# Monitoring Engagement and Motivation Across Learning Environments

**Benjamin Bell, Elaine Kelsey**
**Eduworks Corporation**
**Corvallis, OR**
**benjamin.bell@eduworks.com**
**elaine.kelsey@eduworks.com**

**Benjamin Nye**
**University of Southern California**
**Institute for Creative Technologies**
**Los Angeles, CA**
**nye@ict.usc.edu**

## ABSTRACT

The U.S. military is broadening the use of modeling and simulation to inject greater interactivity in learning environments, an approach likely to be welcome in early-stage, heavily academic training. Qualifying for specific military occupations and specialties requires a learner to master broad foundational knowledge delivered in courses lasting up to several months. Incorporating simulation and gaming is expected to improve learning, retention and training throughput for academically-intensive courses that require learning volumes of material in principally static formats. However, migrating interactivity into academic learning environments brings with it the need to monitor learner engagement and adapt to any detected lapses. In this paper we summarize work performed by Eduworks Corporation and the Institute for Creative Technologies (ICT) aimed at addressing this need. We present an innovative software appliance called the Tracking and Assessing Learner Engagement Toolkit (TALENT). We describe our methodology, design and prototype for providing metrics and persistent assessments to enhance the simulation-based training and education enterprise with adaptive support for learner engagement, and conclude with a discussion of future directions and potential benefits of this work.

## ABOUT THE AUTHORS

**Dr. Benjamin Bell** is the president of Eduworks, where he leads simulation, training, human-machine interaction, and decision support development. His research has addressed the use of simulation for training and education across a spectrum of applications, including K-12, higher education, military, and national security training. He has held faculty positions, chief executive positions in industry, and leadership roles for several international conferences. He is an adjunct professor at Embry Riddle, holds a PhD from Northwestern, and is a graduate of the University of Pennsylvania.

**Elaine Kelsey** is a research engineer and computational linguist with Eduworks Corporation, where she focuses on developing natural language processing and machine learning solutions for a range of government and commercial applications in intelligent tutoring. She most recently led development of Eduworks' automated question generation, assessment, and competency alignment algorithms. She holds multiple bachelor's and master's degrees in computer science, linguistics, molecular biology, and biostatistics, and is currently working towards a combined MS / PhD in machine learning and computational linguistics.

**Dr. Benjamin Nye**, the Director of Learning Science at ICT, leads the ARL Promoting Engagement in Virtual Learning Environments (ENGAGE) project and is co-PI on the ONR Personal Assistant for Life-Long Learning (PAL3) project, which offer insights on affective engagement and social engagement, respectively. His research has been recognized for excellence in tutoring systems and realistic behavior in simulations. His research is on scalable learning technologies and design principles to promote learning, and has yielded over 20 peer-reviewed papers, 12 book chapters, and 1 book, and 5 open-source projects.

# Monitoring Engagement and Motivation Across Learning Environments

**Benjamin Bell, Elaine Kelsey**
**Eduworks Corporation**
**Corvallis, OR**
**benjamin.bell@eduworks.com**
**elaine.kelsey@eduworks.com**

**Benjamin Nye**
**University of Southern California**
**Institute for Creative Technologies**
**Los Angeles, CA**
**nye@ict.usc.edu**

## PROBLEM SUMMARY

To support a complex array of current and envisioned missions, the U.S. military trains and educates a large and diverse uniformed workforce. We focus in this work on the U.S. Air Force, where early-stage training in many Air Force Specialty Codes (AFSC) presents airmen with large corpora of foundational knowledge to master, delivered through several traditional means including live classroom instruction and computer-based training. Many AFSCs related to maintenance, for instance, are required to master the basic principles of mechanics and electronics. This lengthy content is taught at technical training schools in courses lasting up to several months. Enhancing training through motivation and engagement is a necessary element in the recruiting and retention of airmen, helping to preserve and grow a cadre of qualified personnel in mission-critical areas like aerospace maintenance.

The Air Force is broadening its use of interactive activities and games in delivering training and education curricula, both as informal supplemental education and as part of a syllabus. Air Force education and training stakeholders continue to look for ways to engage and motivate their constituencies. In AFSCs that face shortages of critical personnel, the need to present airmen with dynamic, interactive education has mission-ready consequences.

For games and game-play to improve engagement in computer-mediated learning contexts, researchers and training developers must measure engagement and motivation, for two key purposes: (1) to identify which techniques and approaches offer the greatest efficacy; (2) to enable learning systems that can identify and adapt to detected lapses in engagement. Success requires valid constructs, measures, and software to enable application of these metrics across the community of training developers.

## TECHNICAL APPROACH

The goals of the work reported here were: (1) Perform a work analysis to identify requirements and select exemplar AFSCs and representative content; (2) Adapt and synthesize existing research-based models of engagement and motivation; (3) Design metrics/measures of engagement and motivation; (4) Design an adaptive instruction appliance; and (5) Create a proof-of-concept application of metrics using a surrogate online learning activity.

We developed a model of motivation and engagement based on existing research constructs, selecting performance markers that provide metrics defined by that model, designing an architecture for monitoring a learning environment for those metrics, and proposing an appliance that could provide general recommendations in real-time to learning environments to combat detected lapses in motivation and engagement. In this preliminary work, we also demonstrated the potential for using TALENT in concert with a learning environment to detect lapses.

## RESULTS

### Air Force Specialty Code and Content Exemplars

This work focuses on earlier phases of technical training, where the large corpora of general technical and theoretical knowledge make engagement and motivation salient factors in successful outcomes. We explored the training pathways for multiple AFSCs and shreds (an alphanumeric suffix to an AFSC designating an additional specialization). We selected the 2A5XX (Aerospace Maintenance) codes, encompassing four specific codes based on aircraft categories (2A5X1 – Airlift/Special Mission Aircraft Maintenance; 2A5X2 – Helicopter/Tiltrotor Aircraft Maintenance; 2A5X3 – Mobility Air Forces Electronic Warfare Systems; 2A5X4 – Refuel/Bomber Aircraft Maintenance).

We focused our attention on the period of training immediately following receipt of a shred assignment, when airmen are required to continue learning or reviewing electronics and mechanics principles and core theoretical knowledge as a necessary prerequisite to later instruction in code- and shred-specific systems. Our analysis included the Electronics Principles courses required for 2A5XX specializations, large portions of which align with the content in the Navy Electricity and Electronics Training Series (NEETS).

**Engagement and Motivation Model**

We adapted previous research-based models to create an initial model of engagement and motivation suitable for digital learning environments. This TALENT model lends significant insight into representative measures useful in detecting engagement and motivation lapses in Air Force learning tasks, where we focused our investigation on Aerospace Maintenance.

We investigated both models of motivation (describing a cognitive state or trait) and models of engagement (describing resulting behaviors). We compiled an inventory of nine relevant engagement and disengagement models from the literature that emphasizes behavioral indicators (e.g., data from log files or from direct queries to the user) (Core, et al., 2016). These included Intrinsic vs. Extrinsic (Porter & Lawler, 1968); Two Factor Hygiene-Motivator Theory (Gawel, 1997); Motivators from Maslow's Hierarchy (Ibid); Achievement Goal Theory (Pintrich, 2000); D'Mello & Graesser's Engagement model (2012); and Baker's indicators of passive vs. active disengagement (Baker, Corbett, Roll & Koedinger, 2008). From this we synthesized a multi-timescale engagement and motivation model shown in Figure 1.

Our TALENT model emphasizes measurable behavioral indicators, with a particular emphasis on engagement issues central to DoD training needs (e.g. burnout due to long periods of intensive study, and the need to maintain consistent engagement while experiencing variable levels of subjective interest), and which can be measured by multiple indicators across a wide range of digital training environments.

For each of the factors in the unified model, we developed possible indicators, including directly observed or collected behavioral data (e.g. input device logs, sensor data), observed in-system actions (e.g. data from log files of user actions), and direct queries to the user. The goal was to develop a robust model that could perform consistently across a variety of systems and applications, with the assumption that only a subset of data points would be readily available for analysis in any given environment. As such, the system includes multiple possible variants for calculating each factor, based on the data already available in each system it interfaces with. Likewise, where richer data collection exists, the model is able to take advantage of it to further refine and adjust estimates.

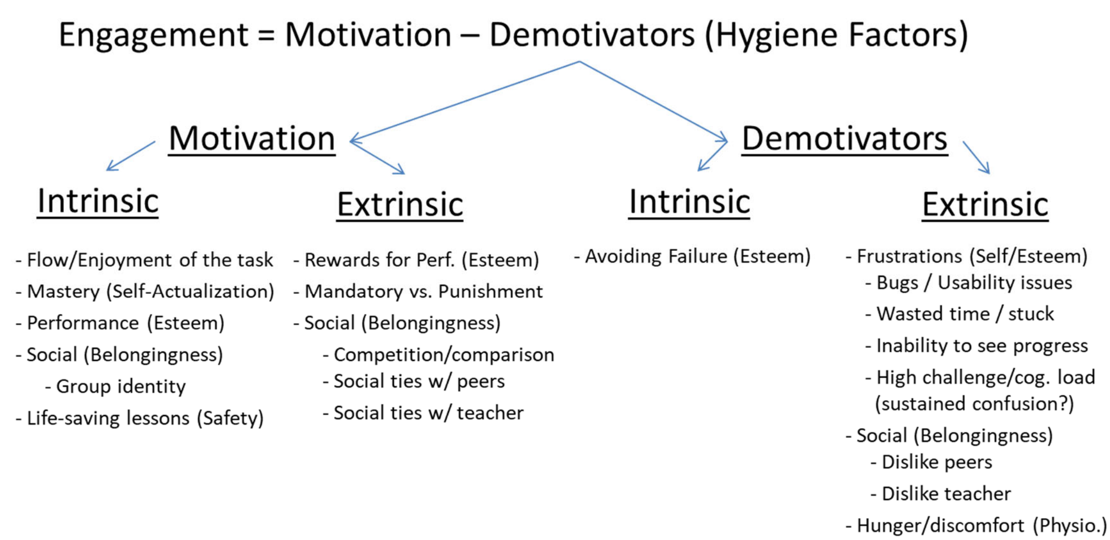


**Figure 1. TALENT Engagement and Motivation Model.**

## Metrics

TALENT requires a suite of metrics to detect engagement lapses from a broad set of measures across diverse learning environments. We compiled an inventory of metrics aligned with this model and derived from our previous research, based on several factors. A first set of factors measures through active involvement, positive affect, or initiative. Researchers have noted, for instance, that engaged people exhibit goal-driven behaviors and express intensity, focus, interest and persistence (Connell, 1990; Furrer & Skinner, 2003). By contrast, apathy and distractedness can signal disengagement. A second set of factors relates to exhibiting initiative. People that express their voice and take initiative in a learning situation are seen as trying to effect change, whereas disengaged people are more passive in accepting external forces governing their task environment (deCharms, 1976; Fiedler, 1975; Koenig, et al., 1977). A third set relates to indirect indicators of engagement, like affect. The metrics calculated for this initial prototype are listed in Table 1.

**Table 1. Metrics for initial prototype. Bold = directly measured; italicized = inferred from direct measures.**

| Name | Level | Motivation Type | Inputs | Description |
|---|---|---|---|---|
| **Mandatory** | Social | Extrinsic | Structural/Self-reported | Required # of hours or completion requirements. Penalty severity for not completing. |
| **Peer Social Ties** | Social | Extrinsic | Structural/Self-reported | Enjoyment or value in participating in activity with peers |
| **Leadership Social Ties** | Social | Extrinsic | Structural/Self-reported | Enjoyment or value in participating due to the leader (e.g., instructor, supervisor) |
| **Peer Mismatch** | Social | Extinsic | Self-report | Dislike of peers or other issues with peers |
| **Instructor Mismatch** | Social | Extinsic | Self-report | Dislike of the instructor or their general approach. Could also compare |
| ***ExpectedUtility*** | Social/Rational | Intrinsic | Self-report, Resource Logs | Self-reported value of content (relevance, need), relative time vs. other content |
| **Motivation-Mastery** | Social/Rational | Intrinsic | Self-report, System Use | Mastery orientation of the learner, as evidenced by self-report and by viewing those functions if optional |
| **Motivation-Achievement** | Social/Rational | Intrinsic | Self-report, System Use | Achievement orientation, as evidenced by self-report, by viewing related system functions, and by lower level of challenge-seeking |
| **Motivation-Evasion** | Social/Rational | (De)Intrinsic | Self-report, System Use | Evasion orientation of the learner, as evidenced by self-report and by avoiding system activities and scoring where possible (e.g., passive resources) |
| *Motivation-Exploration/Gamification* | Rational | Extrinsic | Self-report, System Use | Appreciation of in-activity novelty and rewards (e.g., customization, easter eggs, funny distractions), from viewing in-game functions or self-report |
| **Intent to Use** | Rational | Intrinsic+Extrinsic | Self-report | Stated intention to use the system, in terms of #/week and #h/session if content is |
| **Interest** | Rational | Intrinsic | Self-report, Resource Logs | Self-reported interest in content, relative time on content vs. other content (no way to disentangle the second from utility?) |
| *Evasion* | Rational | Intrinsic | Self-report, Resource Logs | Avoidance behaviors to avoid failure without mastery such as active avoidance (gaming the system, hint abuse) and passive (skipping hard activities) |
| **Lack of Progress/Stuck** | Rational | (De) Extrinsic | Resource Logs, Self-report | Lack of features to display progress on goals that are meaningful to the learner (so need to know their goals, then compare against system features+display events). |
| **Usage: Adjusted Resource Time** | Rational | N/A (Metric) | Resource logs | Total time spent in resources, up to some reasonable max per resource (e.g., up to 1-stddev over the average non-minimal time) |
| **Usage: Longevity** | Rational | N/A (Metric) | Session login/logouts | Time from first login to most recent login. |
| **Usage: Frequency** | Rational | N/A (Metric) | Session login/logouts | Number of logins/time span |
| **Usage: Active** | Rational | N/A (Metric) | Session login/logouts | Time since last login and resulting likelihood of return, based on prior data |
| **Time-on-Task: Session Time** | Cognitive | N/A (Metric) | Session login/logouts | Calculate time breakdown in different activities/sessions, to capture time-on-task. |
| **Time-on-Task: Task Time** | Cognitive | N/A (Metric) | Resource logs, knowledge components | Time for each task, indicating engagement on that task. Normed by user, content, and task (interested in if user is spending more time on task than normal vs. task) |
| *Learning Gains* | Cognitive | N/A | Resource logs, knowledge components | Calculate learning gains between pre-test & post-test, based on question batteries aggregated by arbitrary categories (e.g., knowledge components or other taxonomies). |
| *Interaction Levels* | Cognitive | N/A (Metric) | Interaction logs | Clicks, verbosity, optional inputs, explortation level indicating high levels of |
| **Decision Events/Correctness** | Cognitive | N/A (Metric) | Interaction logs | Calculate metrics about attempts to answer questions or solve problems, including if the attempt was correct or otherwise. |
| *Support Levels: Hint Abuse* | Cognitive | Extrinsic | Interaction logs | Track attempts to game the system such as frequently requesting hints, without trying |
| **Usability Issues** | Cognitive/Affective | (De) Extrinsic | Interaction logs | Problems experienced using system, which prevent interactions from being productive |
| **Time-Waste** | Cognitive/Affective | (De) Extrinsic | Interaction logs | Delays, wait time, and fluff in resources and the system |
| *Repeated Failure* | Cognitive/Affective | (De) Extrinsic | Resource Logs, Self-report | Low successful completion rate of resources over some span (Sustained confusion) |
| **Affect: Engagement** | Affective | (Aggregate) | Self-report | Feeling of engagement with the activity (session or task) |
| *Affect: Confusion* | Affective | (Aggregate) | Self-report | Feeling of confusion during the activity (session or task) |
| *Affect: Frustration* | Affective | (Aggregate) | Self-report | Feeling of frustration during the activity (session or task) |
| *Affect: Boredom* | Affective | (Aggregate) | Self-report | Feeling of boredom during the activity (session or task) |
| **Affect: Anxiety** | Affective | Demotivator | Self-report | Test anxiety and anxiety learning. Likely could be bundled into evasion |

To support implementation of these measures as *computable* metrics, we model the engagement process in terms of multiple loops of cognitive regulation. In this conceptualization, each level builds on the previous: biological responses (e.g., attention and affect) underpin cognition in learning; cognition (e.g., deliberate practice) is necessary for reasoning about the value of tasks or goals; rational decisions (e.g., motivations to return/continue studying) are required to build social engagement (e.g., study teams, help-seeking, identity-formation). Ultimately, higher-level engagement metrics will be calculated by aggregating metrics of lower levels of cognition, including behavioral metrics (e.g., data mining system events), self-report (e.g., motivation, future plans), and records of performance.

**Recommending Adaptive Interventions**

To extend beyond detecting lapses into correcting lapses, we created a preliminary framework for flowing low-level data from our inventory of metrics up through the model, to support adaptive interventions. We designed an extensible framework for recommending these adaptive interventions for a particular learner for a given learning system. The framework (Figure 2) associates each metric that can be consumed by TALENT with triggers and potential adaptations. The design accommodates combinations of metric outputs, which may be used to paint a broader picture of a learner's state, and thus to select and recommend a mixture of adaptations that are appropriate for the learner's state and supported in the target learning system.

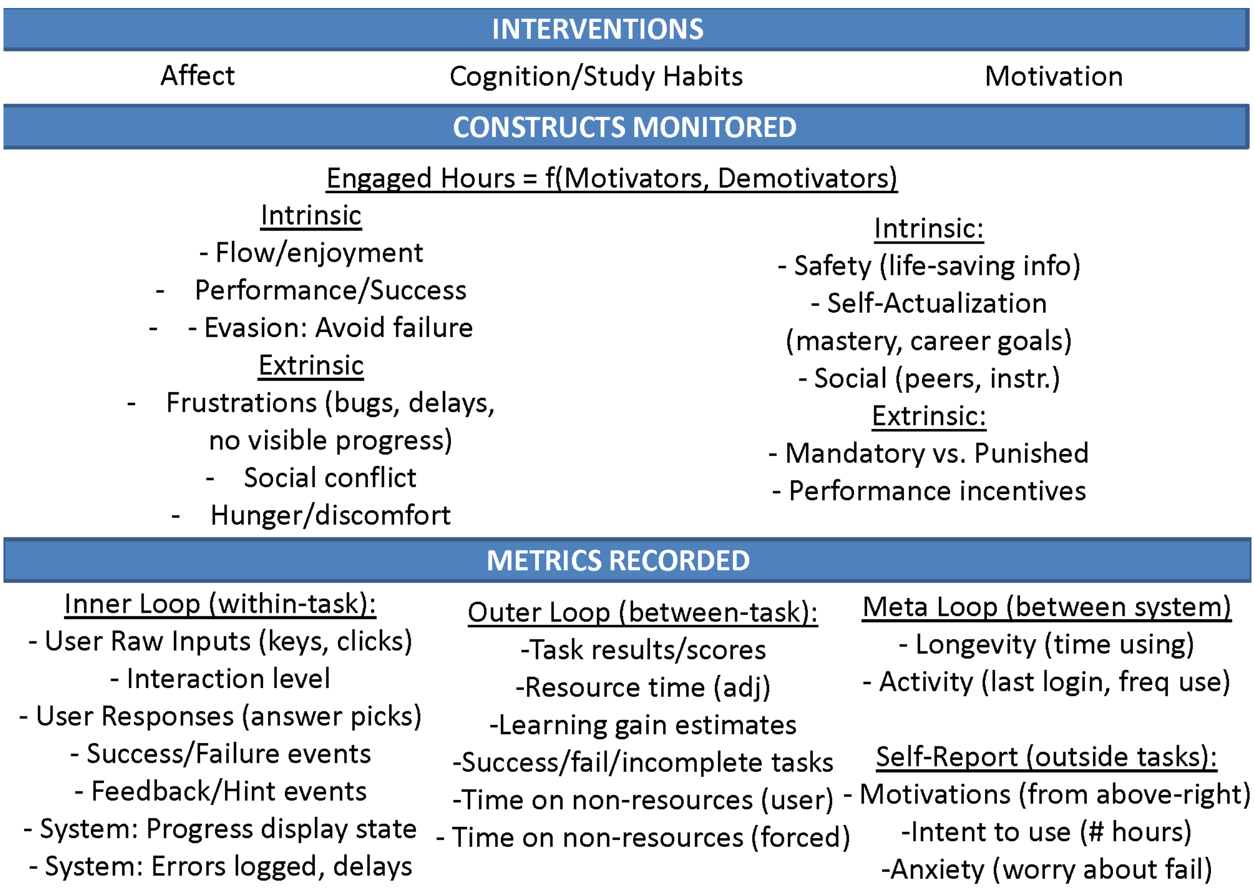


**Figure 2. Interventions (top) informed by constructs (middle) weighted by metrics (bottom).**

We applied Micro-adaptation and Aptitude Treatment Interaction (ATI) theories in order to derive preliminary adaptive interventions. Cronbach & Snow (1977) proposed ATI as a framework for instructional manipulations applied before training begins (aspects of training based on learner interest, learning orientation and styles, and aptitudes such as digital intelligence, cognitive styles, or prior knowledge and experience). ATI is thus well-suited for adaptive training in military domains, because although the learner population is less heterogeneous than the general population, there will usually be differences in experience levels and prior knowledge.

We also considered Micro-adaptive approaches, which respond to specific user actions and responses during a training session, by making incremental and real-time adjustments to aspects of the training (Holland, 1977). The adaptations we considered included switching topics, altering the level of difficulty, or changing the kind of feedback provided to the learner (Goldberg, et. al, 2012).

We ultimately adopted a composite model (Tennyson & Christensen, 1988) that posits a two-step approach where adaptation is based both on learners' prior skills and aptitudes and on their performance during training. This model calls for a pre-training step where ATI techniques establish an appropriate level of difficulty or modify content sequencing and format. Micro-adaptive approaches are then applied during training, to assess performance and monitor behaviors in real time and to use those measurements to adapt training to the learner's current needs.

We applied this composite model in creating TALENT's adaptation handlers. Our design proposes that ATI adaptations can be applied to customize delivery for a learner based on historical performance and profile data, and that Micro-adaptations can be applied to adjust delivery in real time based on changes in performance data detected by TALENT. Table 2 presents our initial adaptation primitives from which TALENT could recommend interventions to learning environments. These interventions have shown some indications of efficacy in various learning systems, though a comprehensive system to apply each of these has not been implemented. As such, our future research will include considering how these interventions should be coordinated together.

**Table 2. Adaptation Primitives**

| Category | Interventions |
|---|---|
| Affective | • Support: Messaging related to emotional support (e.g., frustration, confusion) |
| Cognition & Study Habits | • Difficulty: Easier/harder tasks<br>• Guidance: More/less help or adaptive informational messages<br>• Sequence control: More/less ability to choose next task, to skip the current task, or to return to review a task<br>• Content chunking: Smaller/larger tasks<br>• Task Types: More/less of certain tasks (media, examples, enriched interactive tasks, knowledge checks, realistic tasks)<br>• Messaging addressing evasive behaviors (e.g., skipping resources, text anxiety)<br>• Messaging confronting active disengagement (e.g., overuse of hints, cheating) |
| Motivation | • Mastery Orientation: Salient displays of resource completion and indicators of learning/mastery (e.g., show improvement), keeping in learning activities, messaging related to goals<br>• Social: More/less communication with peers or instructors<br>• Gamification: Presenting internal rewards, achievements, or fanfare for success<br>• Growth Mindset: Messaging aligning difficulty/confusion with later mastery (Dweck, 2010) |

**Architecture and Prototype**

To test and validate the model and corresponding metrics, we created a demonstration prototype consisting of a basic ingestion pipeline for harvesting and transforming information from ICT's Personal Assistant for Life-Long Learning (PAL3). PAL3 (Swartout, Nye, et al., 2016) is an actively developed prototype for guiding Sailors through learning resources to reach specific learning objectives. PAL3 employs adaptive training which attempt to harness multiple mechanisms for motivation: mastery orientation, social cooperation / competition, exploration, and effort-based awards.

Our architecture includes a Learning Record Store (LRS) for long-term aggregation and data format normalization, a dashboard for displaying metrics and viewing stored data records, and a set of algorithms and models for deriving metrics directly from the stored data records (a schematic depiction of this architecture is shown in Figure 3). This proof-of-concept defers implementation of the adaptive recommendations, focusing first on the appliance API and calculation of engagement metrics.

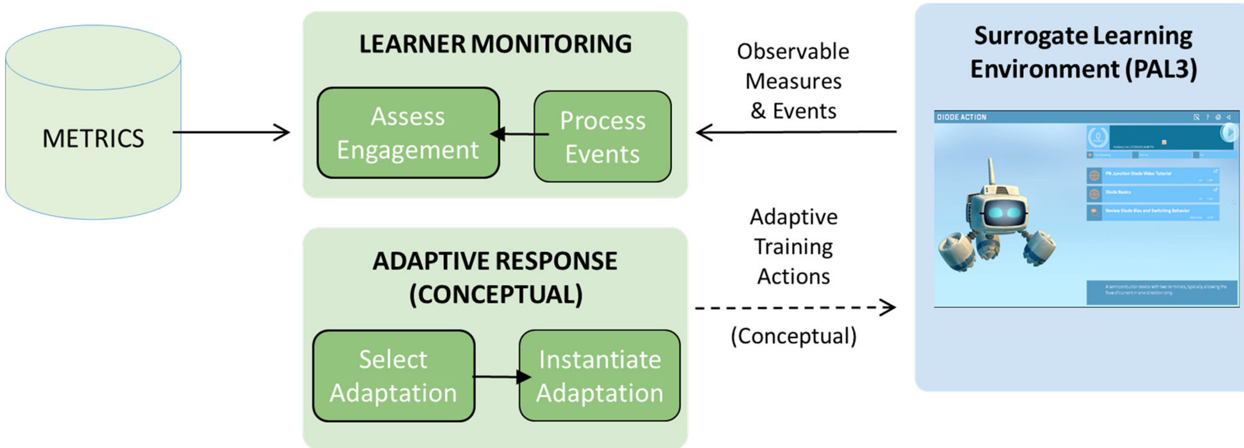


**Figure 3. Schematic depiction of Phase I architecture.**

PAL3 was a useful surrogate for testing initial prototypes of engagement metrics. Our exemplar case of Aerospace Maintenance AFSC is well-aligned with the current PAL3 content addressing foundations for Navy electronics. The PAL3 system is also a useful platform for studying and building engagement metrics, because it records a persistent life-long learning record (xAPI records) suitable for data mining.

The dashboard is intended to accommodate the occasional need for personnel to inspect or configure system components. So although TALENT will operate "behind the scenes", running in real time in concert with a learning environment, the dashboard allows access to historical data, viewing metric results, and configuring an appliance after it has been generated for a specific learning environment. Figure 4 shows the dashboard displaying metric calculations in realtime as TALENT ingests data from PAL3.

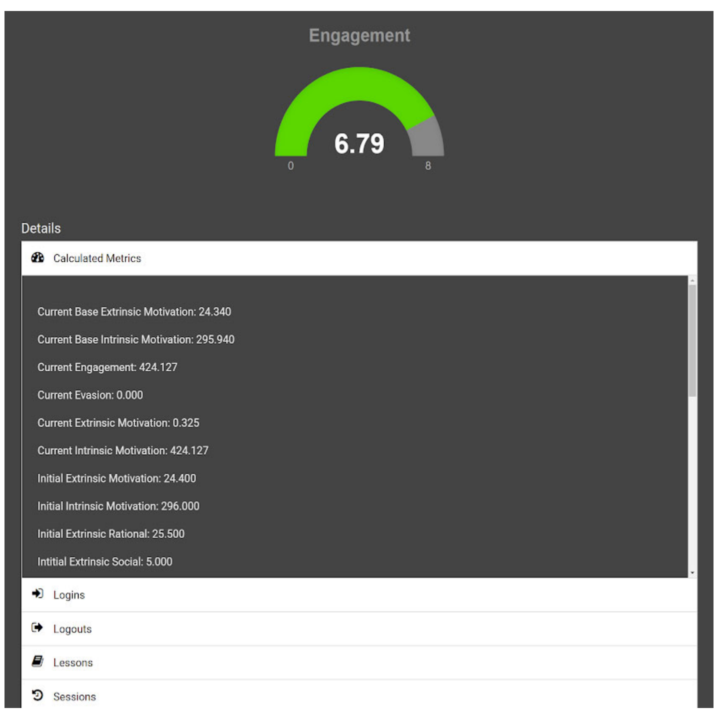


**Figure 4. Dashboard, showing TALENT calculating metrics from PAL3 data in realtime.**

## CONCLUSION AND FUTURE WORK

PAL3 exhibited gaps where the data provided to the TALENT prototype (refer back to Table 1) did not represent optimal indicators for those input elements. For example, PAL3 provided limited log data about users' interaction with the system when completing activities outside the system (in linked learning resources). These gaps proved useful for the purpose of testing the robustness of the engagement model, as we were able to analyze how well the model compensated for missing input elements by extrapolation from other elements for which it had richer data and/or through the use of generic models for these behavioral elements. The prototype was able to generate a near realtime model of learner engagement and to accurately predict real world learner engagement (determined by comparing its projections to a human-determined assessment of actual engagement levels under a series of simulated testing scenarios).

TALENT demonstrates a preliminary model, metrics and general appliance for detecting motivation lapses in learning environments. Our results provide concept validation and establish a development and integration roadmap to develop a service-oriented appliance that client learning applications can employ for detecting lapses in engagement and motivation, and for recommending adaptive interventions. Subsequent work can advance these results to realize general-purpose services, available to a broad range of digital learning environments.

Across the service branches, education and training initiatives must create engaged and motivated warriors, using adaptive instruction and providing data to help training managers track the efficacy of new technologies and paradigms. TALENT will contribute to continued success in these endeavors.

## REFERENCES

Baker, R. S., Corbett, A. T., Roll, I., & Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction, 18*(3), 287-314.

Connell, J.P. (1990). Context, self, and action: A motivational analysis of self-esteem processes across the life-span. In D. Cicchetti (Ed.), The self in transition: From infancy to childhood (pp. 61–97). Chicago: U Chicago Press.

Cronbach, L. J., and Snow, R. E. (1977). Aptitudes and instructional methods. A handbook for research on interactions. New York: Irvington.

Core, M.G., Georgila, K., Nye, B.D., Auerbach, D., Liu, Z.F., DiNinni, R. (2016). Learning, adaptive support, student traits, and engagement in scenario-based learning. In *I/ITSEC, 2016.*

deCharms, R. (1976). Enhancing motivation: Change in the classroom. New York: Irvington.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145-157.

Dweck, C. S. (2010). Even geniuses work hard. Educational Leadership, 68(1), 16-20.

Fiedler, M. L. (1975). Bidirectionality of influence in classroom interaction. Journal of Educational Psychology, 67, 735–744.

Furrer, C., and Skinner, E. A. (2003). Sense of relatedness as a factor in children's academic engagement and performance. Journal of Educational Psychology, 95, 148–162.

Gawel, J. E. (1997). Herzberg's theory of motivation and Maslow's hierarchy of needs. *Practical Assessment, Research & Evaluation, 5*(11), 3.

Goldberg, B., Brawner, K., Sottilare, R., Tarr, R., Billings, D. R., and Malone, N. (2012). Use of evidence-based strategies to enhance the extensibility of adaptive tutoring technologies. In Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Holland, J. G. (1977). Variables in adaptive decisions in individualized instruction. Educational Psychologist, 12, 146–161.Koenigs, S. S., Fiedler, M. L., and deCharms, R. (1977). Teacher beliefs, classroom interaction and personal causation. Journal of Applied Social Psychology, 7, 95–114.

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning & achievement. *J. of Educational Psych, 92*(3), 544.

Porter, L. W., & Lawler, E. E. (1968). Managerial attitudes and performance.

Swartout, B., Nye, B. D., et al. (2016). Designing a Personal Assistant for Life-Long Learning (PAL3). Florida Artificial Intelligence Research Society (FLAIRS) 2016 Conference. AAAI Press.

Tennyson, R.D. and Christensen, D.L. (1988). MAIS: An intelligent learning system. In D. Jonassen (Ed.), Instructional designs for micro-computer courseware (pp. 247–274). Mahwah, NJ: Erlbaum.