

Chapter 11

Multimodal Behavioral Analytics in Intelligent Learning and Assessment Systems

Saad M. Khan

Abstract As the boundary blurs between the real and the virtual in today's learning environments, there is a growing need for new assessment tools that capture behavioral aspects key to evaluating skills such as problem solving, communication, and collaboration. A key challenge is to capture and understand student behavior at fidelity sufficient to estimate cognitive and affective states as they manifest through multiple media, including speech, body pose, gestures and gaze. However, analyzing each of these modalities in isolation may result in incongruities. In addition, the affective states of a person show significant variations in time. To address these technical challenges, this paper presents a framework for developing hierarchical computational models that provide a systematic approach for extracting meaningful evidence from noisy, unstructured data. This approach utilizes multimodal data, including audio, video, and activity log files and models the temporal dynamics of student behavior patterns. To demonstrate the efficacy of our methodology, we present two pilot studies from the domains of collaborative learning and in vivo assessments of nonverbal behavior where this approach has been successfully implemented.

Keywords Machine learning · Multimodal fusion · Hierarchical processing models

11.1 Introduction

To be successful in today's rapidly evolving, technology-mediated world, students must not only possess strong skills in areas such as reading, math, and science, but they must also be adept at 21st-century skills such as critical thinking, communication, problem solving, persistence, and collaboration (Farrington et al., 2012). These skills have been demonstrated to improve learning outcomes and are being

S.M. Khan (✉)
Educational Testing Service, Princeton, NJ, USA
e-mail: skhan002@ets.org

rapidly incorporated in a number of high-stakes standardized assessment systems (Smarter Balanced Assessment Consortium, n.d.). However, the assessment of skills such as collaboration and communication is difficult because often it involves understanding the process used to arrive at a conclusion rather than simply the end product (Bejar, 1984; Romero & Ventura, 2007). Analyzing these processes requires tracking not only the cognitive processes but also noncognitive behaviors, for example, motivation, self-control, and emotional and affective states that influence interpersonal interactions. In addition, much of the infrastructure of assessment design has come of age around traditional multiple-choice tests and self-reports. In contrast, educational simulations and games provide opportunities to expose students to authentic educational tasks and allow them to interact with and explore complex representations of serious academic content (Fisch, 2005; National Research Council, 2011). They do so in a manner that is amenable to capturing rich process data *in vivo*, that is, during the execution of a task involving collaboration, problem solving, and other complex tasks. These data can be multimodal, that is, they can include multiple sensory modalities such as audio, video, and 3D (using depth-sensing devices such as Microsoft Kinect), in addition to traditional forms of computer interaction data such as mouse click streams and keystrokes. The key advantage of using such multimodal data is that it enables high fidelity sensing and tracking of a user's cognitive and noncognitive states, which would otherwise be missed in traditional log files. However, extracting relevant features from these data that can be used as evidence to infer competency in complex constructs such as collaboration is a significant technical challenge for a number of reasons. First, the raw time series of multimodal data often does not have any direct semantic meaning and may not be interpretable by humans as such. As mentioned earlier, it may constitute simulation log files, audio, and visual data, which, without sophisticated computational models, cannot be analyzed for meaningful information. Second, building pattern recognition approaches to detect and recognize sequences and combinations in raw data requires "training data" that may not be readily available. And finally, the inferences and corresponding interpretations from raw multimodal data may contain information at vastly different levels of semantic meaning and abstraction that may not be easily combined in a scoring model, for example, specific facial expressions versus turn taking or user's level of engagement.

This paper is intended to provide a framework and methodology to design and develop computational models that enable analysis of noisy, unstructured, multimodal data for the assessment of complex constructs such as collaboration and communication. Specifically, this paper describes a hierarchical data processing and inference methodology that can help bridge the gap between the raw, low-level multimodal data and the measurement of high-level constructs. To illustrate the efficacy of such a methodology, two example pilot studies are presented where such an approach was implemented to study collaborative learning and *in vivo* measurement of nonverbal behavior using wearable sensors.

11.2 Multimodal Analytics for Studying Student Behavior

Simulations and games in computerized educational environments offer an exciting new paradigm to assess knowledge, skills, and abilities that are difficult to capture with traditional measurement tools such as structured tests and multiple choice items. Such computerized educational environments enable powerful audiovisual interfaces that can be utilized to analyze student's actions, behaviors, and indeed their process in solving the problem, rather than just their final products. Of particular interest are moment-by-moment student affective and cognitive states and how these are related with task performance and learning outcomes in general (D'Mello & Graesser, 2012; Whitehill, Serpell, Lin, Foster, & Movellan, 2014).

A key advantage of using computerized educational environments is that they can enable gathering of rich multimodal data in the form of video streams, audio streams, and simulation log files. These data can be processed and analyzed using multimodal analytics to study performance at individual and group levels. The term *multimodal analytics* (Amer, Siddiquie, Khan, Divakaran, & Sawhney, 2014; Morency, de Kok, & Gratch, 2010; Siddiquie, Khan, Divakaran, & Swahney, 2013) refers to the use of advanced sensor technologies and machine learning systems to track and understand human behaviors. It promises a paradigm shift in learning and assessment that can afford rich, automated, and grounded inferences about human performance from large amounts of multiple sensory data, for example, audio and video. However, developing computational models that can extract meaningful features indicative of performance and skills from the raw, low-level multimodal data is a significant technical challenge. In contrast, when human observers rate task performance, they are quite naturally integrating information from both what they have seen (visual) and what they have heard (auditory). Moreover, the observers' brains translate the visual data into information about body postures, facial expressions, and actions taken. The auditory data are translated into meaningful communication, multiperson verbal exchanges, and tone-of-voice cues. These features are further combined to allow the observers to make judgments about the emotional states, social skills, and technical competencies of the individual performers.

11.2.1 Hierarchical Inference Framework

To address the challenges outlined above, our approach is to build a hierarchical processing and inference framework. As illustrated in Fig. 11.1, raw multimodal data form the first layer of the framework. Data are captured using a multitude of sensors, including audio, video, 3D, and even simulation log files. These data are preprocessed to extract machine features, for example, histogram of oriented gradients (HOG) from visual data, Euler angles from 3D skeleton data, and Mel-frequency cepstral coefficient (MFCC) features from audio data, among others.

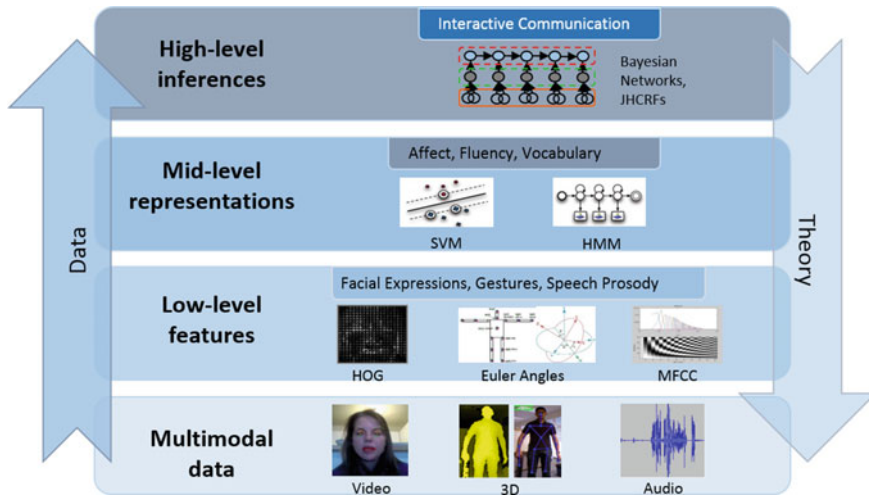


Fig. 11.1 Our framework to bridge the gap between low-level digital data and the measurement of complex constructs. *HMM* hidden Markov model, *HOG* histogram of oriented gradients, *JHCRF* joint hidden conditional random fields, *MFCC* mel-frequency cepstral coefficient, *SVM* support vector machine

We call such machine features *low-level features*, and they reside in the second layer of the hierarchical framework. The output of this layer is descriptive features that may have semantic meanings, such as facial expressions, gestures, or speech prosody. Such descriptive features termed *mid-level representations* reside the next level up in the hierarchical framework.

In this layer, the temporal dynamics of low-level features and mid-level representations are modeled to generate holistic measures of human behavioral states such as affect, engagement, and flow. At the top level of the hierarchy reside the features that make up a theoretical model representing the construct of interest such as communication competency or collaborative skill; these features are called *high-level interpretations*. This layer takes as input assessment of mid-level behavioral features and employs psychometric models to make inference about the competency of interest.

11.2.2 Using Multimodal Analytics to Study Influence of Affect and Noncognitive Behavior on Collaborative Study

Various studies have demonstrated the impact and influence of student affective state and behaviors such as turn taking (Woolley, Chabris, Pentland, Hashmi, & Malone, 2010), entrainment (convergence), and mirroring of affect (Lakin, Jefferis,

Cheng, & Chartrand, 2003) on higher group intelligence and learning outcomes. Here a pilot study is presented that utilizes multimodal analytics to understand the incidence, dynamics, and influence of affect in collaborative problem solving (Luna Bazaldua et al., 2015). Our hypothesis is that performance on collaborative tasks is closely related to participant affective states and behaviors. Therefore, information about such states and behaviors can be important evidence for assessing the overall success of collaboration and individual ability to collaborate, as well as how well different tasks encourage collaboration. To test this approach a study was conducted involving 12 unique dyads collaborating in an online game-like science assessment, Tetralogue (Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2016; Zapata-Rivera et al., 2014). This platform includes both traditional assessment components, such as a set of multiple-choice items on general science topics, a simulation-based assessment, a personality test, and a set of background questionnaires. The simulation task relates to geology topics. The simulation-based task was developed as a task for individual test takers who will interact with two avatars, and as a collaborative task that requires collaboration among two human participants and two avatars in order to solve geology problems. The participants, who may be in different locations, interact through an online chat box and system help requests (i.e., opting to view educational videos on the subject matter).

Figure 11.2 illustrates the use of the Tetralogue collaborative activity platform and data capture system. Multimodal data, including video and activity log files, of each participating dyad were captured. The log files contain behavior that included frequency and content of chat messages between dyads, response to science questions both individually and as a dyad (jointly as a group), and system help requests (i.e., the participant asks to view educational videos on the subject matter to better answer assessment questions). The video data, on the other hand, recorded participant nonverbal behavior, which was analyzed on a frame-by-frame basis using automated facial expression classifiers and annotated by trained human raters on high-level noncognitive behaviors, including: affect display gestures, engagement, anxiety, and curiosity. The data were analyzed at individual and dyad levels and results derived using hierarchical clustering analysis demonstrated statistically significant evidence of cognitive and noncognitive behavioral convergence among dyads (see Sect. 3.2 for details).

11.2.3 Multimodal Data and Low-Level Features

Facial expression analysis of the video data was performed using the FACET SDK, a commercial version of the Computer Expression Recognition Toolbox (CERT; Littlewort et al., 2011). This tool recognizes fine-grained facial features, or facial action units (AUs), described in the Facial Action Coding System (Ekman, Friesen, & Hager, 1977). FACET detects human faces in a video frame, locates and tracks facial features, and uses support vector machine-based classifiers to output

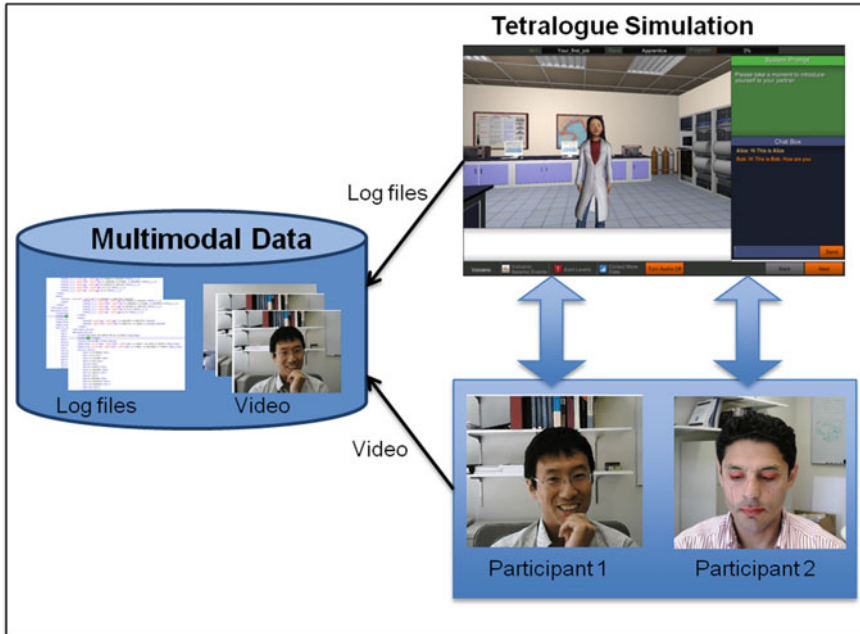


Fig. 11.2 Multimodal data capture including video and action log files while participants engage in collaborative activity on the Tetralogue platform

frame-by-frame detection probabilities of a set of facial expressions: anger, joy, contempt, and surprise.

In addition, seven trained coders reviewed and coded the videos using Anvil software (Kipp, 2001). The video data for each participant were assigned to two raters for annotation; however, in three cases there were three raters coding the same video file, and in two cases only a single rater was available for annotation. The raters followed the same coding scheme during the annotation process, that is, coding data with the labels: *hand touching face*, *expressing engagement*, *expressing anxiety*, or *expressing curiosity*. The behaviors were coded on a binary scale, reflecting whether the behaviors were absent or present. As an outcome of the annotation process, the Anvil software produced extensible markup language (XML) files that were parsed using the XML package (Lang, 2013) in R.

Engagement, anxiety, and curiosity were included in the annotation scheme because of the incidence and relevance with which these three noncognitive states occur in simulation games and online learning systems (Baker, D’Mello, Rodrigo, & Graesser, 2010; Calvo & D’Mello, 2010; D’Mello & Graesser, 2012; Woolf et al., 2009). The coding also included *hand touching face*, an affect display gesture that has been linked to affective and cognitive states such as boredom, engagement, and thinking (Mahmoud & Robinson, 2011; Whitehill et al., 2014).

11.2.4 Mid-level Features and Construct Inference

In order to study evidence of behavioral convergence, features from log files and video data of each of the 24 study participants were represented as a multidimensional behavioral feature vector composed of both the cognitive behaviors: *number_of_messages* and *number_of_help_requests* and the noncognitive behaviors (i.e., fraction of the time each participant exhibited the behavior): *engagement*, *hand_on_face*, *anxiety*, *curiosity*, *anger*, *joy*, *contempt*, and *surprise*.

An unsupervised, agglomerative hierarchical cluster analysis using an average linkage function was performed on a Euclidean distance matrix (i.e., a similarity matrix) computed from the multidimensional behavioral feature data of the study participants. Our hypothesis is that behavioral convergence will manifest in the cognitive and noncognitive features such that members of the same dyad will tend to group together from the beginning of the clustering process, that is, they will be closer to each other in the feature space than to others.

The similarity matrix of behavioral feature distances for participants within and outside dyad clusters was analyzed. Behavioral convergence would imply that, for dyad members, the average distances in feature space is smaller in a statistically significant manner than those of nominal dyad members. Moreover, to study the relative impact of cognitive and noncognitive features, two additional similarity matrices were computed: one using exclusively the cognitive features from log files (number of chat messages and number of system help requests) and the other using exclusively noncognitive features produced from the video data (the four facial expression detectors, and the four features from the coding scheme). All features were normalized to present equivalent scaled values between 0 and 1.

Table 11.1 shows the means and standard deviations of feature similarity distances of participants in dyad and nominal dyad populations. The results consistently show smaller average distances for the dyads (i.e., members within dyads displayed behavior that was more similar to each other than to others), supporting the convergence premise. Additionally, Student's *t* test was used to evaluate the statistical significance of these results. The results show that, when using both cognitive and noncognitive features together, the feature distance between participants belonging to the same dyad was smaller than the corresponding distance

Table 11.1 Average and standard deviation of behavioral feature distances within and outside dyads

Features	Populations	Mean	SD
Cognitive and noncognitive	Dyads	0.572	0.228
	Nominal dyads	0.730	0.243
Cognitive only	Dyads	0.365	0.216
	Nominal dyads	0.571	0.209
Noncognitive only	Dyads	0.411	0.178
	Nominal dyads	0.414	0.225

between nominal-dyads in a statistically significant manner: $t = 2.335$, $df = 11.7$, $p < 0.02$. However, when using noncognitive features alone, a statistically significant pattern of behavioral convergence was not found.

11.3 In Vivo Assessments of Nonverbal Behavior Using Multimodal Wearable Sensors

Human behavior modeling has been studied in a variety of disciplines such as behavioral science, social science, cognitive science, and artificial intelligence, among others. Several researchers have developed models of human behavior, from cognitive and affective states to human activities. This research has also explored the impact and influence of individual personality traits on outcomes of collective group activity. The traditional approach has been to create personality profiles using tools such as Big Five (Tosi, Mero, & Rizzo, 2000) or FACETS (Kyllonen, Lipnevich, Burrus, & Roberts, 2014) and analyzing the outcomes of the group interaction task vis-à-vis individual personality traits. Typically, this entails participants completing pretask or posttask questionnaires, an activity that is time intensive, expensive, and may induce subjective and social biases. Moreover, moment-by-moment activities and interactions in the group task are not captured, and the data are usually too sparse and coarse for an exploratory behavioral analysis. In contrast to this, some exciting new research has focused on measuring and modeling interpersonal behavior using low-level nonverbal behavioral data from environmental and wearable sensors (Olguin & Pentland, 2010). Of particular interest is research on assessing interpersonal skills in tasks such as negotiations, collaboration, leaderless tasks, and so forth, by tapping into a nonverbal, subconscious channel of human communication that Pentland calls honest signals (Pentland, 2008).

Multimodal analytics to conduct such assessments during in vivo group exercises (real world, in person) in a noninvasive manner using wearable sensors. These honest signals influence the outcome of group tasks, and therefore information about such states and behaviors can be important evidence for assessing the overall success of collaboration, individual ability, and interpersonal skills, as well as an alternative way to measure personality traits in and of themselves.

11.3.1 Analyzing Nonverbal Behavior

The subtle, unconscious patterns in which humans interact reveal their attitudes toward each other. These honest signals, as characterized by Pentland, are composed of patterns in physical activity, speech activity, and proximity, among other low-level behavioral cues. This research (Pentland, 2008; Woolley et al., 2010) has

delineated a number of noncognitive, nonverbal behaviors that influence interpersonal interactions and will be the focus of this study. In particular we are interested in the following:

- **Mimicry:** The extent to which people in a conversation are reflexively mirroring each other.
- **Conversational turn-taking:** Participation balance and dominance.
- **Activity:** Measured as body movement or speech energy; increased activity often indicates interest and excitement.

One of the first attempts to measure face-to-face interactions between people using wearable sensors was the sociometer (Choudhury & Pentland, 2003). This wearable sensor package was used to learn social interactions from sensory data and model the structure and dynamics of social networks. Pentland described several statistical learning methods that use wearable sensor data to make reliable estimates of users' interactions. He presented a detailed description of behavior modeling for learning and classifying user behavior from proximity and location data, and influence modeling for predicting the behavior of a subject from another subject's data.

In an ongoing pilot study conducted at ETS, wearable sensors, specifically the Sociometric Badge (Olguin & Pentland, 2010) are utilized to measure nonverbal behavior in human interactions. The Sociometric Badge is a wearable sensing device that can be used to study human behavior and social interactions. Specifically, the badge collects information on (a) speech features such as volume, tone of voice, and speaking time; (b) body movement features such as energy and consistency; (c) information regarding people nearby wearing a Sociometric Badge; (d) the proximity of Bluetooth-enabled devices; and (e) approximate location information. The badges will not record speech or conversational content (unless this option is manually enabled). Figure 11.3 shows an image of the wearable badge. The study consisted of 24 participants that were divided into groups of four to work on a decision-making task. Each group member was given a role (e.g., Vice President [VP] of Finance, VP of Operations) of a hypothetical company, and the groups were tasked with choosing a store location that would be best for their company as it moves into a foreign market. Each member was given positive, negative, and neutral information about each of three potential store locations. Participants wore Sociometric Badges that recorded features from speech and body motion. Figure 11.4 illustrates preliminary analysis that easily shows evidence of turn taking and dominance of social interactivity. In the top left image, each participant is represented by a node (colored circle) in a connected graph. The thickness of the edges connecting any pair of nodes represents the number of speaking turns between the participant pair. It can be clearly seen that the individuals represented by the red, blue, and orange nodes had more turns between themselves than with the person represented by the green node. A similar picture emerges in the top right image, which shows a pie chart of individual speaking time, and the bottom image, which shows a timeline of speech onsets and offset from each of the four participants.

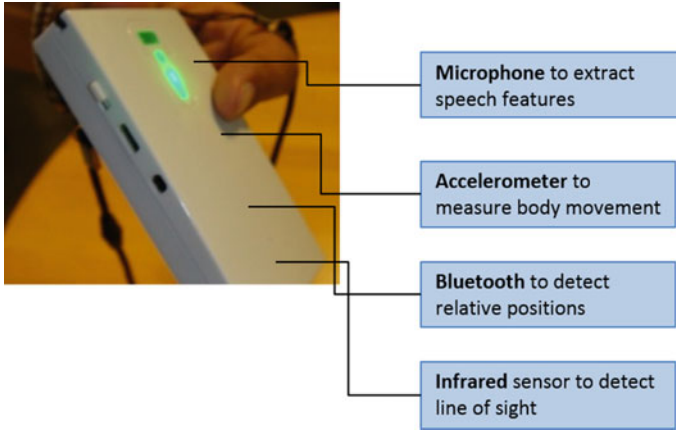


Fig. 11.3 The wearable sociometric badge

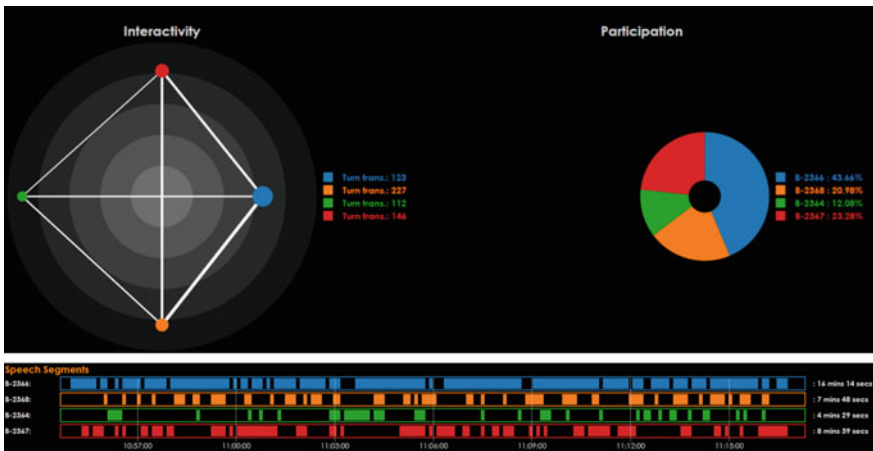


Fig. 11.4 Speech frequency and segments of a four-person group measured with sociometric badges

11.4 Concluding Remarks

This paper presents a framework to design and develop computational models that enable analysis of noisy, unstructured, multimodal data for the capture, analysis, and measurement of complex human behavior. This approach utilizes multimodal data including audio, video, and activity log files and constructs a hierarchical analysis methodology to model temporal dynamics of human behavior and the integration of multiple data modalities. The efficacy of such a methodology is

demonstrated with two pilot studies where this approach was implemented to study collaborative learning and in vivo measurement of nonverbal behavior using wearable sensors.

Acknowledgements This research has benefited from help in data collection efforts, technical insight and valuable feedback the author received from colleagues including Diego Luna Bazaldua, Alina von Davier, Jiangang Hao, Robert Mislevy and Ketly Jean Pierre.

References

- Amer, M. R., Siddiquie, B., Khan, S., Divakaran, A., & Sawhney, H. (2014). Multimodal fusion using dynamic hybrid models. In Institute of Electrical and Electronics Engineers (Ed.), *2014 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 556–563). Los Alamitos, CA: IEEE.
- Baker, R. S., 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.
- Bejar, I. I. (1984). Educational diagnostic assessment. *Journal of Educational Measurement*, *21*(2), 175–189.
- Calvo, R. A., & D’Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, *1*(1), 18–37.
- Choudhury, T., & Pentland, A. (2003). Sensing and modeling human networks using the sociometer. In Institute of Electrical and Electronics Engineers (Ed.), *Proceedings of the 7th IEEE International Symposium on Wearable Computers* (pp. 216–222). Los Alamitos, CA: IEEE.
- D’Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157.
- Ekman, P., & Friesen, W. V. (1977). *Facial action coding system*.
- Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., et al. (2012). *Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago, IL: Consortium on Chicago School Research.
- Fisch, S. M. (2005). Making educational computer games “educational.” In Association for Computing Machinery (Ed.), *Proceedings of the 4th International Conference for Interaction Design and Children* (pp. 56–61). New York, NY: Association for Computing Machinery.
- Kipp, M. (2001) Anvil—A generic annotation tool for multimodal dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)* (pp. 1367–1370).
- Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill based, and affective learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*, 311–328.
- Kyllonen, P. C., Lipnevich, A. A., Burrus, J., & Roberts, R. D. (2014). *Personality, motivation, and college readiness: A prospectus for assessment and development* (Research Report No. RR-14-06). Princeton, NJ: Educational Testing Service.
- Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, *27*(3), 145–162.
- Lang, D. T. (2013). *XML: Tools for parsing and generating XML within R and S-Plus* (R package version 3.98–1.1). Retrieved from <http://CRAN.R-project.org/package=XML>

- Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J. R., et al. (2011). The computer expression recognition toolbox (CERT). In Institute of Electrical and Electronics Engineers (Ed.), *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 298–305). Los Alamitos, CA: IEEE.
- Liu, L., Hao, J., von Davier, A., Kyllonen, P., & Zapata-Rivera, D. (2016). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on computational tools for real-world skill development*. Hershey, PA: IGI-Global.
- Luna Bazaldua, D. A., Hao, J., Khan, S., Liu, L., von Davier, A. A., & Wang, Z. (2015). On convergence of cognitive and non-cognitive behavior in collaborative activity. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, ... M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining Conference* (pp. 496–499). Madrid, Spain: International Educational Data Mining Society.
- Mahmoud, M., & Robinson, P. (2011). Interpreting hand-over-face gestures. In S. D’Mello, A. Graesser, B. Schuller, & J-C Martin (Eds.), *Proceedings of the International Conference on Affective Computing and Intelligent Interaction* (pp. 248–255). New York, NY: Springer.
- Morency, L. P., de Kok, I., & Gratch, J. (2010). A probabilistic multimodal approach for predicting listener backchannels. *Autonomous Agents and Multi-Agent Systems*, 20(1), 70–84.
- National Research Council. (2011). *Assessing 21st century skills*. Washington, DC: National Academies Press.
- Olguin, D. O., & Pentland, A. (2010). Assessing group performance from collective behavior. In Association for Computing Machinery (Ed.), *CSCW 2010 Workshop on Collective Intelligence in Organizations*. New York, NY: Association for Computing Machinery.
- Pentland, A. (2008). *Honest signals: How they shape our world*. Cambridge, MA: MIT Press.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 35, 135–146.
- Siddiquie, B., Khan, S., Divakaran, A., & Sawhney, H. (2013, July). Affect Analysis in natural human interaction using joint hidden conditional random fields. In Institute of Electrical and Electronics Engineers (Ed.), *2013 IEEE International Conference on Multimedia and Expo (ICME 2013)* (pp. 1–6). Los Alamitos, CA: IEEE.
- Smarter Balanced Assessment Consortium. (n.d.). *Thermometer crickets: Grade 11 performance task*. Retrieved from <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/09/performance-tasks/crickets.pdf>
- Tosi, H. L., Mero, N. P., & Rizzo, J. R. (2000). *Managing organizational behavior* (4th ed.). Cambridge, MA: Blackwell Publishers.
- Whitehill, J., Serpell, Z., Lin, Y. C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, 5(1), 86–98.
- Woolf, B., Bursleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: recognizing and responding to student affect. *International Journal of Learning Technology*, 4(3–4), 129–164.
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686–688.
- Zapata-Rivera, D., Jackson, T., Liu, L., Bertling, M., Vezzu, M., & Katz, I. R. (2014). *Assessing science inquiry skills using dialogues*. *Lecture Notes in Computer Science* (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 8474, pp. 625–626). Berlin, Germany: Springer.