# Towards Detection of Engagement and Affect in a Simulation-based Combat Medic Training Environment

Ryan S. Baker, Jeanine A. DeFalco, Jaclyn Ocumpaugh, Luc Paquette

Teachers College, Columbia University, New York, NY

baker2@exchange.tc.columbia.edu, jad2234@tc.columbia.edu, jo2424@tc.columbia.edu, paquette@tc.columbia.edu

**Abstract.** Recently, there has been increasing research in intelligent tutoring systems on improving learning by automatically detecting affect and providing affective support. Systems have been developed both to detect affect from physiological sensors, and from students' interactions with an online learning system. In this paper, we discuss ongoing efforts to construct sensor-free detectors of trainee affect and behavioral engagement for vMedic, an immersive software used to train military trainees in combat medic field procedures. This paper reviews rationale and plans to develop models of affect and engagement based on log files and field observations from a September 2013 study at West Point conducted jointly by the Army Research Lab, Teachers College, and North Carolina State University. Additionally, this paper discusses the preparation for the distillation process, plans for development of the affect detector model, and the implications of future integration into the GIFT Framework's Pedagogical Module.

**Keywords:** GIFT, Educational data mining, affect detection, affective computing, ITS

# 1 Introduction

In recent years, developers of ITSs have turned their attention toward understanding student affect and engagement. It is widely known that the interaction of affect and engagement shapes learning in complex ways (Baker, D'Mello, Rodrigo, & Grasser, 2010; Baker et al., 2011; D'Mello, Taylor & Grasser, 2007; Dragon et al., 2008; Lee et al, 2011; Sabourin, Rowe, Mott, & Lester, 2011). Take boredom, an affective state replicably shown to be associated with poorer learning and educational outcomes (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013). For instance, boredom can lead either to gaming the system (Baker, D'Mello, Rodrigo, & Grasser, 2010) and off-task behavior (Baker et al., 2011). When boredom leads to gaming, boredom persists, creating a vicious cycle (Baker, D'Mello, Rodrigo, & Grasser, 2010). By contrast, off-task behavior relieves boredom (Baker et al., 2011).

Since gaming the system is much more strongly correlated with learning and educational outcomes (negatively) than off-task behavior (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; Cocea, Hershkovitz & Baker, 2009) the different student responses to boredom matter. So, too, the affective states of confusion and frustration can be associated with positive or negative learning outcomes, depending on the length of their duration (Lee et al., 2011; Liu, Pataranutaporn, Ocumpaugh, & Baker, 2013). Affect also influences learning through its effects on memory, attention, and strategy use (Pekrun, 1992; Schunk & Zimmerman, 2007). Learners who experience positive emotions are better able to retrieve information connected with such feelings (Forgas, 2000). Research has also shown that negative states may trigger greater cognitive load, which reduces working memory (Linnebrink & Pintrich, 2000). While these processes are not thoroughly understood, it is likely that affective states such as frustration and anxiety can draw cognitive resources away from the task at hand to focus on the source of the emotion (Zeidner, 1998). Because of the range of impacts that affect can have on learning, learning systems that explicitly consider affect have been able to positively impact students' engagement and learning (Baker, D'Mello, Rodrigo, & Graesser, 2010).

Researchers have relied on a variety of strategies when constructing models of engagement and affect. One approach that has been popular and successful is to utilize physical sensors such as video cameras that capture facial expressions, posture sensors that detect when a student is shifting positions, and galvanic skin response sensors that detect sweating (Conati, 2002; Mohammad & Nishida, 2010; Alzoubi, Calvo, & Stevens, 2009). The use of sensors has yielded successful detectors of a range of affective states (Alzoubi, Calvo, & Stevens, 2009), but it is not always feasible to implement the resulting detectors into the field. As such, other researchers have turned their analysis towards the fine-grained interaction logs produced by online learning systems, developing sensor-free detectors of affect and engagement.

Previous research on sensor-free detection has shown that this modeling strategy can be successful and affords analysis of learner actions at multiple levels. Successful results have been obtained for a variety of systems, including straightforward problem-solving intelligent tutoring systems (e.g. Baker et al., 2012; Rodrigo & Baker, 2009), dialogue tutors (e.g. Litman & Forbes-Riley, 2006), science simulations (e.g. Paquette et al., in press), and narrative-based virtual environments (e.g. Baker, Ocumpaugh, Gowda, Kamarainen, & Metcalf, in press). Across these systems, models have been developed for affective states such as boredom, frustration, confusion, and engaged concentration, and for disengaged behaviors such as gaming the system and off-task behavior. Although the resulting models are often quite complex, detectors constructed solely from interaction log data offer a significant advantage: they can be used in settings where physical sensors are unavailable, enabling greater scale.

So far, one of the important lessons of developing sensor-free models of affect and engagement for a range of different environments is that the types of behaviors associated with the same affect may differ substantially between systems. For example, affective models developed for ASSISTments (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013), an ITS that provides scaffolded instruction for middle-school mathematics problems, demonstrate that the timing of learner actions is almost as important as the actions themselves. On the other hand, models constructed for EcoMUVE (Baker, Ocumpaugh, Gowda, Kamarainen, & Metcalf, in press), an immersive virtual environment that teaches environmental science, relied more heavily on the type of action the student was making to predict affect, such as whether the student was repeating specific kinds of measurement, and what information pages the student was accessing in the virtual data manual (Baker, Ocumpaugh, Gowda, Kamarainen, & Metcalf, in press).

Understanding the different behaviors commonly associated with affect and engagement in different systems is an important step towards the development of general frameworks for affect/engagement detection that facilitate the development of affect and engagement detectors for new learning environments.

In this paper, we discuss plans to build upon this body of previous research in the context of vMedic, an immersive ITS integrated with the Generalized Intelligent Framework for Tutoring (GIFT) (Sottilare, Goldberg, & Holden, 2012) provided through the GIFT framework's Trainee Module. We provide an overview of the methods that will be used to construct these detectors, developed through a combination of quantitative field observation (QFO) methods and educational data mining (EDM) techniques. These detectors will serve as an example of how to integrate this type of model into GIFT, will offer insight into the sorts of behavior that are correlated with learner engagement, will be useful for automated interventions, and will serve as a springboard for further research on learning and engagement within the vMedic software.

## 2 Data and Methods

Two sources of data will be used for this study: log file data produced by learners using the vMedic software and quantitative field observations of those learners while using the system. This section describes both sources of data, providing information about learning system (and module) that the subjects will participate in, the subjects themselves, and the field observations.

It then describes the data mining techniques that will be used to generate our automated detectors.

## 2.1 Learning System and Subjects

We will model engagement within the context of hemorrhage control learning materials in vMedic (a.k.a. TC3Sim) serious game used to train US Army combat medics and lifesavers on tasks associated with dispensing tactical field care and care under fire. It is part of the US Army's objective to devise portable learning platforms that can be deployed quickly, effectively, and inexpensively to US warfighters stationed around the world. vMedic has been integrated with the Generalized Intelligent Framework for Tutoring (GIFT) (Sottilare, Goldberg, & Holden, 2012) that monitors real-time interaction and can trigger feedback scripts for participants based on actions in the game and performance.

Designed to support real-time performance messages, current and predicted cognitive and affective states, GIFT is modular and service-oriented. GIFT contains the components of sensor, trainee, pedagogical, learning management system (LMS), and domain modules. The automated detectors developed in this project will enhance the trainee module, towards providing support for creating pedagogical interventions (driven by the pedagogical model) associated with game-specific actions in vMedic. By better modeling trainees' states, and understanding the roles different affective states and disengaged behaviors play during learning, we can generate pedagogical support tailored to individual trainees, helping to realize ARL's vision of tailored, self-regulated, individual tutoring experiences for U.S. Army trainees.

Compared with other systems, game-based environments like vMedic, where trainees interact with a virtual world through an avatar, place fewer constraints on learner actions than problem-solving systems (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; Baker et al., 2012) or dialogue tutors (Litman & Forbes-Riley, 2006). Some virtual environments may present more constraints on learner behavior than others. vMedic is more restrictive than EcoMUVE (Baker, Ocumpaugh, Gowda, Kamarainen, & Metcalf, in press) where students have considerable freedom to explore the virtual world as they please. While vMedic allows a considerable amount of learner control, scenarios impose structure on trainee experiences through events that are triggered within the scenarios independent of the participant's actions (e.g., explosions and injuries occur and require attention irrespective of the actions of the participant within the scenario). These scenarios provide help in focusing the participant's attention to the objectives of the game (administering care) and implicitly guide trainee experiences toward key learning objectives.

For this study, predominantly first year cadets from West Point will be observed at the beginning of the academic year.  The cadets' voluntarily participation in a one-hour session will be held in a West Point computer laboratory. At the start of the session, cadets will be fitted with Q-sensors and synchronized with Kinect depth sensors (for subsequent analyses, outside the scope of this paper, of the relative value of sensor-based detectors and interaction-based detectors). Cadets then initiate a unique user profile by logging into GIFT. Cadets will be asked to answer questionnaires on self-efficacy, complete a pre-test on hemorrhage control, and review a PowerPoint presentation on hemorrhage control. Following the PowerPoint presentation, the cadets engage in approximately 20 minutes of the vMedic game, and then complete a post-test.

## 2.2 Quantitative Field Observations (QFOs)

In this study, Quantitative Field Observations (QFOs) will be collected using the Baker-Rodrigo Observation Method Protocol (BROMP) (Ocumpaugh, Baker, & Rodrigo, 2012). As mandated by BROMP, coders must be certified, achieving an adequate inter-rater reliability of Kappa = 0.6 with another BROMP-trained coder. BROMP has been used for several years to study behavior and affect in educational settings (Baker, D'Mello, Rodrigo, & Grasser, 2010; Baker et al., 2011; Baker et al., 2012; Rodrigo & Baker, 2009) and has been used as the basis for successful automated detectors of affect (Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; Baker et al., 2012). Observations in this study will be conducted by two BROMP-certified coders.

Within the BROMP protocol, behavior and affective states are coded separately but simultaneously using the Human Affect Recording Tool (HART), an application developed for the Android platform (and freely available as part of the GIFT distribution). HART enforces a strict coding order, determined at the beginning of each session. Students or trainees are coded individually, and coders are trained to rely primarily on peripheral vision in order to minimize observer effects. The coder has up to 20 seconds to categorize each trainee's behavior and affect, but records only the first thing he or she sees. In situations where the trainee has left the room, where his or her affect or behavior do not match any of the categories in the current coding scheme, or when the trainee can otherwise not be adequately observed, a '?' is recorded, and that observation is eliminated from the training data used to construct automated detectors.

The typical coding schemes used by BROMP will be modified to accommodate the unique behaviors and affect that manifest for this specific cadet population. Affective states observed will include frustration, confusion, engaged concentration, boredom, surprise, and disdain. Behavioral categories will consist of on-task behavior, off-task behavior, psychopath (friendly fire, killing bystanders; may not be observed in practice), and WTF ("without thinking fastidiously") behavior (Goldberg, Holden, Brawner, & Sottilare, 2011).

## 3 Feature Distillation

In order to distill a feature set for our affect detectors, trainee actions within the software will be synchronized to the field observations. During data collections, both the handheld computers and the GIFT server will be synchronized to the same internet NTP time server. Actions during the 20 seconds prior to data entry by the observer will be considered as cooccurring with the observation. It is anticipated that the following features will be engineered using data from the actions that co-occurred with or precede the observations:

- Changes in the state of the casualty, both recent and since injury, including:
    - Blood pressure and volume
    - Heart rate
    - Bleed rate
    - Lung efficiency
- Attempts by player, both during clip and overall, to treat patient
    - Application of tourniquet
    - Checking vitals
    - Conducting a blood sweep
    - Communicating
    - Requesting medevac
- Player state in terms of attackers
    - Is player under fire?
    - Is player under cover? (Are they currently taking cover?)
    - Is player with unit?
    - Is casualty safe from further attack?
- Time between actions

These features will be initially constructed within Microsoft Excel as a rapid prototyping method. After we have determined which features are predictive of affect, these features will be integrated into the GIFT platform and automatically distilled. Other features may also be developed through the process of feature engineering and studying their effects when integrated into the system.

## 4 Machine Learning Process

It is anticipated that separate detectors will be built for each affective and behavioral construct studied, although detectors for very rare behaviors or affective states may not be developed, due to lack of data.

Each detector will be evaluated using leave-one-out trainee-level cross-validation. In this process, a detector is built using data from every trainee except one before being tested on that student. By cross-validating at this level, we increase confidence that detectors will be accurate for new trainees. In addition, re-sampling will be used to make the class frequency more equal for detector development. However, all performance calculations will be made with reference to the original dataset, as in (Baker et al., 2012).

## 5 Conclusion

It is the intent of the investigators on this project that the results will contribute to the Army Research Lab's goal of further developing the GIFT framework (Sottilare, Goldberg, & Holden, 2012). Developing automated detectors that can integrate sensor and interaction (including performance and history of the trainee) but can function effectively absent sensor data, is an important step in developing affect interventions for U.S. Army trainees across a range of settings.

Optimally leveraging these detectors will depend on several additional steps. First, it is essential to study the relationship between affect, engagement, and outcome variables, towards understanding which affective states and engagement variables need to be responded to in a suite of optimally effective computer-based tutoring systems for Army use. The data collected to develop these detectors can be expected to also be useful to accomplishing this goal. Second, interventions will need to be developed which leverage the detectors to effect change in trainees. There is an extensive literature on interventions for behavioral disengagement and affect, in online learning systems; however, it is not fully known which of these approaches will be effective with military trainees, a different population than the much younger populations these interventions have typically been developed for. Once interventions have been developed and tested, integrating automated detectors and interventions into vMedic through GIFT's Trainee Module and Pedagogical Module will provide a valuable example of how to respond to trainees' negative affect and disengagement, a valuable contribution in improving vMedic and similar interventions used by the U.S. Army.

**Acknowledgments.** This research was supported by a grant by the Army Research Lab, Grant #W9NF-13-2-0008. We thank our research colleagues at the Army Research Lab, Dr. Robert Sottilare, Dr. Benjamin Goldberg, Dr. Keith Brawner, and Dr. Heather Holden, as well as at North Carolina State University, Dr. James Lester, Dr. Bradford Mott, and Dr. Jonathan Rowe. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of ARL or NCSU.

**References**


Alzoubi, O., Calvo, R.A., & Stevens, R.H. (2009). Classification of EEG for affect recognition: An adaptive approach. In *AI 2009: Advances in Artificial Intelligence*, Springer, Berlin Heidelberg, Germany: Springer.

Baker, R.S.J.d., 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. In *International Journal of Human-Computer Studies,* 68, 223-241.

Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., & Rossi, L. (2012). Towards sensor-free affect detection in cognitive tutor algebra. In *Proceedings of Educational Data Mining 2012*, 126-133.

Baker, R.S.J.d., Moore, G., Wagner, A., Kalka, J., Karabinos, M., Ashe, C., Yaron, D. (2011). The Dynamics Between Student Affect and Behavior Occuring Outside of Educational Software. In *Proceedings of the 4th bi-annual International Conference on Affective Computing and Intelligent Interaction*.

Baker, R.S., Ocumpaugh, J., Gowda, S.M., Kamarainen, A., & Metcalf, S.J. (in press). Extending log-based affect detection to a multi-user virtual environment for science. To appear in *Proceedings of the 22nd Conference on User Modeling, Adaptation, and Personalization*.

Cocea, M., Hershkovitz, A., Baker, R. S. J. d. (2009). The impact of off-task and gaming behaviors on learning: Immediate or aggregate? *In Proceedings of the 14th international conference on Artificial Intelligence in Education*, 507-514.

Conati, C. Probabilistic assessment of user’s emotions in educational games. In *Journal of Applied Artificial Intelligence*, 16, 555-575.

D'Mello, S.K., Taylor, R., Grasser, A.C. (2007). Monitoring affective trajectories during complex learning. *In proceedings of the 29th Annual Cognitive Science Society*, 203-208.

Dragon, T., Arroyo, I., Woolf, B.P., Burleson, W., Kaliouby, R.e., Eydgahi, H. (2008). Viewing student affect and learning through classroom observation and physical sensors. In *Proceedings of the International Conference on Intelligent Tutoring Systems*, 29-39.

Forgas, J.P. (2000). Managing moods: Toward a dual-process theory of spontaneous mood regulation. In *Psychological Inquiry*, 11, 172-177.

Goldberg, B., Holden, H., Brawner, K., & Sottilare, R. (2011). Enhancing performance through pedagogy and feedback: Domain considerations for ITSs. In *Interservice/Industry Training, Simulation, and Education Conference, I/ITSEC*.

Lee, D.M., Rodrigo, M.M., Baker, R.S.J.d., Sugay, J., Coronel, A. (2011). Exploring the relationship between novice programmer confusion and achievement. In *Proceedings of Affective Computing and Intelligent Interaction 2011*, 175-184.

Linnenbrink, E.A., & Pintrich, P.R. (2000). Multiple pathways to learning and achievement: The role of goal orientation in fostering adaptive motivation, affect, and cognition. In *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance*. C. Sansone & J.M. Harackiewicz (Eds.). San Diego, CA: Academic Press. (2000), 195-227.

Litman, D.J., & Forbes-Riley, K. (2006). Recognizing student emotions and attitudes on the basis of utterances in spoken tutoring dialogue with both humans and computer-tutors. In *Speech Communication*, 48, 559-590.

Liu, Z., Pataranutaporn, V., Ocumpaugh, J., Baker, R. S. (2013). Sequences of frustration and confusion, and learning. *In Proceedings of the 6th International Conference on Educational Data Mining*, 114-120.

Mohammad, Y., & Nishida, T. (2010). Using physiological signals to detect natural interactive behavior. In *Spring Science & Business Media*, 33, 79-92.

Ocumpaugh, 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. In *EdLab*. New York, NY & Manila, Philippines: Ateneo Laboratory for the Learning

Paquette, L., Baker, R.S., Sao Pedro, M.A., Gobert, J.D., Rossi, L., Nakama, A., & Kauffman-Rogoff, Z. (in press). Sensor-free affect detection for a simulation-based science inquiry learning environment. To appear in *Proceedings of the 12th International Conference on Intelligent Tutoring Systems*.

Pardos, Z., Baker, R.S.J.d., San Pedro, M.O.Z., Gowda, S.M., Gowda, S. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *In Proceedings of Leaning Analytics and Knowledge 2013*, 117-124.

Pekrun, R. (1992). *Educational environments and intrinsic/extrinsic academic motivation: Expectancy value theory and longitudinal evidence.* In paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Rodrigo, M.M.T., & Baker, R.S.J.d. (2009). Coarse-grained detection of student frustration in an introductory programming course. In *Proceedings of the International Computing Education Research conference 2009: the International Computing Education Workshop*.

Sabourin, J., Rowe, J., Mott, B., Lester, J. (2011). When off-task is on-task: The affective role of off-task behavior in narrative-centered learning environments. In *Proceedings of Artificial Intelligence in Education 2011*, 534-536.

Schunk, D.H., & Zimmerman, B.J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. In *Reading and Writing Quarterly* 23, 7-25.

Sottilare, R.A. Goldberg, B., & Holden, H. (2012). The Generalized Intelligent Framework for Tutoring (GIFT).

Wixon, M., Baker, R.S.J.d., Gobert, J., Ocumpaugh, J., & Bachmann, M. (2012). WTF? Detecting students who are conducting inquiry Without Thinking Fastidiously. In *Proceedings of the 20th International Conference on User Modeling, Adaptation and Personalization, UMAP (2012*), 286-298.

Zeidner, M. (1998). *Test Anxiety: The State of the Art. Berlin, German.*