both of which provide a high degree of structure to the tasks, and are restricted to interactions that occur in front of a computer screen. In this paper, we argue that multimodal learning analytics can offer new insights into student learning trajectories in more complex and open-ended learning environments. We present several examples of this work and its educational applications.

Keywords: Learning analytics, multimodal interaction, constructivism, constructionism, assessment

1 INTRODUCTION

The same battle is fought in every field of educational research and practice: the champions of the direct instruction of well-defined content pitted against those who encourage student-centred exploration of ill-defined domains. These wars have taken place repeatedly over past decades, and partisans on each side have been reborn in multiple incarnations. The first tradition tends to be aligned with behaviourist or neo-behaviourist approaches, while the second favours constructivist-inspired pedagogies. In language arts, the battle has been between phonics and the whole word approach. In math, war is waged between teaching algorithms versus instruction in how to think mathematically. In history, they clash over the relative merits of critical interpretations and the memorization of historical facts. In science, they clash about inquiry-based approaches versus direct instruction of formulas and principles.

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

The educational research community has always maintained that the debate would end when research results inevitably demonstrated the superiority of one of the sides. Yet this conclusion has eluded scholarship for decades. One of the reasons for this interminable contest is that the underlying rationale for the differences concerns individual values and societal beliefs and will not be resolved by a purely scientific approach. In fact, the whole debate may serve the educational research community in quite a different capacity. More specifically the debates may reveal the underlying visions for what education should be about, for different groups, and we might more profitably re-examine the nature and purpose of our schools. More fundamentally, is education a tool for filtering, ranking, emancipation, social equalization, economic progress, meritocracy, or for the promotion of social Darwinism?

Educational scholars would greatly differ in their answers — rendering the question of “which approach is better,” and “what evidence suffices,” pointless. As with the debates on public healthcare and fiscal policy, despite our best efforts to generate reliable research, the “best” way to conduct education will always be controversial and dependent on larger societal and political winds. But the fundamental problem, and the motivation for this article, is that the prevailing issue is not who “wins” the debate, but rather the existence of a healthy debate. Fostering a healthy debate requires some level of symmetry. However, as it stands, the playing field is not symmetrical. The “direct instruction” approach is inherently easier to test and quantify using currently available tools that include mass-production of content and decades of research concerning psychometrics and standardized testing strategies. Meanwhile, the constructivist side counts on laborious interventions, and complex mixed-mode research methods. The result of this asymmetry is that public systems, more dependent on high-profile research results, are left, by inertia, to the designs of the proponents of traditional approaches, while only affluent schools, private or public, who can experiment more, can afford to implement modern, constructivist approaches to learning.

Learning analytics could deepen this asymmetry, or help eliminate it. The elimination of the asymmetry could re-establish a healthy public debate around education, where both sides would have comparable and credible results to show, and policy makers would be able to make choices based on their values and visions for education. However, the deepening of this asymmetry could be a significant impediment to progressive education and the vision of creating alternative learning environments that can reach a more diverse population of learners. Should public education succumb to the temptation of the fiscal benefits supposedly offered by total automatization and its much lower baseline for cost and quality, all other options would be driven into the ground as economically unfeasible: who could compete with virtually free computerized tutors and videos? How many years would the debate take, while children caught in the “experimental” years are being victimized?

Consequently, we propose that an important goal of learning analytics is to equalize the playing field by developing methods that examine and quantify non-standardized forms of learning. We suggest that this need for a level playing field is more necessary than ever, given the increasing demand for scalable project-based, interest-driven learning and student-centred pedagogies (e.g., Papert, 1980). Within our

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

increasingly interconnected societal and economic environment — which has become pervaded by technology and threatened by challenging global problems such as climate change — both K–12 and university-level engineering education (Dym, 1999) demand higher-level, complex problem-solving as opposed to performance in routine cognitive tasks (Levy & Murnane, 2004). Approaches that place premiums on student-centred, constructivist, self-motivated, self-directed learning have been advocated for decades (e.g., Dewey, 1902; Freire, 1970; Montessori, 1965; Barron & Darling-Hammond, 2010) but have failed to become scalable and prevalent, and have come under attack during the last decade (e.g., Kirschner, Sweller, & Clark, 2006; Klahr & Nigam, 2004).

New high-frequency data collection technologies and machine learning could offer new insights into learning in tasks in which students are allowed to generate unique, personalized artifacts, such as computer programs, robots, movies, animations, and solutions to engineering challenges. To date most of the work on learning analytics and educational data mining has focused on tasks that are computer-mediated and are more structured and scripted. In this work, we argue that *multimodal data collection and analysis techniques* (“multimodal learning analytics” or MMLA) could yield novel methods that generate distinctive insights into what happens when students create unique solution paths to problems, interact with peers, and act in both the physical and digital worlds.

Assessment and feedback is particularly difficult within these open-ended environments, and these limitations have hampered many attempts to make such approaches more prevalent. Automated, fine-grained data collection and analysis could help resolve this tension in two ways. First, such capacities would give researchers tools to examine student-centred learning in unprecedented scale and detail. Second, these techniques could improve the scalability of these pedagogies since they make feasible both assessment and formative feedback, which are typically very complex and laborious in such environments. They might not only reveal students’ trajectories throughout specific learning activities, but they could also help researchers design better supports, pedagogical approaches, and learning materials.

At the same time, in the well-established field of *multimodal interaction*, new data collection and sensing technologies are making it possible to capture massive amounts of data in all fields of human activity. These technologies include the logging of computer activities, wearable cameras, wearable sensors, biosensors (e.g., that permit measurements of skin conductivity, heartbeat, and electroencephalography), gesture sensing, infrared imaging, and eye tracking. Such techniques enable researchers to have unprecedented insight into the minute-by-minute development of a number of activities, especially those involving multiple dimensions of activity and social interaction. However, the technologies just mentioned have not yet become popular in the field of learning analytics. We propose that *multimodal learning analytics* could bring together these multiple techniques in more comprehensive evaluations of complex cognitive abilities, especially in environments where the processes or outcomes are unscripted. Thus, the goal of this paper is to demonstrate the feasibility and power of novel assessment techniques in several modalities and learning contexts.

ISSN 1929-7750 (online). The Journal of Learning Analytics works under a Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

2 STATE OF THE FIELD

In considering the current sensing and assessment modalities possible using MMLA, we see three non-mutually exclusive areas: assessing student knowledge, assessing student affect and physiology, and assessing student intentions or beliefs. At the crux of all these forms of student characterization is the underlying invocation of data analysis to generate useful models from large sets of quantitative data. Hence, what varies in the different forms of student assessment is the source of the raw data and how that data is translated into computable data. Once the translation has been completed, the data is processed using a collection of machine learning algorithms. In what follows, we present several methods being used to capture and process student data. There are several techniques — web data mining, user data mining, simple web-based surveys, etc. — but the following technologies have been selected for inclusion because they live on the cutting-edge of technology and help promote the notion of “natural” assessment (Zaïane, 2001). Furthermore, while each of these technologies represents a research contribution in and of itself, our interest in including them is to bring to the forefront a wider variety of non-traditional approaches that education researchers and educational data scientists can begin to combine in their learning analytics research. For the first three techniques we mention — text analysis, speech analysis, and handwriting analysis — our discussion will be very cursory, as these represent areas of research that have received considerable attention with the computer science community, and have started to get traction within the learning analytics community. Nonetheless, we want to make the reader aware of some of the current capabilities and research in these areas. For the latter analyses that we discuss, we will engage in a more detailed and descriptive explanation of each, as these domains remain relatively new, even among the computer science community.

2.1 Text Analysis

While text analysis, or natural language processing, has been around for decades it is only in recent history that education has begun to benefit from this technology, and researchers have targeted learners’ text explicitly. Despite the fact that text itself is not multimodal, text analysis allows for the interpretation of open-ended writing tasks, differently from multiple-choice tests. Given that collecting text from students is unproblematic both technically and logistically, it constitutes one of the most promising modalities for MMLA: text can be easily gathered from face-to-face and online activities, from tests and exams, and from expert-generated prose from textbooks and online sources (often used as baseline). For example, Sherin (2013) has been doing pioneering work in the analysis of text in the learning sciences community. He uses techniques from topic modelling and clustering to study the progression of students’ ideas and intuitions as they describe the explanation for the existence of the four seasons (Sherin, 2013). More specifically, he shows that, as students explain the seasons, invoking different types of scientific explanations, it is possible to identify which type of explanation each student is referring to at different points in time. He also goes beyond this to show how students can be accurately clustered without using a pre-defined set of exemplar responses, but instead by using automatically derived topics models from the corpus itself. This approach of clustering segments of

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

students' text based on the descriptions of their peers is a powerful tool that can allow researchers and practitioners to draw meaningful commonalities and differences among large populations of students, without having to explicitly read and compare the entirety of each transcript. Given the prevalence of text-based assessment and the intensive use of text in face-to-face and online learning, this promising method will likely accelerate discourse-based research, and open new possibilities for large-scale analysis of open-ended text corpora.

2.2 Speech Analysis

Speech analysis shares many of same goals and tools as text analysis. Speech analysis, however, further removes the student from the traditional assessment setting by allowing them to demonstrate fluency in a more natural setting. For example, Worsley & Blikstein (2011) studied how elements of student speech, as inferred by linguistic, textual, and prosodic features, can be predictive for identifying students' level of expertise on open-ended engineering design tasks. In addition to traditional linguistic and prosodic features, speech signals can be analyzed for a wealth of other characteristics. Various research tools have been developed to help researchers in the process of extracting these features, however, several challenges remain in knowing how to analyze student learning appropriately using said features.

Other researchers have moved away from raw analysis of the speech signal to leverage speech recognition capabilities. In particular, Beck and Sison (2006) demonstrated a method for using speech recognition to assess reading proficiency. As an extension of Project LISTEN — an intelligent tutor that helps elementary school students improve their reading skills — researchers completed a study that combines speech recognition with knowledge tracing, a form of probabilistic monitoring. By having a language model that was largely restricted to the content of each book being learned, the work required for doing automatic speech recognition, and subsequent accuracy classification, was greatly simplified. Outside of the education domain there have been decades of work in developing speech recognizers and dialogue managers. However, to date, such technologies are still not widely used in education because of the challenges associated with building a satisfactory language model that can reliably recognize speech. Munteanu, Peng, and Zhu (2009) have made some progress in this area by showing how to improve speech recognition of lectures in college-level STEM class. A primary consideration in the area of speech recognition, therefore, will be to identify the most effective ways to use this technology in real-world educational settings. Although using it to transcribe lectures might be feasible, the challenge of collecting and interpreting student data seems extremely difficult. Differently from other applications of speech recognition (smartphones, personal assistants, dictation), educational applications need to address simultaneously classroom noise, multiple overlapping speakers, and logistical difficulties in voice training — very ambitious challenges that have not been solved yet.

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

2.3 Handwriting Analysis

A different form of text analysis is handwriting analysis, which is important in educational settings because a considerable part of the work done by students is still handwritten. Anthony, Yang, and Koedinger (2007) highlight the affordances of combining handwriting recognition with intelligent tutors for algebra. Based on their study of high school and middle school students, introducing handwriting recognition halved the time students needed to complete tutoring activities because students no longer had to deal with cumbersome keyboard and mouse-based entry. This is significant because it enables students to focus on understanding the material using familiar forms of interaction as opposed to struggling to learn a new interface. Accordingly, handwriting recognition can facilitate more effective learning by eliminating the barriers to using certain computer-based interfaces. It also permits the student to learn in such a way that more closely parallels the usual mathematics environment (i.e., utilizing a writing tool as opposed to a keyboard), which may increase transfer.

Researchers also studied the use of handwriting recognition technology among school-aged children (Read, 2007), examining the length and quality of stories produced by students using different input methods. A primary finding of this work was that students were more willing to engage in the writing process when using digital ink than when using traditional keyboard input. However, the team still found that handwriting recognition technology was not yet comparable to traditional paper and pencil. Similarly to Anthony et al. (2007), Read (2007) emphasizes the affordances of handwriting as a more natural form of authorship that may help students better engage in learning.

More recent work extends handwriting recognition to mid-air “writing” that achieves high levels of accuracy by utilizing a combination of computer vision, multiple cameras, and machine learning (Schick, Morlock, Amma, Schultz, & Stiefelhagen, 2012). This approach highlights some of the more recent opportunities in handwriting recognition in novel learning environments and contributes to the discussion around the expansive possibilities available away from traditional keyboards and screens.

2.4 Sketch Analysis

Whereas handwriting analysis is primarily concerned with looking for words, others researchers have embarked on work that looks at both text-based and graphic-based representations. Fundamental work on object recognition and sketches is that of Alvarado, Oltmans, and Davis (2002) and Alvarado and Davis (2006). These researchers developed a framework for performing multi-domain recognition of sketches using Bayesian networks and a predetermined set of shapes and patterns for each domain. With the predefined shapes and patterns, their algorithm is able to decipher messy sketches from the domains of interest.

Ken Forbus and colleagues also describe seminal work in the development of both systems and techniques for analyzing and comparing sketches among learners. For example, Jee, Gentner, Forbus,

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

Sageman, & Uttal (2009) explain the design and implementation of CogSketch, a tool used to study how students of different levels of experience describe common scientific concepts in geology through sketches (Forbus, Usher, Lovett, Lockwood, & Wetzel, 2011). CogSketch pays particular attention to both the content and the process of the sketches being developed. Chang and Forbus (2012) extend this work on qualitative sketching to include quantitative analysis of sketching, which allows them to garner a more accurate representation and understanding of the sketches.

Sketching is particularly important given the current focus on conceptual learning in STEM. One of the most popular forms of eliciting student knowledge in science has been the creation of diagrams and concept maps. From this prior work, it is apparent that a number of research groups have demonstrated the ability to do meaningful analyses of sketches in order to study cognition and learning.

2.5 Action and Gesture Analysis

Action recognition has recently received considerable attention within the computer vision community. For example, work by Weinland, Ronfard, & Boyer (2006) and Yilmaz and Shah (2005), among others, has demonstrated the ability to detect basic human actions related to movement. The work of Weinland et al. (2006) involved developing a technique that could capture user actions independent of gender, body size, and viewpoint. The work of Yilmaz and Shah (2005) involved human action recognition using uncalibrated moving cameras, which might prove useful for the dynamic nature of classrooms and/or laboratories.

This kind of work is currently being applied to classroom settings as well. Raca, Tormey, and Dillenbourg (2014), for instance, are pioneering ways of capturing student engagement and attention by conducting frame-by-frame analyses of videos taken from the teacher's position. They show that students' motion and level of attention can be estimated using computer vision, and that individuals with lower levels of attention are slower to react than focused students. This line of work opens the door for new kinds of feedback loops for teachers, by providing not only real-time information about students but also aggregate measures of their levels of attention over time.

Other work in the area of gesture recognition has leveraged infrared cameras and accelerometers that are affixed to the research subject. Using infrared, one avoids some of the complications that may exist with camera geometry, lighting, and other forms of visual variance. Using this approach Schlömer, Poppinga, Henze, and Boll (2008) demonstrate the ability to construct a gesture recognition system by capturing and processing accelerometer data from a Nintendo Wiimote. Their technique allows them to reliably capture gestures for squares, circles, rolling, the shape "Z," etc.

More recent work has taken advantages of the Microsoft Kinect sensor and simple infrared detectors as low cost tools for capturing and studying human gestures. The Mathematical Imagery Trainer (Howison, Trninic, Reinholz, & Abrahamson, 2011) uses hand gestures captured by the Kinect sensor as a way for studying student understanding of proportions. Students use their hands to indicate the relationship

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

between two values, and benefit from visual feedback on the correctness of their hand placement. This system also enables teachers to give students real-time, immediate feedback and change their instruction as they perceive students’ difficulties (and not only after the fact), and points to one of the main benefits of multimodal learning analytics. As an even more basic example, the Kinect sensor can be used to give teachers and students immediate feedback about the amount of gesticulation that they are doing. Hence, without requiring a set of recommended actions, low-cost sensing of movement could be useful in helping students and teachers be more aware of their own behaviours.

Related to the measurement of student gesticulation, early work in this domain by Worsley and Blikstein (2013) involved a comparison of hand/wrist movement between experts and novices as they completed an engineering design task. In particular, the researchers used hand/wrist movement data from a Kinect sensor to examine the extent of two-handed action, and found that experts were much more likely to employ two-handed actions than novices. These preliminary results aligned with theories associated with two-handed inter-hemispheric actions, and provided initial motivation for studying gestures in complex learning environments.

In a similar line of work, Schneider and Blikstein (2014) used a Kinect sensor to evaluate student strategies when interacting with a Tangible User Interface (TUI): their task was to learn about the human hearing system by interacting with 3D-printed organs of the inner ear. Using clustering algorithms, the authors found that students’ body postures fell into three prototypical positions (Figure 1): an active, semi-active, or passive state. The amount of time spent in the active state was significantly correlated with higher learning gains, and the time spent in the passive state was significantly correlated with lower learning gains. More interestingly, the number of transitions between those states was the strongest predictor of learning. Those results suggest that successful students went through cycles of reflection and action, which helped them gain a deeper understanding of the domain taught. This approach shows the potential of using clustering methods on gestures data to find recurring behaviours associated with higher learning gain.

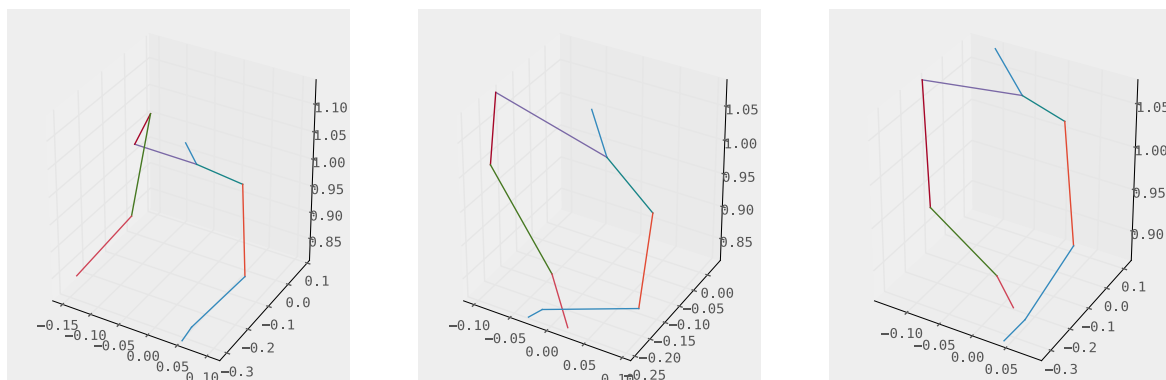


Figure 1: Using k-means on student body posture (Schneider & Blikstein, 2014). The first state (left) is active, with both hands on the table; the second (middle) is passive, with arms crossed; the third (right) is semi-active, with only one hand on the table.

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

As a whole, the advances in action and gesture recognition, and the introduction of low-cost, high-accuracy sensors is creating additional opportunities for action and gesture recognition to be included in education research.

2.6 Affective State Analysis

Studying students' affective states can often be challenging and hard to validate. However, several studies have demonstrated that identifying affect can be done consistently, and that affect is an important marker in studying and understanding learning.

2.6.1 Human Annotated Affective States

Baker, D'Mello, Rodrigo, & Graesser (2010) and Pardos, Baker, San Pedro, Gowda, & Gowda (2013) are examples of work using human annotated affective states. In Pardos et al. (2013), the researchers used the Baker-Rodrigo Observation Method Protocol (BROMP) (Ocumpaugh, Baker, & Rodrigo, 2012) to correlate student behaviour and affect while participating in cognitive tutoring activities with performance on standardized tests. They found that the learning gains associated with certain affective states, namely boredom and confusion, are highly dependent on the level of scaffolding that the student is receiving. This finding builds on prior work that studies affective state as students participate in cognitive tutoring activities (e.g., Litman, Moore, Dzikovska, & Farrow, 2009; Forbes-Riley, Rotaru, & Litman, 2009).

2.6.2 Automatically Annotated Affective State

Other work, using the Facial Action Coding System (FACS), has demonstrated that researchers can recognize student affective state by simply observing their facial expressions. In the case of Craig, D'Mello, Witherspoon, and Graesser (2008), researchers were able to perceive boredom, stress, and confusion by applying machine learning to video data of the student's face throughout the tutoring experience. Data was collected while students interacted with AutoTutor, an intelligent tutoring system for learning science. The technique that Craig et al. (2008) validated is a highly non-invasive mechanism for realizing student sentiment, and can be coupled with computer vision technology to enable machines to detect changes in emotional state or cognitive-affect automatically. Worsley and Blikstein (2015) utilize the Facial Action Coding System to compare two different experimental conditions. More specifically, the authors compared the frequency and rate of transitions among four automatically derived affective states that are conjectured to be important for learning. In particular, they were able to show that the two experimental conditions expressed significantly different rates of confusion and differed in how frequently they transitioned from neutral to surprise, and from neutral to confusion. Being in, or transitioning to a confused expression was generally associated with good outcomes, whereas being more surprised, or transitioning to an expression of surprise was generally associated with less favourable outcomes.

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

Researchers have also used conversational cues to realize students' emotional states. Similar to Craig et al. (2008), D'Mello, Craig, Witherspoon, McDaniel, & Graesser (2008) designed an application that could use spoken dialogue to recognize the states of boredom, frustration, flow, and confusion. Researchers were able to resolve the validity of their findings through comparison to emote-aloud (a derivative of talk-aloud where participants describe their emotions as they feel them) activities while students interacted with AutoTutor.

2.6.3 *Physiological Markers of Affective State*

More recent work in this space was able to accurately predict the affective state, and the source of the change in affective state for users as they interact with a computer-based tutoring system (Conati & MacLaren, 2009). In particular, the system was able to predict when students experienced joy, distress, and admiration effectively. In the past years, other researchers have expanded the detection of affect within educational contexts to leverage physiological markers (Hussain, AlZoubi, Calvo, & D'Mello, 2011; Chang, Nelson, Pant & Mostow, 2013).

Especially when dealing with web-based and tutoring activities, identifying the intensity and the time-occurrence of the emotional state is an important clue to distinguish an affective learning process from a pleasant, but not learning-effective, computer-based activity. Seeking clarity on this distinction, Muldner, Burleson, and VanLehn (2010), used physiological (skin conductance sensor and pupil dilatation), behavioural (speaking aloud protocol, posture in the chair, and mouse clicks) and task-related data to predicted moments of excitement associated to learning, referred to as a "yes!" moment. They found that the "yes!" moment was associated with more reasoning, effort, and investment in solving the task, suggesting that the intensity of this positive emotion after the achievement of a goal may be a predictor of increased learning.

This same physiological approach is also useful to identify negative feelings and reactions, which in turn is associated with lower performance in cognitive tasks. An increase in physiological reactivity was observed by Lunn and Harper (2010) to be associated with a frustrating web-based activity. Moreover, Choi et al. (2010) demonstrated that tense emotions induced by an external stimulus have a negative effect on performance in a subsequent cognitive task.

The various studies of student affect emphasize the potential for empowering educators through student sentiment awareness. Using one, or more, of the modalities of speech, psychophysiological markers, and computer vision, researchers are able to better understand the relationship between affect and learning, and at a much more detailed level.

2.7 **Neurophysiological Markers**

Though briefly mentioned in the previous section, there is a growing cadre of researchers doing work on psychophysiology, and its relationship to cognition and learning. Burt and Obradović (2013) provide an

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

overview of this domain, while also pinpointing key areas for researchers to pay attention to when doing this work. Other researchers, such as Stevens, Galloway, and Berka (2007), describe the IMMEX system used to study the electroencephalograms (EEGs) of students as they participate in a computer-based environment. In their work, they also present preliminary findings on the relationship between EEG and cognitive load, distraction, and engagement. One unexpected finding of the research was that even as a student's skill level increased, the workload remained the same. This unexpected result highlights one of the key affordances of these new multimodal modes of analysis: they may challenge researchers to question previously held assumptions or intuitions about student learning. The study of Stevens et al. (2007) is only one among a host of cutting-edge publications that examine cardiovascular physiology (Cowley, Ravaja, & Heikura, 2013), mid-frontal brain activity (Luft, Nolte, & Bhattacharya, 2013), and other connections between cognition and physiology (Burt & Obradović, 2013).

Moreover, studies vary in the number of sensors used as well as in the types of analyses. Using a single channel portable EEG device, Chang, Nelson, Pant & Mostow (2013) were able to distinguish easy and difficult sentences read by children and adults. In a more complex task, nine EEG channels were used to identify differences from solutions created by students when solving a maze problem that required physics concepts. Students with better solutions (reduced number of leans used) had higher theta power in the frontal areas of the brain, which is related to mental effort, concentration, and attention (She et al., 2012). Neuroimaging techniques increased the comprehension about brain mechanisms involved in learning as well in learning disabilities. Understanding brain mechanisms required for cognitive processing and learning is important to either adapt learning methodologies to specific topics or create interventions for students with specific needs.

2.8 Eye Gaze Analysis

Another area applicable to educational research is eye tracking and gaze analysis. While this technology has long been used within the field of research on consumer electronics and software usage, recent work in a variety of learning environments has shown eye tracking can be useful for understanding student learning. One of the constructs more related to eye gaze is attention. For example, Gomes, Yassine, Worsley, and Blikstein (2013) captured eye-tracking data from high school students as they completed a collection of engineering design games. By using machine learning to cluster the students based on their gaze patterns, the team identified that the highest performing students used very similar patterns in where they looked, how longed they looked, and their level of systematicity.

Data from eye tracking also helps to understand what kind of approaches are useful in helping students to enhance learning. Mason, Pluchino, Tornatora, and Ariasi (2013) demonstrated that using pictures in a scientific text is better than using only text. However, based in the number of fixations in the final part of the text, the authors conclude that using an abstract picture that represents the topic studied (physics phenomena) appears to be more efficient, i.e., same performance but less cognitive load than using a concrete illustration about the same topic,

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

However, de Koning, Tabbers, Rikers, and Paas (2010) argue that looking at specific stimulus can represent the student's shifting of attention to possible areas of interest, but does not always mean that they are learning. In their study, students looked longer and more often at an instructional animation with cues compared to the same animation without cues, but the authors could not confirm that giving cues would reduce the student's cognitive load or even increase conceptual understanding.

Notwithstanding, the most promising use of eye-tracking technology in education has been to study small collaborative learning groups. The overall framework for this type of work is to synchronize two eye-trackers and compute the number of times a particular group achieves joint visual attention (JVA). JVA has been studied extensively in a variety of disciplines (developmental psychology, communication, learning sciences) and is known as a strong predictor of a group's quality of collaboration. Richardson and Dale (2005), for instance, found that the degree of gaze recurrence between individual speaker–listener dyads (i.e., the proportion of alignment of their gazes) was correlated with the listeners' accuracy on comprehension questions. In a remote collaboration, Jermann, Mullins, Nüssli, and Dillenbourg (2011) describe how “good” programmers tend to have a higher recurrence of joint visual attention when having productive interactions, compared to less proficient programmers. Additionally, recent work by Schneider and Pea (2013) suggests that JVA is not just a proxy for predicting collaboration, but can also be influenced to improve communication between students. They designed an intervention in which students worked in pairs (in different rooms). In one condition, the two participants could see each other's gaze; in the other condition, no such augmentation was provided. Their task was to study a set of diagrams to learn about the human visual system. Those who could see the gaze of their partner in real time on the screen achieved significantly higher learning gains and had a higher quality of collaboration. Those findings highlight the potential of using *gaze-awareness* tools for augmenting student interactions in various learning environments and settings. It should be noted that those examples are limited to remote collaborations. Schneider et al., (2015) extends this line of work to co-located settings. Using mobile eye-trackers and computer vision algorithms, they were able to replicate the findings above: in a side-by-side collaboration, JVA was found to be a significant predictor of student learning gains and performance on a problem-solving task.

Finally, Schneider and Pea (2014) are expanding what can be predicted when combining JVA, network analysis and machine learning. In this work, they describe networks where nodes represent visual fixations and edges represent saccades. Their findings suggest that when those networks characterize a dyad (i.e., the size of a node represents the amount of joint visual attention on one particular area of the screen), different properties of the network are associated with different facets of a good collaboration. For instance, the extent to which students reach consensus during a problem-solving task is associated with the average size of the strongly connected components of the graphs. They found that other dimensions of a productive collaboration (sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, task management, technical coordination,

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

reciprocal interaction, individual task orientation) could similarly be predicted by applying machine-learning algorithms on the features of those graphs.

These studies suggest interesting opportunities to understand and enhance collaborative learning using eye-tracking data. More specifically, they provide new ways to study small-group visual coordination and its relationship to productive learning strategies. Recent work is generalizing this line of inquiry across various settings, which opens promising new doors for predicting and influencing collaboration among students.

2.9 Multimodal Integration and Multimodal Interfaces

Having considered several example modalities currently being used by researchers to study student learning individually, we now turn to a final example that entails analysis using multiple modalities. As previously noted, Multimodal Learning Analytics also builds on the idea of multimodal integration and multimodal interfaces. Multimodal integration is the synchronous alignment and combination of data from different modalities (or contexts) in order to get a clearer understanding of the learning cues that students are producing. Worsley (2014) and Worsley and Blikstein (2014) discuss and employ various multimodal learning analytics techniques. Worsley (2014) considers the impact of using different multimodal data fusion approaches. Specifically, the paper highlights naïve fusion, low-level (or data-level) fusion and high-level (or quasi feature-level) fusion as having differing levels of utility, and as being associated with different underlying research questions. Naïve fusion was the label given to multimodal analyses that built machine-learning classifiers from the summary statistic generated from each of the data streams or features. In many cases, these features are first subjected to feature selection in order to reduce the feature space down to something reasonable. Low-level fusion (or feature fusion) involved synchronizing the data at each time step and conducting analyses on the features after they have been fused together. Finally, high-level fusion is described as extracting one or more semantic level features from one or more data streams before fusing them with the other data streams. An example of this would be to do gesture recognition or speech recognition before aligning the hand/wrist movement and/or audio channels with the other data sources available for analysis.

In Worsley and Blikstein (2014), the authors present a multimodal comparison from a two-condition experiment, in which students worked in pairs to complete an engineering design challenge. By using hand/wrist movement, electro-dermal activation, and voice activity detection, the authors were able to identify a set of representative multimodal states that students used, and subsequently used those states to model each student's design approach. Interestingly, students in the two experimental conditions used markedly different approaches. In this way, then, the analysis served to reveal some of the behavioural differences associated with the two different experimental conditions. The analysis also revealed that the multimodal behaviours observed had clear correlations with prior work on epistemological framing (Russ, Lee, & Sherin, 2012).

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

The strategies used in Worsley (2014) and Worsley and Blikstein (2014) represent a small fraction of the work being done in the multimodal analysis community, which spans a variety of complex approaches for doing multimodal fusion at different levels of analysis, as well as using a variety of algorithms, data representations, and strategies for training and testing said algorithms (see Song, Morency, & Davis, 2012; Scherer et al., 2012; and Ngiam et al., 2011 for more details.) A particular challenge, however, is reconciling the complexities of these computational approaches with actionable ideas and theories for learning.

Taken together, the prior research points to a wealth of technology and methodologies that can be used for doing multimodal analysis of student learning across a diversity of environments. By studying learning through these different lenses we can better identify how students are changing, and make more sense of their changes. Furthermore, multimodal analysis enables researchers to get far more nuanced and complex understandings of student learning processes, something that we have only begun to study at scale.

3 CONCLUSION

In this article, we have presented a review of the literature on what we have termed “multimodal learning analytics” — a set of techniques employing multiple sources of data (video, logs, text, artifacts, audio, gestures, biosensors) to examine learning in realistic, ecologically valid, social, mixed-media learning environments.

The incorporation of multimodal techniques, which are extensively used in the multimodal interaction community, should enable researchers to examine unscripted, complex tasks in more holistic ways. In particular, we have focused on describing a set of modalities that have been the topic of multimodal analysis for decades, as well as modalities that have recently emerged as new data streams through which researchers can study human interaction and behaviour.

REFERENCES

- Alvarado, C., & Davis, R. (2006). Dynamically constructed Bayes nets for multi-domain sketch understanding. *Proceedings of the ACM SIGGRAPH 2006 Courses* (A. 33). <http://dx.doi.org/10.1145/1281500.1281544>
- Alvarado, C., Oltmans, M., & Davis, R. (2002). A framework for multi-domain sketch recognition. *Proceedings of AAAI Spring Symposium on Sketch Understanding*, Palo Alto, California (pp. 1–8). Retrieved from <http://rationale.csail.mit.edu/publications/Alvarado2002Framework.pdf>
- Anthony, L., Yang, J., Koedinger, K. (2007). Adapting handwriting recognition for applications in algebra learning. *Proceedings of the ACM Workshop on Educational Multimedia and Multimedia Education (EMME 2007)*, 47–56. <http://dx.doi.org/10.1145/1290144.1290153>

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

- Baker, R., D’Mello, S. K., Rodrigo, M. M. T., Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners’ cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human–Computer Studies*, 68(4), 223–241. <http://dx.doi.org/10.1016/j.ijhcs.2009.12.003>
- Barron, B., & Darling-Hammond, L. (2010). Prospects and challenges for inquiry-based approaches to learning. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The nature of learning: Using research to inspire practice* (pp. 199–225). Paris: OECD.
- Beck, J. E., & Sison, J. (2006). Using knowledge tracing in a noisy environment to measure student reading proficiencies. *International Journal of Artificial Intelligence in Education*, 16(2), 129–143.
- Burt, K. B., & Obradović, J. (2013). The construct of psychophysiological reactivity: Statistical and psychometric issues. *Developmental Review*, 33(1), 29–57. <http://dx.doi.org/10.1016/j.dr.2012.10.002>
- Chang, K. M., Nelson, J., Pant, U., & Mostow, J. (2013). Toward exploiting EEG input in a reading tutor. *International Journal of Artificial Intelligence in Education*, 22(1), 19–38. http://dx.doi.org/10.1007/978-3-642-21869-9_31
- Chang, M., & Forbus, K. (2012). Using quantitative information to improve analogical matching between sketches. In *Proceedings of the 24th Annual Conference on Innovative Applications of Artificial Intelligence (AAAI-12)*, Toronto, Canada, 2269–2274.
- Choi, M. H. C., Su-Jeong Lee, S. J. L., Jae-Woong Yang, J. W. Y., Ji-Hye Kim, J. H. K., Jin-Seung Choi, J. S. C., Jang-Yeon Park, J. Y. P., ... & Dae-Woon Lim, D. W. L. (2010). Changes in cognitive performance due to three types of emotional tension. *International Journal of Bio-Science and Bio-Technology*, 2(4), 23–28. http://dx.doi.org/10.1007/978-3-642-17622-7_26
- Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 19(3), 267–303. <http://dx.doi.org/10.1007/s11257-009-9062-8>
- Cowley, B., Ravaja, N., & Heikura, T. (2013). Cardiovascular physiology predicts learning effects in a serious game activity. *Computers & Education*, 60(1), 299–309. <http://dx.doi.org/10.1016/j.compedu.2012.07.014>
- Craig, S. D., D’Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with AutoTutor: Applying the facial action coding system to cognitive-affective states during learning. *Cognition & Emotion*, 22(5), 777–788. <http://dx.doi.org/10.1080/02699930701516759>
- D’Mello, S., Craig, S., Witherspoon, A., McDaniel, B., & Graesser, A. (2008). Automatic detection of learner’s affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1), 45–80. <http://dx.doi.org/10.1007/s11257-007-9037-6>
- de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction*, 20(2), 111–122. <http://dx.doi.org/10.1016/j.learninstruc.2009.02.010>
- Dewey, J. (1902). *The school and society*. Chicago, IL: University of Chicago Press.

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

- Dym, C. L. (1999). Learning engineering: Design, languages, and experiences. *Journal of Engineering Education*, 88(2), 145–148. <http://dx.doi.org/10.1002/j.2168-9830.1999.tb00425.x>
- Forbes-Riley, K., Rotaru, M., & 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), 11–43. <http://dx.doi.org/10.1007/s11257-007-9038-5>
- Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2011). CogSketch: Sketch understanding for cognitive science research and for education. *Topics in Cognitive Science*, 3(4), 648–666. <http://dx.doi.org/10.1111/j.1756-8765.2011.01149.x>
- Freire, P. (1970). *Pedagogia do Oprimido*. Rio de Janeiro: Paz e Terra.
- Gomes, J. S., Yassine, M., Worsley, M., & Blikstein, P. (2013). Analysing engineering expertise of high school students using eye tracking and multimodal learning analytics. In S. K. D’Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM 2013), (pp. 375–377). Retrieved from http://www.educationaldatamining.org/EDM2013/papers/rn_paper_88.pdf
- Howison, M., Trninic, D., Reinholz, D., & Abrahamson, D. (2011). The Mathematical Imagery Trainer: from embodied interaction to conceptual learning. In G. Fitzpatrick, C. Gutwin, B. Begole, W. A. Kellogg & D. Tan (Eds.), *Proceedings of the annual meeting of CHI: ACM Conference on Human Factors in Computing Systems* (CHI 2011), (Vol. "Full Papers", pp. 1989–1998). <http://dx.doi.org/10.1145/1978942.1979230>
- Hussain, M. S., AlZoubi, O., Calvo, R. A., & D’Mello, S. K. (2011). Affect detection from multichannel physiology during learning sessions with AutoTutor. In G. Biswas, S. Bull, J. Kay, A. Mitrovic (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence in Education* (AIED), (pp. 131–138). http://dx.doi.org/10.1007/978-3-642-21869-9_19
- Jee, B., Gentner, D., Forbus, K., Sageman, B., & Uttal, D. (2009). Drawing on experience: Use of sketching to evaluate knowledge of spatial scientific concepts. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (CogSci 2009), (pp. 2499–2504). Retrieved from http://blog.silccenter.org/publications_pdfs/jee_et_al_CogSci2009_Final.pdf
- Jermann, P., Mullins, D., Nüssli, M.-A., & Dillenbourg, P. (2011). Collaborative gaze footprints: Correlates of interaction quality. In H. Spada, G. Stahl, N. Miyake, N. Law (Eds.), *Proceedings of the 11th International Conference on Computer-Supported Collaborative Learning* (CSCL 2011), (Vol.1, pp. 184–191). Hong Kong: International Society of the Learning Sciences.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75–86. http://dx.doi.org/10.1207/s15326985ep4102_1
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction. *Psychological Science*, 15(10), 661. <http://dx.doi.org/10.1111/j.0956-7976.2004.00737.x>

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

- Levy, F., & Murnane, R. J. (2004). *The new division of labor: How computers are creating the next job market*. Princeton, NJ: Princeton University Press.
- Litman, D., Moore, J., Dzikovska, M., & Farrow, E. (2009). Using natural language processing to analyze tutorial dialogue corpora across domains and modalities. *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling (AIED '09)*, (pp. 149–156). The Netherlands: IOS Press. Retrieved from http://people.cs.pitt.edu/~litman/paper_125.pdf
- Luft, C. D., Nolte, G., & Bhattacharya, J. (2013). High-learners present larger mid-frontal theta power and connectivity in response to incorrect performance feedback. *The Journal of Neuroscience*, 33(5), 2029–2038. <http://dx.doi.org/10.1523/JNEUROSCI.2565-12.2013>
- Lunn, D., & Harper, S. (2010). Using galvanic skin response measures to identify areas of frustration for older web 2.0 users. *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A '10)*, (p. 34). <http://dx.doi.org/10.1145/1805986.1806032>
- Mason, L., Pluchino, P., Tornatora, M. C., & Ariasi, N. (2013). An eye-tracking study of learning from science text with concrete and abstract illustrations. *The Journal of Experimental Education*, 81(3), 356–384. <http://dx.doi.org/10.1080/00220973.2012.727885>
- Montessori, M. (1965). *Spontaneous activity in education*. New York: Schocken Books.
- Muldner, K., Burleson, W., & VanLehn, K. (2010). “Yes!” Using tutor and sensor data to predict moments of delight during instructional activities. In P. de Bra, A. Kobsa, D. Chin (Eds.), *Proceedings of 18th International Conference, UMAP 2010: User modeling, adaptation, and personalization (Lecture Notes in Computer Science Series)*, (Vol. 6075, pp. 159–170). http://dx.doi.org/10.1007/978-3-642-13470-8_16
- Munteanu, C., Penn, G., & Zhu, X. (2009). Improving automatic speech recognition for lectures through transformation-based rules learned from minimal data. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (Vol. 2, Vol. 2)*, (pp.764–772). Stroudsburg, PA: ACM.
- Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, (pp. 689–696). Retrieved from http://machinelearning.wustl.edu/mlpapers/papers/ICML2011Ngiam_399
- Ocuppaugh, J., Baker, R. S. J. d., Rodrigo, M. M. T. (2012). *Baker-Rodrigo Observation Method Protocol (BROMP) 1.0: Training Manual. Version 1.0*. [Technical Report]. New York: EdLab. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York: Basic Books.
- Pardos, Z. A., Baker, R. S. J. d., San Pedro, M. O. C. Z., Gowda, S. M., Gowda, S. M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (LAK '13)*, 117–124. <http://dx.doi.org/10.1145/2460296.2460320>

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

- Raca, M., Tormey, R., & Dillenbourg, P. (2014). Sleepers' lag-study on motion and attention. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge (LAK '14)*, 36–43. <http://dx.doi.org/10.1145/2567574.2567581>
- Read, J. C. (2007). A study of the usability of handwriting recognition for text entry by children. *Interacting with Computers*, 19(1), 57–69. http://dx.doi.org/10.1007/1-84628-062-1_9
- Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science*, 29(6), 1045–1060. http://dx.doi.org/10.1207/s15516709cog0000_29
- Russ, R. S., Lee, V. R., & Sherin, B. L. (2012). Framing in cognitive clinical interviews about intuitive science knowledge: Dynamic student understandings of the discourse interaction. *Science Education*, 96(4), 573–599. <http://dx.doi.org/10.1002/sce.21014>
- Scherer, S., Glodek, M., Layher, G., Schels, M., Schmidt, M., Brosch, T., ... & Palm, G. (2012). A generic framework for the inference of user states in human computer interaction. *Journal on Multimodal User Interfaces*, 6(3–4), 117–141.
- Schick, A., Morlock, D., Amma, C., Schultz, T., & Stiefelwagen, R. (2012). Vision-based handwriting recognition for unrestricted text input in mid-air. *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)*, 217–220. New York: ACM. <http://dx.doi.org/10.1145/2388676.2388719>
- Schlömer, T., Poppinga, B., Henze, N., & Boll, S. (2008). Gesture recognition with a Wii controller. *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction (TEI '08)*, 11–14. <http://dx.doi.org/10.1145/1347390.1347395>
- Schneider, B., & Blikstein, P. (2014). Unraveling Students' Interaction Around a Tangible Interface using Gesture Recognition. In J. Stamper, Z. Pardos, M. Mavrikis, B. Mclaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining (EDM'14)*, (pp.320-323). Retrieved from <http://educationaldatamining.org/EDM2014/proceedings/EDM2014Proceedings.pdf>
- Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, 8(4), 375–397. <http://dx.doi.org/10.1007/s11412-013-9181-4>
- Schneider, B., & Pea, R. (2014). Toward Collaboration Sensing. *International Journal of Computer-Supported Collaborative learning*, 9(4), 371-395. <http://dx.doi.org/10.1007/s11412-014-9202-y>
- Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, A. D. (2015). 3D tangibles facilitate joint visual attention in dyads. In O. Lindwall, P. Hakkinen, T. Koschmann, P. Tschounikine, S. Ludvigsen (Eds.), In *Proceedings of the International Conference on Computer Supported Collaborative Learning 2015: Exploring the Material Conditions of Learning (CSCL'15)*, (Vol.1, pp. 158–165). Gothenburg, Sweden: The International Society of the Learning Sciences.
- She, H. C., Jung, T. P., Chou, W. C., Huang, L. Y., Wang, C. Y., & Lin, G. Y. (2012). EEG dynamics reflect the distinct cognitive process of optic problem solving. *PLOS ONE*, 7(7), e40731. <http://dx.doi.org/10.1371/journal.pone.0040731>

(2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <http://dx.doi.org/10.18608/jla.2016.32.11>

- Sherin, B. (2013). A computational study of commonsense science: An exploration in the automated analysis of clinical interview data. *Journal of the Learning Sciences*, 22(4), 600–638. <http://dx.doi.org/10.1080/10508406.2013.836654>
- Song, Y., Morency, L. P., & Davis, R. (2012). Multimodal human behavior analysis: Learning correlation and interaction across modalities. *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)*, 27–30. <http://dx.doi.org/10.1145/2388676.2388684>
- Stevens, R. H., Galloway, T., & Berka, C. (2007). EEG-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills. In C. Conati, K. McCoy, G. Paliouras (Eds.), *Proceedings of the 11th International Conference on User Modeling*, (pp. 187–196). Berlin Heidelberg: Springer Link. http://dx.doi.org/10.1007/978-3-540-73078-1_22
- Weinland, D., Ronfard, R., & Boyer, E. (2006). Free viewpoint action recognition using motion history volumes. *Computer Vision and Image Understanding*, 104(2), 249–257. <http://dx.doi.org/10.1016/j.cviu.2006.07.013>
- Worsley, M. (2014). Multimodal learning analytics as a tool for bridging learning theory and complex learning behaviors. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge (MLA '14)*, 1–4. <http://dx.doi.org/10.1145/2666633.2666634>
- Worsley, M., & Blikstein, P. (2011). What's an expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th Annual Conference on Educational Data Mining (EDM 2011)*, (pp. 235–240).
- Worsley, M., & Blikstein, P. (2013). Toward the development of multimodal action based assessment. *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. (2014). Deciphering the practices and affordances of different reasoning strategies through multimodal learning analytics. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge (MLA '14)*, 21–27. <http://dx.doi.org/10.1145/2666633.2666637>
- Worsley, M., & Blikstein, P. (2015). Using learning analytics to study cognitive disequilibrium in a complex learning environment. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge (LAK '15)*, 426–427. <http://dx.doi.org/10.1145/2723576.2723659>
- Yilmaz, A., & Shah, M. (2005). Actions sketch: A novel action representation. In C. Schmid, S. Soatto, C. Tomasi (Eds.), *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, (Vol. 1, pp. 984–989). Los Alamitos, CA: IEEE Computer Society. <http://dx.doi.org/10.1109/CVPR.2005.58>
- Zaiane, O. R. (2001). Web usage mining for a better web-based learning environment. In *Proceedings of the Conference on Computers and Advanced Technology in Education (CATE 2001)*, (pp. 60–64). <https://webdocs.cs.ualberta.ca/~zaiane/postscript/CATE2001.pdf>