INSTITUTE FOR DEFENSE ANALYSES

Affect-Sensitive Instructional Strategies for Synchronous Distance Learning

John E. Morrison
Emily A. Fedele

September 2020

Approved for public release;
distribution is unlimited.

IDA Document D-14245

Log: H 20-000230

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

IDA

About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, Project AI-2-4637, "Evaluation of Emotive Recognition Technology for Instructional Feedback and Assessment," Navy Education and Training Command/Research, Modeling and Analysis (NETC/OPNAV N1T) and the Office of the Under Secretary of Defense, Research and Engineering. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

For More Information

John E. Morrison, Project Leader
jmorriso@ida.org, 703-578-2839

Leonard J. Buckley, Director, Science and Technology Division
lbuckley@ida.org, 703-578-2800

Copyright Notice

INSTITUTE FOR DEFENSE ANALYSES

IDA Document D-14245

Affect-Sensitive Instructional Strategies for Synchronous Distance Learning

John E. Morrison
Emily A. Fedele

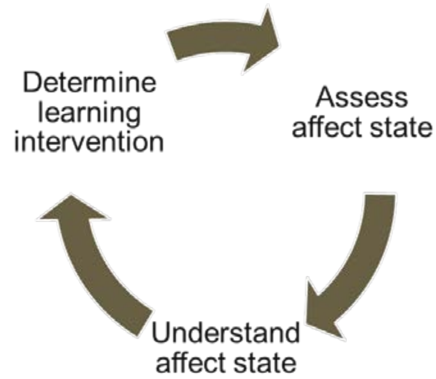
Executive Summary

Problem and Objective

The Submarine Learning Center employs a synchronous distance-learning environment based on virtual-world technology. This learning technology makes it possible to provide expert instruction to sailors located around the world, allowing them to interact in an engaging, virtual learning environment. In this system, students are represented as avatars interacting with each other and virtual objects in a compelling 3D simulation. Although virtual-world technology provides audio and text communication between students and instructors, it does not provide nonverbal student feedback to the instructor indicating the student's emotional state. Such emotive cues provide the instructor valuable information to adjust and adapt the pace and content of instruction to the students' affective and cognitive states.

The emerging technology of automated affect recognition provides an innovative approach to providing nonverbal instructional feedback. Over the last 20 years, there have been significant advances in the capability of detecting student affect states as they learn. However, to take full advantage of this technology, an instructional system must not only detect the affective states of students but also respond appropriately to those states. Thus, the development of an affect-sensitive learning system must address three separate problems. The first is how to dynamically collect cognitive and affective information from the learner to *assess affective state*. The second is how to understand and model the implications that those affective states have on instruction. Given that understanding, the third is to choose an appropriate instructional intervention for individual students and contexts. Once the intervention is deployed, student affect is reassessed and the cycle restarts. All three problems or goals for developing affect-sensitive instruction are depicted in the figure as an iterative loop.

The objective of this report is to examine and assess the maturity of the science and technology behind the three problems faced in developing affect-sensitive instruction: assess state, understand state, and determine intervention.



Three Problems of Affect-Sensitive Instruction Represented as an Iterative Loop

Understand Affect State

Seven prominent theories of emotion were examined in an attempt to understand the function and purpose of emotions and to enumerate affective states that are relevant to the Submarine Learning Center learning environment: (1) Ekman's Basic Emotion Theory; (2) Russell's Circumplex Model of Affect; (3) Watson, Clark, and Tellegen's Positive and Negative Affect Schedule (PANAS); (4) Ortony, Clore, and Collins (OCC) Model of Emotions; (5) Pekrun's Concept of Academic Emotions; (6) Graesser and D'Mello's Learning Centered Emotions; and (7) Keltner's Consensual Taxonomy. For each of these theories, we provided a brief synopsis its theoretical constructs and taxonomy of affective states.

Emotion research is an active area of investigation, and researchers disagree on some issues. Nevertheless, researchers concur on some basic capabilities of emotions and the functions that they serve. In particular, four points of agreement have implications for emotion-recognition technology:

- Emotions arise through an unconscious appraisal process. Unless a person is consciously trying to deceive others, the emotion that he or she displays is the emotion that he or she is experiencing.
- Emotional experiences are brief and transitory. This implies that assessment or measurement of affect must occur in real time or be explicitly time linked.
- There are distinct and detectable cues for determining the emotions that someone is experiencing. Any particular emotion is brought about by a distinct set of antecedent conditions or events and has unique behavioral and physiological expressions.
- Humans are good at recognizing and interpreting the emotional cues experienced from other people. Researchers view this ability as an important basic skill in forming social relations.

We found little agreement among the theories in their enumeration of affective states. In the end, we identified 43 affective states that could be considered as candidate emotions for feedback from students. Not all these emotions could or should be monitored for any educational application. Whether an emotion should be monitored depends on (1) the accuracy of the technology that is used to detect the affective state, and (2) the effect that the affective state has on learning processes and outcomes.

Assess Affect State

We examined the state of affective computing and its application to synchronous distance learning for the Submarine Learning Center. Affective computing is generally defined as computing involving or arising from human emotion. We highlighted the major problem areas in affective computing and central research findings in the field. Our analyses were based on published research literature and interviews with 12 subject-matter experts who actively work in the field of affective computing.

We started with an examination of the three goals defined in the affect computing loop (see the figure). One of our interviewees, James Lester, offered his assessment of the technological maturity of the three problems, maintaining that the first goal (assess or recognize affect state) is the most mature, the third goal (determine learning intervention) is fairly mature, and the second goal (understand affect state) being the least mature.

We then examined whether emotion recognition should be based on a single mode of expression (unimodal) or more than one mode (multimodal). The multimodal approach is generally desired. The data suggest that any multichannel model that includes facial features with contextual cues (e.g., dialogue) is the best emotion-detection strategy. Note, however, there was almost unanimous agreement among our experts that if one were to select a single signal to use to detect affect, eye-tracking is the most effective method.

The next section considered the distinction between sensor-free and sensor-based approaches to affect detection. Sensor-free affect-detection research efforts focus on developing affect-detection techniques that recognize affect using non-sensor-based sources, such as log files of user interactions with a computer-based learning environment. Sensor-based affect detection relies on physical sensors (e.g., eye-tracker, facial-feature detector, EEG wearables, heart rate monitor, skin response, body-posture detector) to capture behavioral manifestations of emotion through physiological response, facial expression being the most popular in research.

Then we discussed the methodological consideration of establishing ground-truth measures of affect to build automatic emotion-detection platforms. There are three established methods for establishing ground truth of affect during learning: observation methods (e.g., the Baker-Rodrigo Observation Monitoring Protocol or BROMP), self-report methods, and log-file annotation (less popular). The BROMP method is perhaps the most accepted approach to determining the ground truth of the affective states of students while learning.

Finally, we discussed the extent to which commercial learning-management systems (LMSs) could support automatic affect recognition in a synchronous distance-learning environment. There are at least two examples where affect-detection systems based on student responses have been incorporated into LMSs. These systems suggest integrating rating processes of student interaction with the LMS (e.g., completing assignments on time) as a way to understand their affect with learning material.

Determine-Learning Intervention

With regard to affect-sensitive learning systems, researchers have used two general approaches to learning intervention. The first is proactive, in that learning systems are designed to induce positive emotional states or impede negative states before instruction begins. In comparison, reactive systems are those that detect and respond to affective states as they arise. Reactive systems have been further divided into those employing task-loop adaptivity, which focuses on selecting learning tasks or problems that are appropriate to the individual learner's states or traits and those employing step-loop adaptivity, which pertains to changes within a task or learning activity based on the learner's momentary state.

Proactive systems are particularly appropriate to the collaborative environment in synchronous distance learning because they focus on the similarities, rather than the differences, between students. For example, if student affect data identify course elements that are particularly confusing or frustrating to most or many students, the course would then be redesigned to remove those sources of confusion.

Reactive systems are more difficult to apply to synchronous distance learning because they call for learning adjustments at the individual student level. It is conceivable, however, that aspects of reactive systems could be incorporated into synchronous distance learning. For instance, suppose that an instructor has reason to believe that students are likely to be confused at a certain point in the lesson. The instructor could pose multiple-choice questions designed to probe those points of confusion, and the student responses could be displayed on an instructor dashboard. If there are only a few items that students answered incorrectly, the instructor could explain why those particular answers were incorrect, thereby removing the source of confusion.

Conclusions and Recommendations

We presented five conclusions concerning the theories and technologies of emotion recognition and affect-sensitive instruction:

1. Successful implementation requires solving three separate problems—assessing the affective state, understanding the affective state, and determining the learning intervention.
2. Emotion recognition is based on well-established science though some theoretical issues remain contentious.
3. The evidence is inconclusive about whether affect-sensitive instruction improves learning outcomes.
4. Student disengagement is not necessarily counterproductive.
5. Emotional traits, as well as states, are also relevant to learning.

We offered six specific recommendations for applying emotional-recognition technologies to the synchronous distance-learning environment used, or planned to be used, at the Submarine Learning Center:

1. Use multiple modes of affect detection.
2. Focus on most relevant and detectible emotions.
3. Use student audio and video feeds to detect emotions.
4. Use proactive approaches to design courses that induce positive emotions, impede negative emotions, or both.
5. Incorporate instructional features to enhance engagement.
6. Machine-learning models show promise, but have caveats.

Contents

1. Introduction

A. Problem

The Navy Education and Training Command (NETC) is responsible for the training, education, and professional development of the Navy’s 400,000 active duty and reserve sailors from accession and continuing throughout their careers. NETC is a global leader in rapid development and delivery of effective, leading-edge training for naval forces.

In its mission to prepare undersea warriors, NETC’s Submarine Learning Center (SLC) has a synchronous distance-learning system that employs virtual-world technology. This learning technology makes it possible to provide expert instruction to sailors located around the world, allowing them to interact in an engaging virtual learning environment. In this system, students are represented as avatars through which they can interact with each other and with virtual objects in a compelling 3D simulation. Although virtual-world technology provides audio and text communication between students and instructors, current systems do not provide nonverbal student feedback indicating the student’s emotional state to the instructor. Nonverbal feedback, such as facial expressions and body posture, can provide valuable cues about student engagement and comprehension. Such emotive cues are valuable information the instructor can use to adjust and adapt the pace and content of instruction to the students’ affective and cognitive states.

B. Report Objective and Organization

The emerging technology of automated affect recognition provides an innovative approach to providing nonverbal instructional feedback. Over the last 20 years, there have been significant advances in the capability of detecting students’ affect states as they learn. The objective of this report is to examine and assess the maturity of the science and technology of affect recognition as it relates to synchronous distance learning. This chapter provides a general overview of affect-recognition technology. Subsequent chapters examine how this technology could be applied to synchronous distance learning. The final chapter presents conclusions relevant to the SLC distance-learning environment and specific recommendations for implementing emotion-recognition technology in that environment.

C. Background

1. Affective Computing

New developments in human-computer interaction technology seek to close the communication gap between the human and the machine. A key component needed to meet these requirements is the ability of computer systems to assess user affect (emotions, moods, feelings). The knowledge of a user's affect can provide useful feedback regarding the degree to which a user's goals are being met, enabling dynamic and intelligent adaptation. This area of research, called affective computing, grew largely from Picard's (1997) highly influential book on the topic, which triggered a wave of research focusing on creating technologies that can monitor and appropriately respond to affective states of a user. Affective Computing is generally defined as computing involving or arising from human emotion (Calvo & D'Mello, 2010; D'Mello, Kappas, & Gratch, 2018; Picard 1997).

Picard's (1997) work on affective computing shaped the research space in this area for several years on key issues (at that time) in affective computing. In Picard's terms, affective computing is computing that relates to, arises from, or influences emotions. And in her work, she defines important issues related to affective computing, suggests models for affect recognition, and presents ideas for new applications of affective computing to computer-assisted learning, perceptual information retrieval, arts and entertainment, and human health and interaction. Picard discusses how, and shows that, affect plays a key role in understanding phenomena such as attention, memory, and aesthetics, and she goes on to support the idea that if computers are to interact naturally and intelligently with humans then computers need the ability to at least recognize affect.

Following Picard (1997), there has been considerable research on incorporating user affective states into the decision cycle of a computer interface (e.g., D'Mello & Graesser, 2010b). The inclusion of emotions into the decision cycle of computer interfaces is motivated by the complex interplay between cognition and affect, where cognitive activities such as causal reasoning, planning processes, and goal appraisal operate continually throughout the experience of affect (Baker et al., 2010; D'Mello & Graesser, 2010b; Russell, 2003). A computer interface that is sensitive to this complex interplay between cognition and affect is expected to be more usable, useful, naturalistic, social, and fun (D'Mello & Graesser, 2010b). For example, an affective-sensitive learning environment that detects and responds to students' frustration in the classroom is expected to increase motivation and improve learning gains compared with a system that ignores student affect (Calvo & D'Mello, 2010).

Affective computing, an interdisciplinary field of research owing much of its roots to decades of emotion research in psychology and cognitive science, is of interest to a number of fields, including machine learning, linguistics, computer vision, psychology, signal processing, education, and neuroscience. Affective computing relies on a computer's ability to reliably detect a user's affective state. This is a challenging problem because affect and emotions are psychological constructs with noisy and fuzzy boundaries, and they encompass neurobiological changes, physiological responses, bodily expression, and cognitive and metacognitive states (D'Mello, Kappas, & Gratch, 2018; Calvo & D'Mello, 2010). Adding to the affective computing challenge is the varying nature of affect from person to person and from context to context (e.g., D'Mello & Graesser, 2010b).

Researchers have made considerable advancements in affective computing as it relates to student engagement and learning goals. Traditional measures of student engagement in the classroom include self-report questionnaires, online observations, teacher ratings, and video recording (D'Mello, Dieterle, & Duckworth, 2017). More recent computational measures of engagement include affect-aware systems that measure fine-grained components of engagement in an automated fashion using two methods: (1) biometric sensors (e.g., eye tracking) and (2) semantically meaningful interactions (i.e., analysis of the way people use language) with the software being used (Baker & Ocumpaugh, 2015). The ultimate goal is to model the assumed link between internal affective state, engagement, and human behavior by means of machine-readable biometric signals or human-computer interactions (D'Mello, Kappas, & Gratch 2018; Calvo et al. 2015; Baker & Ocumpaugh, 2015).

2. Key Problem Areas in Affect Recognition

Technologies can support learning in a wide variety of contexts and sociocultural environments, and they are most useful when designed to meet specific needs and contexts of the learning community of interest (NASEM, 2018). Specifically, technology provides users *affordances*, or opportunities that a technology makes possible, related to learning and instruction. For example, distance learning has different features like text boxes, interactive dialogue boxes, and spoken messages, affording the user important learning opportunities (NASEM, 2018). There are eight identified key affordances of learning technologies: interactivity, adaptivity, feedback, choice, nonlinear access, linked representations, open-ended learner input, and communication with other people. Table 1 lists these affordances with their definitions.

Table 1. Key Affordances of Learning Technologies

Affordance	Description
Interactivity	The technology systematically responds to actions of the learner.
Adaptivity	The technology continually adapts information that is contingent on the behavior, knowledge, and characteristics of the learner.
Feedback	The technology gives feedback to the learner on the quality of the learner's performance, sometimes including how the quality could be improved.

Affordance	Description
Choice	The technology gives options for what to learn and how to learn.
Nonlinear Access	The technology allows the learner to select or receive learning activities that deviates from a set order.
Linked Representations	The technology provides quick connections between representations for a topic that emphasize different conceptual viewpoints, pedagogical strategies, and media, such as between spoken messages, texts, diagrams, videos, and interactive simulations.
Open-ended learner input	The technology allows learners to express themselves through natural language and other forms of open-ended communication that encourage active learning.
Communication with other people	The learner communicates with one or more other "persons" who may range from peers to subject-matter experts.

Source: NASEM (2018).

Not all the affordances listed in Table 1, such as the affordance of choice or nonlinear access, are applicable to the problem set at hand (i.e., SLC's unique problem set). But there are some specific affordances that align not only with type of learning conducted in an environment like that of the SLC, but also with the main goals of affective computing as it relates to learning. To accommodate the type of learning environment at the SLC, learning technologies that are engaged need to support deep learning (i.e., understanding complex concepts and systems and integrating information from a variety of inputs (NASEM, 2018); open-ended learner input and linked representations support this kind of learning. The quote below from the National Academies (2018) describes a deep-learning situation where different technology affordances are central to learning:

> The value of technology for representing a situation from multiple linked perspectives is evident in the example of helping learners understand a system, such as an electronic circuit. An intelligent technology can allow a learner quick access to perspectives, including a picture of the circuit as it appears in a device, a functional diagram of the components and connections, descriptions of the properties of each component, formulas that specify quantitative laws, explanations of device behavior, and the simulated behavior of the circuit as a whole when one component in the circuit is modified (p. 168).

There are three main objectives of affective computing and learning that are aligned with the key affordances of learning technologies (DeFalco et al., 2018; Woolf et al., 2009; D'Mello, Picard, & Graesser, 2007; Keith Brawner, personal communication; Jeanine DeFalco, personal communication; James Lester, personal communication). As depicted in Figure 1, the objectives comprise what is called the *affective loop*:

1. Affect recognition (interactivity; recognizing students' affective state).

2. Affect understanding (interactivity, adaptivity; recognize and adapt to different states appropriately with learning material).
3. Affect-instruction synchrony (feedback; quality and type of instruction affecting learning goals).

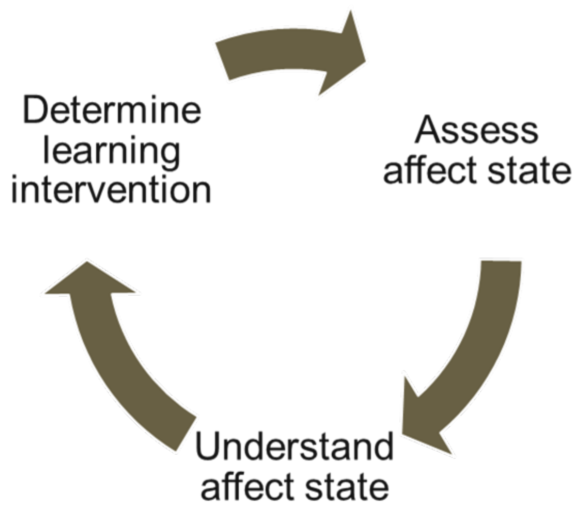


Figure 1. Affect Computing Loop

Table 2 highlights this loop with case studies of students' affective state and the subsequent tutor intervention based on this state.

Table 2. Case Studies of Affective States and Tutor Interventions

Cognitive Clues	Recognize Affect	Detect Need for Learning Intervention
Student makes an error	Student appears curious and focused	*No intervention needed*; student engaged in learning
	Student is frowning, fidgeting, looking around	*Alternative actions are needed*; student is confused
Student has not made progress	Evidence of stress, fidgeting, high arousal	*Alternative actions are needed*; student is under stress
	Evidence of boredom and confusion	*Interventions using off-task activities are needed*; student is not engaged
	Student is not frustrated	*No intervention needed*; student is curious and engaged in learning
Student is solving problems correctly	Student is not frustrated and is engaged	*No intervention needed*; student is in control, concentrated, and focused
	Student is bored—problems are too easy	Escalate challenge for a bored student

Source: Woolf et al. (2009, 135).

In essence, the three problems posed by the affective-computing loop for learning are to (1) dynamically collect cognitive and affective information from the user to assess affective state (affect recognition, interactivity), (2) detect a need for learning interventions (affect understanding, i.e., align detected affective state with need for learning intervention), and (3) determine which interventions are the most successful for individual students and contexts. The following chapters assess the current state of science and technology to address these three problem areas.

2. Assess Affect

This chapter describes the first goal identified in the affective loop—dynamically collect information to determine the user's affective state. Over the years there have been many approaches to computationally assessing affective state, and in this chapter, we attempt to outline the field and various ways one could approach assessing affective state. We begin with a description of unimodal and multimodal approaches to affect detection. This is followed by a discussion of how affect is detected using sensor-based and sensor-free approaches. Next, we explain the need to establish ground truth. Finally, we discuss how learning-management systems engage with this problem in learning environments. In addition, this chapter highlights the major problem areas in affective computing and the central research findings in the field, primarily focusing on assessing affective states. Although the findings in this chapter are based primarily on published documents in the open research literature, some of the findings were derived from interviews of 12 subject-matter experts (SMEs) who actively work in the field of affective computing. Appendix A provides details of those interviews.

A. Affective Computing Loop Methodology

This subsection briefly highlights some methodological approaches to assessing affective states. Note that the three main objectives of affective computing as it relates to learning are difficult to achieve, and each piece of the loop is actively being explored in current research (see Chapter 1 for overview of loop). Specifically, this subsection discusses two methodological approaches. The first is by Woolf et al. (2009), who uses human coders, sensors, and machine-learning approaches to detect affective state (step 1 of the loop). The second methodological approach is AutoTutor by D'Mello et al. (2010), who uses spoken dialogue to address each aspect of the loop, but here, we will focus on assessing affective state.

Woolf et al. (2009) explore three options of data collection that are generally adopted by a wide variety of researchers in affective computing to detect affect: human coders, sensors, and machine-learning approaches. Data collection via human coders is the most time-intensive approach, but it can be more nuanced than other data-collection techniques. It is the approach adopted first by Woolf et al. (2009) in an observation experiment with 34 students that used the intelligent-tutoring software Wayang Outpost (described later in this chapter; Arroyo et al., 2009) over a 3-week period for mathematics. Human observers (i.e., coders) coded one student at a time in a classroom, each observation period lasting 15–20 seconds per student. To gain an understanding and assessment of a students' affective state, coders looked for expressed affect and recorded facial expressions, physical expression, and verbal behavior, and whether the student was on or off task. Since distinguishing different affective states is difficult, coders tried to identify valence (positive/negative emotional energy) and arousal (physical activity). Results from coders noted 58% of students being in a preferred state of high valence and low arousal (i.e., concentrated). The second way they explored collecting affective state information was by using hardware sensors—specifically, a facial-expression system, posture-analysis seat, pressure mouse, and wireless skin-conductance sensor, combined in an in-house-developed platform. These sensors were tested on 100 students, and subsequent data analyses were able to predict 60% of the variance of a student's emotional state. Using all sensor data, the accuracy varied for different emotions; frustration was the most accurate (89%). Finally, the researchers also explored machine-learning techniques (e.g., Bayesian networks) to recognize and assess affect. Depending on the data set, context, and technique, these techniques can account for student affect with up to 75% accuracy. The three approaches (human coding, sensors, and machine learning) all highlight different methodological approaches to assessing affective state.

The second highlighted methodological approach to assessing affective state and also closing the affective loop is Affective AutoTutor (D'Mello et al., 2012; they try to implement each component of the loop by using interactive dialogue). The original AutoTutor (which lacks affect detection) is an intelligent tutoring system that interacts with a learner via text conversation and is sensitive to a student's affective state. The current AutoTutor is a fully automated tutor for Newtonian physics, computer literacy, and critical thinking (Graesser et al., 2004; Graesser et al., 2005). AutoTutor is dialogue based, meaning that users need to articulate three- to seven-sentence expressions in response to challenging questions regarding the above topics. AutoTutor encourages dialogue by taking turns with the user and providing feedback (e.g., "good job"; "not quite"), giving hints, correcting misconceptions, and answering questions.

Affective AutoTutor enhances the original intelligent tutor by being able to recognize and actively monitor the presence of boredom, confusion, and frustration. It attempts to alter these states with responses that are empathetic, encouraging, and motivational, and ultimately intervene and suggest alternative directions in learning. The effectiveness of recognizing these more negative emotions and attempting to course-correct a student was studied by D'Mello and Graesser (2010b), who showed that for lower domain knowledge learners, the Affective AutoTutor was more helpful than for a high-domain knowledge learner (i.e., certain students need learning interventions based on their affective state, but others do not); however, across learning sessions, the Affective AutoTutor showed learning gains higher than using the non-affective tutor. This is evidence that recognizing (and attending to) affective states (i.e., closing the affective loop) during learning is important for positive changes in perception toward learning material.

Other researchers have tried tackling these issues using a single sensor, such as eye tracking; using a combination of sensors, such as eye tracking and facial recognition; or using no sensors, focusing instead on dialogue or interaction logs with a system. The following subsections continue to explore affect-recognition approaches, beginning first by addressing the difference between unimodal and multimodal affect recognition, and then addressing sensor-based and sensor-free affect recognition. Chapters 3 and 4 address the other two steps of the loop.

B. Unimodal and Multimodal Detection

Affective states are typically accompanied by a degree of physiological arousal detectable from facial expressions, acoustic-prosodic cues (i.e., stress and intonation patterns in spoken dialogue), body movements, gesture, contextual cues, text and discourse, and measures of peripheral and central physiology (e.g., eye-tracking, heart rate, galvanic skin response) (D'Mello, 2013). Much of the first decade or so of computational affect-recognition focused on a *single modality* to detect affect, such as facial expressions or intonation patterns of speech. This approach is termed "unimodal affect" (D'Mello & Graesser, 2010b; Poria et al. 2017). Unimodal affect recognition has three significant problems identified in the literature. First, it's unclear if all emotions, for example boredom or engagement, can be expressed and reliably detected via facial expressions or patterns of speech (D'Mello & Graesser, 2010b; Craig et al., 2008). Second, users can control and deceive via facial expressions and speech (D'Mello & Graesser, 2010b; Jonathan Gratch, personal communication, November 2010). Last, naturalistic emotional expression is rarely unimodal. From this tradition, researchers began fusing, or combining, different affect-appropriate signals in an effort to understand which combinations are strongest at classifying (i.e., assessing) affective state. This approach is termed "multimodal affect." The next two subsections look at unimodal affect recognition and multimodal affect recognition in turn.

1. Unimodal Affect Recognition

Unimodal approaches to assessing affective state have also been used by a number of researchers. Studies have shown that dialogue data and facial-recognition data showed evidence for being strong enough signals to capture an affective state alone. Facial expressions are the most common signal to explore for unimodal recognition and will be primarily discussed in this subsection.

Research enterprises have focused on solely using facial expressions for automated feedback in teaching. As the field of machine learning continues to advance and computational power continues to increase, techniques like pattern recognition are strong approaches to detecting affect signals in single-source data, like facial expressions. For example, Whitehill, Bartlett, and Movellan (2008) explored affect feedback based on automatic recognition of a student's facial expression in an effort to measure perceived material difficulty and to determine preferred speed of lesson material presentation. In their pilot study, eight participants viewed various video lectures (e.g., an introductory physics lecture, a university lecture on philosophy) at adjustable speeds while their facial expressions were automatically recorded. Participants also rated lectures based on their difficulty and took six quizzes. The machine-learning-based, automatic facial-expression-recognition system analyzed facial expression using FACS (Ekman & Friesen, 1978; see Chapter 3), identifying the eyes, mouth, and nose. Results show a high degree of inter-subject variability with regards to which facial muscles (action units or AUs) correlated with difficultly and viewing speed, the only consistent correlation across participants was eye blink where difficult sections of the lectures had lower blink rates (difficulty or cognitive load is typically associated with lower rates of blinking). Due to the high inter-subject variability in which facial movements correlated with difficult and viewing speed, Whitehill, Bartlett, and Movellan (2008) suggest that machine learning models should be subject specific models trained on an individual's specific facial expression in order to be useful in predicting perceived difficulty (also echoed across many interviews IDA conducted with experiments in this field, specifically, Shri Narayan and Brandon Booth).

Which signal is best for *unimodal* analysis? Historically, the psychological literature regarding facial signals as strong indicators of affect has two camps: proponents of basic emotions (e.g., Ekman and Friesen, 1978) who support that facial features of basic emotions are innate, universal, and cross-cultural and opponents who suggest that emotional expression are always modulated by context and might be understood better via valance-arousal models. Critically, though, D'Mello and Graesser (2010b) showed that emotions can be expressed with or without facial cues and that the use of facial cues as a signal for affect is depending on the type of judgment being conducted (i.e., signal × judgment type interaction); for fixed judgments (i.e., static), dialogue is the best signal of affect whereas for spontaneous judgments (i.e., fluid), facial movements are the best. Whereas Whitehill, Bartlett, and Movellan (2008) showed promise in just using facial expressions, but ultimately advocated for subject-specific models in order to more accurately predict affect.

Based on IDA's interviews conducted with SMEs in the field of affective computing, there was almost unanimous agreement that if one were to select a signal to use to detect affect, eye-tracking is the most effective method. Eye movements and pupil dilation (i.e., pupillometry) are considered strong indicators of cognitive processes and show visual attention and mind wandering. This has been confirmed in a number of studies across several fields. Specifically, via pupil dilation one could assess cognitive load of a student, but more simply, seeing where a student is looking on a screen is indicative of attention (e.g., they're looking at and interacting with targeted information vs. eyes wandering around the screen not looking at material). While eye movements are generally a good measure of attention, it should be noted that people interact and attend to information differently. For example, someone might look off screen when they're thinking about and integrating information during learning, this eye-wandering might indicate someone isn't paying attention to work when in fact it's the opposite (Jonathan Gratch, Mohammad Soleyami, personal communication). A proxy for eye movements as indicative of attention is mouse-tracking. Mouse-tracking is a technique used to monitor where students are moving their mouse and clicking on material. While this doesn't capture affective state, it does capture attention and engagement with material, arguably more important than affective state depending on the student population (Ben Nye, Jonathan Gratch, Mohammad Soleyami, personal communication). Finally, as indicated in D'Mello and Graesser (2010b), dialogue (text) is also a powerful signal.

2. Multimodal Affect Recognition

Taking a multimodal approach to affect detection requires the integration of multiple affect-appropriate signals into one multisensory emotion classification system (i.e., implementing machine learning techniques). The main hypotheses surrounding multichannel affect detectors concern *super-additivity*, *additivity*, *redundancy*, and *inhibitory* effects (D'Mello & Graesser, 2010b). *Super-additivity* means that classification performance (e.g., classifying a user as confused or frustrated) from multiple channels (e.g. eye tracking and facial features) will be superior to just an additive combination of individual affect signals. *Additivity* is where the performance of multiple channels is equivalent to an additive combination of individual channels. *Redundancy* means there are negliable gains to combing multiple affect signals together (i.e., multimodal recognition shows no improvement to unimodal detection). Finally, the last hypothesis concerning multimodal affect recognition concerns *inhibitory effects*, which is where combing multiple channels results in substantially lower classification rates (i.e., multimodal recognition performs *worse* to unimodal detection). These hypotheses regarding multimodal recognition are addressed below in D'Mello and Graesser (2010b).

D'Mello and Graesser (2010b) tested a multimodal classification detector of boredom, engagement/flow, confusion, frustration and delight (i.e., learning centered emotions) by using AutoTutor. Data was collected from 28 participants randomly assigned to topics in computer literacy (hardware, internet, or operating systems), three streams of information were recorded during the participant's interaction with AutoTutor: the participant's face, posture patterns, and audio and video of the participant's entire tutoring session. In order to appropriately code the affective states of participants, two trained judges independently coded the learners' facial features using the Facial Action Coding System (FACS). As discussed in Chapter 3, FACS was developed by Ekman and Friesen (1978). Judgments were made in two ways: (1) *fixed* (asynchronous, static, freeze-framed) judgments based on video streams of the participant's face captured at 20 s intervals and

(2) *spontaneous* judgments where judges can pause the video whenever they chose to judge the affective state of the participant.

Looking at multimodal streams of data, the trained judges in D'Mello and Graesser (2010b) found that the most common judgment type for the *fixed* 20 s judgments was neutral (i.e., no emotion), followed by confusion, engagement/flow, and boredom. The best multichannel model using fixed judgements includes two streams of data—facial expression and dialogue. *Spontaneous judgments* had a different distribution; the most prominent affective state was confusion, followed by frustration, delight, and boredom. Confusion is prominently detected for both types of judgment, fixed and spontaneous. The best multichannel model for *spontaneous* judgments was similar to *fixed* judgements in that facial features and dialogue were the best multimodal stream of data. One difference is that posture can also aid in assessing affective state when using *spontaneous* judgments in the creation of the classification model.

Looking at each stream of data independently (i.e., unimodal analysis), D'Mello and Grasser (2010b) found differing results, depending on the type of judgment done to collect the data, that is, fixed or spontaneous. For *fixed* affect judgments, facial features did not provide sufficient cues to discriminate subtle emotional expressions. In fact, the dialogue data (i.e., log file data with AutoTutor) had the highest agreement between coders in fixed judgments. For *spontaneous* judgments, the best single indicator of affect state was facial features (exactly opposite that of fixed judgments). Therefore, there is a signal × judgment-type interaction for determining the best single signal for affect recognition.

To summarize, it seems that any multichannel model that includes facial features with contextual cues (e.g., dialogue) is the best emotion-detection strategy, similar to previous findings on multimodal affect recognition (e.g., Arroyo et al., 2009). To answer the question about how significant fusing different data streams for affect recognition is, it seems that a multimodal channel is *super-additive* (i.e., each feature by itself is not capable of detecting affect alone, whereas together the fused signal is more accurate).[1]

As the field continues to advance and machine-learning techniques are implemented at the person-level, unimodal approaches to affect recognition are not out of the question. In fact, unimodal systems act as primary building blocks for well-performing multimodal frameworks, therefore, research in this area will continue to flourish and form the foundation for more sophisticated affect-recognition models (Poria et al., 2017). However,

---

[1] There are more significant details of the study showing signal differences between fixed and spontaneous data being used in a classifier. The scope of these details is outside the focus of the current work, but note that while a multimodal approach to assessing affect shows advantages over unimodal approaches (depending on the combination of signals), unimodal approaches also show promising detection abilities.

incorporating other sources of information (log information, student performance, dialogue) only provides a better picture of student engagement in the classroom.

C. Sensor-Free and Sensor-Based Affect Detection

The previous section looked at which data streams are the most useful in assessing affective state. Another, not mutually exclusive approach, divides affect recognition into sensor-free detection and sensor-based detection. Sensor-free affect-detection research efforts focus on developing affect-detection techniques that recognize affect from user log files of user interaction with a computer-based learning environment. Sensor-based affect detection relies on physical sensors (e.g., eye-tracker, facial-feature detector, EEG wearables, heart-rate monitor, skin response, body-posture detector) to capture behavioral manifestations of emotion via physiological response, facial expression being the most popular in research (Calvo & D'Mello, 2010). For any affect channel, there are advantages and disadvantages in using the specific affect channel as an appropriate signal (listed in Calvo and D'Mello, 2010):

1. Validity of the signal as a natural way to identify an affective state.
2. Reliability of the signals in real-world environments.
3. Time resolution of the signal in real-world environments.
4. Cost and intrusiveness for the user.

This section discusses important research for sensor-free and sensor-based approaches.

1. Sensor-Free Affect Detection

This subsection focuses on describing major research paradigms on unobtrusive measures of measuring affect. In many domains of affect recognition, researchers have relied on physical sensors (e.g., wearables, heart-rate monitor); however, not every educational setting is capable of using physical sensors (e.g., because of financial restrictions, classified environments). As a result, there is continued effort and interest in developing affect detectors that rely on interaction data between the student and computer (Baker & Ocumpaugh, 2015). This approach aims to detect learner behaviors associated with engagement and affect by inferring patterns of student behavior based on interaction with education software only (DeFalco et al., 2018; Baker & Ocumpaugh, 2015). The selection of intelligent tutoring systems below is taken from Baker and Ocumpaugh (2015), who provide a succinct, yet comprehensive review of sensor-free systems and how collected data are used in machine-learning algorithms. To highlight the flavor of systems developed for sensor-free affect detection, only a small subset of those reported in Baker and Ocumpaugh (2015) are presented here. For example, Crystal Island, an important system, is presented in Chapter 4.

a. Why-2Atlas

Why-2Atlas is a *text-based* qualitative physics tutor that teaches students physics by having them write paragraph-long explanations of simple mechanical phenomena. The tutor uses deep syntactic analysis and abductive reasoning, eventually converting the student's paragraph to a physics proof. The proof allows the tutor to uncover misconceptions as well as to detect missing correct parts of the explanation (Baker & Ocumpaugh, 2015; Ai et al., 2006). The idea is that dialogue systems use emotion detection to discover problematic points in learning in written and spoken dialogue (lexical and prosodic features). Ai et al. (2006) used ITSPOKE (a spoken-dialogue tutor built on top of the Why2-Atlas text-based tutoring system) to incorporate automatically obtained system/user performance features into machine-learning experiments to detect student emotion based on tutoring sessions with 20 students. As mentioned above, a student is prompted to write a paragraph on a conceptual physics question. Then, after analyzing the essay, the ITSPOKE talks through misconceptions with the student. Finally, the student revises the essay and resubmits it, ending the tutoring session. The machine-learning algorithm used lexical and prosodic features in addition to gender, student ID, problem ID, turn sequence, and other performance measures to predict emotion conveyed in each student turn. The model using all of these features show a classification improvement of 8.08% over the baseline (i.e., without using any features like lexical prosody) with an accuracy of 59.41% in determining how certain students are of their answers.

b. AutoTutor

AutoTutor has been discussed many times already (i.e., Nye, Graesser, & Hu, 2014), but due to its central role in the development of affect-recognition intelligent tutors, it deserves its own subsection of review. AutoTutor is a family of intelligent-tutoring systems that share the same theoretical principles and features of AutoTutor's design (e.g., natural-language processing algorithms, conversational agents). Some of systems in this "family" include AutoTutor, Affective AutoTutor (D'Mello & Graesser, 2012), and Gaze Tutor (D'Mello et al., 2012). AutoTutor systems have also recently been granted a U.S. patent (Graesser & D'Mello, 2019).

AutoTutor is a natural-language-based tutoring system developed at The Tutoring Research Group at the University of Memphis. The intelligent tutor uses an animated talking head to help students learn skills in computer literacy, physics, and critical thinking by holding conversations with students. The disciplines that AutoTutor focuses on force students to use deep-learning skills, since mastering difficult technical material requires students to confront difficult concepts, anomalous events, and obstacles (D'Mello, Picard, & Graesser, 2007; Nye, Graesser, & Hu, 2014). The dialogue-based tutoring system consists of various subtopics within each main topic area; the animated tutor has a certain set of expectations that a student needs to fulfill (e.g., number of dialogue moves, corrections of misconceptions, other conversational cues and exchanges) to move to the next subtopic question. AutoTutor maintains a log file capturing the student's response, assessments of the conceptual quality of the response, the feedback provided, and the tutor's next move. AutoTutor provides feedback on what a student types (positive, negative, neutral), pumps the student for more information (e.g., "What else?"), prompts the student to fill in missing words, gives hints, fills in missing information with assertions, identifies and corrects misconceptions, answers questions, and summarizes topics. A full answer to a question normally takes between 30 and 200 turns. AutoTutor compares student responses with a curriculum script on the topic and uses Latent Semantic Analysis (a statistical technique that measures conceptual similarity between texts) to judge whether the response is expected or misconstrued. The AutoTutor discourse moves are summarized with examples in Table 3 (taken from Nye, Graesser, & Hu, 2014).

Table 3. Examples of AutoTutor Discourse Moves

Move Type	Description	Example(s)
Main Question	A question that starts off the dialogue, focused on a particular topic or goal	"If the man drops his keys just as the elevator falls, how do the objects move relative to each other? Explain why."
Pump	Asking the student to provide more information.	"Anything else?"
Hint	Leading question or statement that attempts to direct the user to answer the main question	"What do you think about the gravitational force on this object?"
Prompt	Leading the student to express a missing word from an important idea for the main question.	"The force on the objects from gravity acts in which direction?
Short feedback	Signaling about the quality of the student's last statement.	"Great!" (Positive), "Okay." (Neutral), "Not quite." (Negative)
Correction	Correcting a misconception or incorrect statement by the learner.	"No, the force of gravity on both objects is equal." (After student claims one is greater)
Assertion	Presenting an important idea within the problem or the answer to the problem.	"The force of gravity on both objects is equal."
Answer	Response to a learner's question about the definition of a concept.	"A vector is a quantity with both a magnitude and a direction." (In response to "What is a vector?")
Summary	Presents the full answer to the main question or problem.	"The magnitude of the force of gravity on each object is equal, and all force vectors point down…"

Over the years, more than a dozen experiments using AutoTutor have been conducted, comparing AutoTutor tutoring treatments with different types of approaches to computer-based training (e.g., students not using a tutoring system). Results consistently show that AutoTutor improves student learning and shows significant learning gains compared to a pretest—on average about a .08 standard deviation learning gain above controls who read static materials (Nye, Graesser, & Hu, 2014).

c. Prime Climb

Prime Climb (Conati & Maclaren, 2009) is a game-based learning system for mathematics with an affective agent for sixth- and seventh-grade students. Unlike AutoTutor and Why-2Atlas, Prime Climb does not require text dialogue, but instead a numerical based response from a student. Students were rewarded or penalized based on success of response (e.g., if a student responds incorrectly, it will cause the student to fall down the mountain). Conati and Maclaren (2009) rely on the Ortony, Clore, and Collins (OCC) theory to develop the affective model employed in Prime Climb. In Conati and Maclaren's (2009) research, more than 60 sixth and seventh graders used Prime Climb for approximately 10 minutes to learn about factorization; students were encouraged to indicate their emotional state on a slider-interface. The Prime Climb agent would intervene to help students during the game. Results show that the agent was moderately able to predict joy from distress (32% change), but when the data were re-validated, these results disappeared. The results highlight not only that it is difficult to create an accurate affective computing loop but also that it is crucial that data be tailored to the learning goals at hand.

d. Wayang Outpost

Arroyo et al. (2009) used Wayang Outpost, a multimedia adaptive-tutoring system for geometry and other mathematics topics typical on the SAT, to explore the use of sensors in intelligent tutors to detect students' affective states and also provide emotional support for students. Wayang Outpost is a web-based tutoring system that helps students learn to solve math problems typical of those on achievement tests. Since it is a web-based program, students log into the site and begin a session with receiving a problem. Each math problem is presented as a flash movie with decisions and hints made by the intelligent tutor. Data were collected for two different studies using four sensors (mouse, posture chair, video camera, skin-conductance bracelet), and questionnaire data were collected by probing students' affective states while they used Wayang Outpost. While the study uses sensors (sensor-based), some of the emotion results focused on questionnaire data (sensor-free). The ultimate objectives were to see if students' affective states influence their learning, motivation, and attitudes toward math, and to trace students' emotional states in a real-world classroom. The two experiments independently looked at high school students (n = 38) and undergraduates (n = 29) who used Wayang Outpost for 4–5 days and took math tests and a survey about their perception of math before and after using the software. During use, the system iterated through different topics and problems, adapting to student performance; the system also prompted students every 5 minutes or after every question about how they felt. Post-test results show that students improved about 10% in math performance. Further, their self-reports of emotion showed that the emotion depended on the event that occurred in the previous problem and not on incoming beliefs of ability or motivation.

2. Sensor-Based Affect Detection Using Biometric Sensors

Picard's (1997) influential work supports the idea that if computers are to interact naturally and intelligently with humans they need the ability to at least recognize affect; specifically, wearable computers that can perceive physiological information can gather powerful data for advancing results in cognitive and emotion theory. Research on affective computing has used a wide range of data streams to detect affect, many researchers focusing on physical sensors because physical sensors are able to capture physiological manifestations of affect (DeFalco et al., 2018). This sensor-based affect detection has been developed using physical sensors to detect affect via facial expressions (Arroyo et al., 2009; Bosch et al., 2015), voice (Lee et al., 2015; Lee & Narayanan, 2005), posture (D'Mello & Graesser, 2010b), and galvanic skin response (Arroyo et al, 2009).

Facial expressions are the most popular signal for detecting emotions, but not necessarily the best. One needs trained coders to identify facial actions associated with affective states (a costly and time-consuming pursuit), and many automatic systems rely on acted facial expressions (i.e., not naturalistic). Other popular signals for affect include voice (speech features). The literature supports the idea that affective information is encoded in speech patterns—sadness, anger, and fear are best recognized through voice (Calvo & D'Mello, 2010). Speech data are also often collected in real-life settings (e.g., tutoring sessions or call center logs), these are a richer, more accurate data source for affect models.

Arroyo et al. (2009), which is discussed in the context of Wayang Outpost in the previous section, employed a number of different sensors. The researchers focused on two data-collection techniques, one being sensor-free, wherein students were queried about their affective state. The other data-collection technique was sensor based, where four sensors (mouse, posture chair, video camera, skin-conductance bracelet) were used to determine the extent of the benefit of using sensor data to detect students' emotions. The results from the studies showed that emotion can be predicted from what happened in the previous problem. In addition, the regression models built by analyzing the contribution of each individual sensor show that the camera sensor accounts for 52% of the variance and helped predict confidence, excitement, and being interested. The seat sensor significantly sensed frustration—similar to what D'Mello & Graesser (2010a) found (see below) and accounted for 68% of the variance (i.e., posture was very helpful in predicting student affect as defined in this study).

Turning to another sensor, D'Mello and Graesser (2010a) investigated postural patterns associated with naturally occurring episodes of boredom, flow/engagement, confusion, frustration, and delight during a tutoring session with AutoTutor. Posture is an interesting signal for affect detection that is often overlooked. Human bodies are large, and movements are usually unconscious and unintentional and therefore are less susceptible to control than facial expressions or prosodic features. D'Mello and Graesser focused on data mining, using data collected from 28 students using AutoTutor to answer *how* the body conveys affect through posture articulations. Recall that in D'Mello and Graesser (2010b), student affect was measured via retrospective judgment from learners themselves (self-report), peers, and judges (i.e., affect judged after the fact) in two ways, *fixed* at 20 s intervals, and *spontaneous* at any time point between the 20 s intervals.

Here, Body Pressure Measurement Systems, an automated system developed by Tekscan, was used to capture the pressure exerted by the learner on the seat. For the purposes of looking at body posture as a signal for affect, D'Mello and Graesser (2010a) relied on the *affective-arousal* framework in which heightened pressure in a seat relates to positioning one's body toward the source of stimulation (i.e., high attentiveness), whereas an increase in pressure on the back of a seat suggests someone is detaching from the stimulus (i.e., low attentiveness). The different pressure-related features of body movements were computed by examining the pressure map during an emotional judgment, and these pressure features were associated with an emotional category based on the human judges' affect ratings. Logistic-regression analyses were used to systematically explore relationships between posture features and affective states, specifically, to distinguish between each affective state (boredom, confusion, delight, flow, frustration) and neutral.

Results indicate posture features explained about 11% of the variance in discriminating affective states from neutral on average. Based on logistic-regression results on the change in seat pressure for each affective state, boredom shows significant body disengagement indicated by an increase in pressure in the back of the seat compared with a neutral affective state. In addition, boredom also shows an increased rate of change in seat pressure movement, indicating fidgeting. Delight and flow showed increased attentiveness by learners leaning forward in a seat, and confusion and frustration states also had learners leaning forward, but more upright in posture. In summary, D'Mello and Graesser (2010a) discovered relationships between body position and learners' affective state. Boredom is associated with leaning back; delight and flow are associated with leaning forward; confusion and frustration are also associated with leaning forward, but in a more alert, upright position.

Adding to body posture, Bosch et al. (2015) used cameras to capture facial expressions to assess affect. For the most part, affect-sensitive intelligent tutoring systems have been largely developed in controlled experimental settings, devoid of distractions (see also Arroyo et al., 2009). Bosch et al. (2015) investigated affect detection, not in an experimental setting, but in a classroom computer lab, using facial expression and body posture. Videos of 137 students' faces, affect labels, and labels of on-task, off-task behavior were collected while students interacted with Physics Playground, a game-based physics-education environment, using computers equipped with a webcam. Physics Playground is a two-dimensional game that requires the player to relate physics principles to different challenges (e.g., guiding a green ball to a red balloon). The affect states of interest were learning-centered emotions—boredom, confusion, delight, engaged concentration, and frustration—monitored by trained observers who used the Baker-Rodrigo Observation Protocol (BROMP) labeling method to record students' affect (Ocumpaugh, Baker, & Rodrigo 2015). These labels, which served as ground truth of affect, were used to train the automatic affect detectors along with facial expression data, which was collected using FACET, a commercial computer vision software for facial feature extraction based FACS. Bosch et al. (2015) built classification models to detect overall five-way discrimination between bored, confused, delighted, engaged, and frustrated, in addition to a separate detector for each affective state. Ultimately, the authors showed that machine-learning models can be developed and validated to learn student affect based on facial expression recorded from a camera signal in a noisy school environment. Specifically, the best models for each classification showed that the overall five-way discrimination between all affective states performed above chance, and each individual detector for each affective state (e.g., frustration vs. delight) all performed above chance.

One significant drawback to using physical sensors is that these kinds of instruments are usually applied in controlled environments and not in real time. Some researchers have also noted that biometric instruments can negatively influence users' affective state if they feel they are being monitored or become frustrated with the biometric sensor and not the learning material at hand.

From our expert interviews, the best biometric sensors to use are eye-tracking devices, mouse tracking (a proxy for eye tracking), and some sort of speech/text recognition. While multi-modal channels can be stronger (as discussed previously), if one needs to select a single sensor (due to environment, learning situation, or budget constraints), these three are the strongest alone.

D. Establishing Ground Truth

When developing automatic affect-detection programs, programmers must consider the methodological issue of how to establish ground-truth measures of affect to build such detection platforms. Our review of the affective research studies in this report shows that the quality of results obtained by such studies (i.e., obtained by affect-sensitive models) is inextricably linked to the quality and contextual nature of the data that are collected to build such models (Baker & Ocumpaugh, 2015). The data that are used to build affect-sensitive systems must accurately reflect the underlying construct and also be related to the predictive goals of the model. For example, collecting ground-truth data on students' affective states while watching television and then using these data to build a predictive model for students' affective state in the classroom will lead to undesirable results because the ground-truth data do not match the use case. Further, most of the subject-matter experts the IDA team interviewed noted that data collected for one learning task are not generalizable to other learning contexts; that is, learning and learning environments are context specific with different goals and incentives for performance.

There are three existing methods for establishing ground truth of affect during learning: observation methods (e.g., BROMP), self-report methods, and log-file annotation (less popular). Observation methods, like BROMP or FACS, can be conducted live or using video (e.g., D'Mello & Graesser 2010b), but there are differences between the approaches. For example, observing video data to determine ground-truth affect is usually more definitive, but has lower interrater reliability.

BROMP, an observation-based ground truth method, is an objective coding paradigm where trained field observers repeatedly observe students in a predetermined order (to avoid biases toward interesting activities); field observers observe one student at a time for up to 20 seconds and record the first affect the student displays. The observers are trained to look holistically for a range of behaviors, including physical and verbal demonstrations of affect (Baker & Ocumpaugh (2015); Ocumpaugh, Baker, & Rodrigo, 2015). In-person observations might have limitations depending on the learning environment, and in particular, Baker and Ocumpaugh (2015) note that conducting this kind of coding is most effective if the field observers are drawn from approximately the same cultural background as the students. BROMP has been employed in a large number of studies with success for establishing ground truth for affective state (e.g., DeFalco et al., 2018).

The second method for establishing ground truth is to obtain data via self-reports either in the form of an emote-aloud procedure or a questionnaire (Baker & Ocumpaugh, 2015). An emote-aloud procedure is where a person (e.g., participant, student) verbalizes his or her affective states during learning after being instructed on what each affective state represents (e.g., emote "anger" is to have a strong feeling of displeasure). For example, in Craig et al. (2008), an emote-aloud procedure was used to discover the facial action units that were present during learning-centered affective states (i.e., anger, boredom, confusion, contempt, curiosity, disgust, eureka, and frustration). In this study, seven undergraduates used AutoTutor and provided verbalized affective states whenever they experienced one. The results show that the emote-aloud methodology helped pinpoint at what point during learning affective states occurred.

Questionnaires are an emote-aloud procedure that can be used to gain ground-truth affective state data. Arroyo et al. (2009) used this method. When using the tutoring system, students were queried every 5 minutes and after they finished a problem: "How [interested/excited/confident/frustrated] do you feel right now?" Conati and Maclaren

(2009) also collected student affective states during interaction with Prime Climb, using a slider-based interface to prompt students about their affective and motivational states.

There are several shortcomings of the emote-aloud methods. Irregularly interrupting students can change the student's affect in a way unrelated to the learning content. It can also place extra cognitive demands on the student. Finally, if the students are to verbalize emotion, doing so can be disruptive in a classroom setting with many students (Baker & Ocumpaugh, 2015).

E. Learning Management Systems and Synchronous Learning

The discussion so far has focused on *automated* synchronous tutoring systems, where the tutor is usually some sort of talking head or avatar character, The reason for this focus is that the research done in affective computing and learning has not only focused primarily on automated systems but also served as the foundation for other forms of education and distance-learning research, for example, research on Learning Management Systems (LMSs), which draws on work from Picard (1997) and the D'Mello and Graesser enterprise. LMS research grapples with similar issues to automated affective tutors—motivation, engagement, and student affect (e.g., Farman Ali Khan et al., 2009; Rodrigues, Fdez-Riverola, & Novais, 2011).

An LMS is an online learning platform for learners and instructors. An LMS typically includes discussion forums, class content like homework or lecture notes, and creation of learning content. LMSs, such as Moodle (Modular object-oriented dynamic environment) and Blackboard, are very successful in electronic and distance education, but do not fully support or accommodate adaptivity for synchronous learning. While LMSs have been around since the mid 90s, it is only relatively recent that certain LMSs have adopted the capability of collaboration and synchronous teaching. For example, Blackboard Inc. acquired Elluminate, Inc. in 2010 forming Blackboard Collaborate, which provides a real-time, synchronous virtual classroom. In today's climate, video-chat systems like Zoom and Google's G-suite also provide synchronous virtual classrooms, but they lack the classroom content organization of an LMS.

Some researchers have attempted to apply the concepts and approaches for detecting affect employed in automatic affect recognition to LMSs. Khan et al. (2009) presented an approach that investigates patterns of behavior in LMSs that correspond to students' different affective states to provide students with more individualized support with LMSs. Interestingly, due to the nature of LMS systems (i.e., students navigate the system, use it for syllabi, discussion forums etc.), the learning-centered emotions commonly detected in affect computing are not applicable. Instead, Khan et al. (2009) looked at only four affective states, *confidence*, *effort*, *independence*, and *confusion*, which really represent characteristics of how students relate to commonly used features of LMSs. Since LMSs are meant to be used differently than intelligent tutoring systems, relevant affective states need to be appropriately defined for LMS use. In this regard, affective states and behavior are primarily defined by the ways in which a student interacts with an LMS. For example, academic confidence is measured by factors of studying (e.g., visiting content, making outlines), understanding (e.g., attempting exercises), verbalizing information (e.g., discussion groups), clarifying (e.g., visiting assignments), and attendance (e.g., counting discussion posts, replying to others). Effort in terms of LMSs is defined as attempting a high number of self-assessment tests and exercises with a high number of correct answers, visiting a high number of postings related to content, and submitting assignments before they are due. Independence is defined as having intentional behavior with learning material, for example, planning, organizing, monitoring, and evaluating one's performance. Finally, confusion in an LMS is defined by students performing a low number of self-assessment tests and exercises, leaving high numbers of questions unanswered, spending more time on content, and interacting with discussions forums only to inquire about how work is to be completed (Khan et al. 2009).

Rodrigues, Fdez-Riverola, and Novais (2011) have also attempted to design an affect-sensitive LMS approach using Moodle. One of the most popular LMSs used worldwide, Moodle is primarily module driven in that each aspect of the LMS is interactive (e.g., forums, chats, quizzes) and can be personalized to specific learning environments. Moodle is open source, meaning that users can design different plug-ins that integrate with Moodle. Rodrigues, Fdez-Riverola, and Novais (2011) proposed a framework for Moodle where an external module is linked to Moodle, enabling the detection of student's affective states and learning styles based on the affective-loop cycle discussed previously. This affective module that includes two sub-modules, an explicit affective-state detector and an implicit affective-state detector. The explicit affective-state detector gathers information by directly posing questions to the student about their affective state. The implicit affective-state detector monitors the interactions between the student and LMS to infer the student's affect using facial analyses, mouse and keyboard analysis, and log files in an LMS. Similar to Khan et al. (2009), Rodrigues, Fdez-Riverola, and Novais (2011) suggest integrating student interaction with the LMS as a way to understand student affect with learning material; however, they advance the affective abilities of an LMS by including sensor-based measures like keyboard use and mouse use (e.g., keystrokes or clicks per minute).

3. Understand Affect State

This chapter examines emotional theories and taxonomies in an attempt to understand the function and purpose of emotions and to enumerate affective states that are relevant to the Submarine Learning Center (SLC) learning environment. In this chapter, we attempt to answer the question: What affective states are relevant to instruction delivered by synchronous distributed-learning systems? Most researchers use “affective states” as a general term covering a variety of affect-related constructs, including emotions, feelings, preferences, and attitudes. In this report, we refer specifically to emotions, which are defined as brief states that can be distinguished by distinct physiological, subjective, and behavioral signals.

A. Review of Models

To determine relevant affective states, we reviewed scientific models of emotion that have important implications for emotional expression, for artificial intelligence, and for learning. Based on these criteria, we selected seven prominent theories of emotion to help understand and model the function and purpose of emotions. These theories are presented in the chronological order of their development:

- Ekman’s Basic Emotion Theory.
- Russell’s Circumplex Model of Affect.
- Watson, Clark, and Tellegen’s Positive and Negative Affect Schedule (PANAS).
- Ortony, Clore, and Collins (OCC) Model of Emotions.
- Pekrun’s Concept of Academic Emotions.
- Graesser and D’Mello’s Learning Centered Emotions.
- Keltner’s Consensual Taxonomy.

Each of these theories provides a taxonomy that defines and organizes discrete emotive states. We compare these taxonomies in an attempt to determine the emotional states that are relevant to learning and education. The following sections provide a synopsis of each model and its theoretical constructs and taxonomy of affective states. Also discussed is relevance of the theory to the synchronous distance-learning system employed at the Navy’s SLC.

1. Basic Emotion Theory

In his 1989 review, Ekman noted that research from the late 1960s and early 1970s demonstrated that humans in literate and preliterate cultures are able to correctly identify emotional states from pictures of human facial expressions. Based on such findings, Ekman and Friesen (1978) hypothesized that the ability to detect emotional states from facial patterns of others is a skill that is universal across human cultures. These ideas evolved into Basic Emotion Theory (BET), which has become the theoretical foundation and research paradigm for the science of emotion and remains today as "central narrative" in this now-established discipline (Keltner, Sauter, Tracy, & Cowen, 2019).

a. Synopsis of the Theory

According to Eckman (1992), emotions have evolved across human cultures and even animal species. The primary function of emotions is mobilizing the organism to deal quickly with fundamental life tasks, but primarily interpersonal encounters. From this viewpoint, Ekman derived nine characteristics, described in Table 4, that identify and differentiate the basic emotions. The first characteristic (distinctive universal signals) refers to all types of physical expressions, such as vocal expressions, but Eckman (1992) maintained that the "strongest evidence for distinguishing one emotion from another comes from research on facial expressions" (p. 175). In addition, the third (distinctive physiology) and fourth (distinctive antecedent events) characteristics distinguish one emotion from another. The remaining characteristics serve to distinguish basic emotions from other related affective states, such as moods (e.g., euphoria or irritation), emotional traits (e.g., hostile or melancholic), emotional attitudes (e.g., love or hatred), and emotional disorders (e.g., depression or anxiety).

Based on these characteristics, Ekman (1992) concluded that there was good empirical evidence for six basic emotions (1) *anger*, (2) *fear*, (3) *sadness*, (4) *enjoyment/happiness*, (5) *disgust*, and (6) *surprise*. These emotions have been commonly recognized by emotion researchers as the "Big Six." In addition, Ekman contended there is "some" evidence for five additional basic emotions, although the evidence is not as strong as that for the Big Six: (1) *contempt*, (2) *shame*, (3) *guilt*, (4) *embarrassment*, and (5) *awe*.

Table 4. Common Characteristics of Basic Emotions

Characteristic	Description
1. Distinctive universal signals	The emotion is associated with a visual signal to other animals. This is often a distinct facial expression but is not limited to this type of signal.
2. Presence in other primates	There are comparable expressions of the emotion in nonhuman primates and other animals.

Characteristic		Description
3.	Distinctive physiology	There is a distinct physiological pattern associated with the emotion.
4.	Distinctive universals in antecedent events	There is a common set of conditions or events that evoke the emotion.
5.	Coherence in emotional response	There is a systematic relationship between the expression of the emotion and the physiological changes that occur during the emotional experience.
6.	Quick onset	The emotion begins so quickly that it happens before a person knows it has started.
7.	Brief duration	To prevent the deleterious effects of sustained arousal, the emotion does not endure for a long period.
8.	Automatic appraisal	There is an automatic appraisal mechanism that selectively attends to certain antecedent conditions and associates it with an emotional response.
9.	Unbidden occurrence	The response to emotional conditions is involuntary.

b. Relevance to Synchronous Distance Learning

Not only did BET provide the scientific basis for emotion research, it provided the first practical technology for detecting the emotional states of humans. Elaborating on methods originally developed by Swedish anatomist Carl-Herman Hjortsjö (1970), Ekman and Friesen (1978) developed the Facial Action Coding System (FACS). FACS is a method for classifying facial expressions by 46 movements, called action units (AUs), that are associated with specific facial muscles. FACS was originally designed to be implemented by human coders as a method to comprehensively code facial expression in an objective fashion—that is, the coding scheme was developed independently of any contextual goal, making the codes applicable to a wide variety of situations. Nevertheless, computer-automated versions have been developed for tracking for faces on video. For instance, Hamm, Kohler, Gur, and Verma (2011) developed a sophisticated automated system for analyzing facial expression in neuropsychiatric disorders. This system tracks faces on video, extracting geometric and texture features, and profiling facial movements. Less complex automated systems using consumer-type webcams have been developed to detect and interpret facial expressions in real time and in naturalistic settings (e.g., EyeSee, 2019).

Just as the technology to automatically detect facial expression was starting to mature, research began to emerge that undermines some of the basic assumptions of these technologies. Based on an extensive review of the literature, Barrett, Adolphs, Marsella, Martinez, and Pollak (2019) recently concluded that there was insufficient evidence to support the commonly held assumption that the emotional state of humans can be inferred from their facial expressions. They suggest that the emerging facial-recognition technology be functionally described as systems that detect facial movements, not emotional expressions. Barrett and colleagues maintain that systems designed to determine emotional states from facial movements alone, without considering the context, "are at best incomplete and at worst entirely lack validity, no matter how sophisticated the computational algorithms" (p. 48).

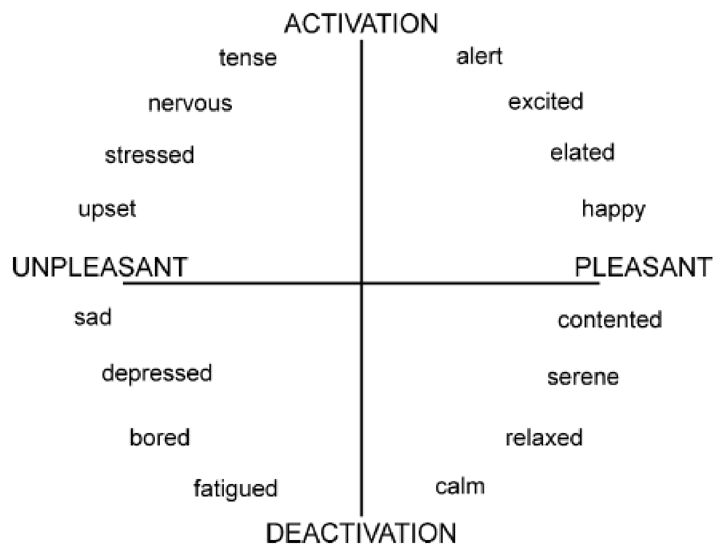
Today, researchers generally agree that there is not necessarily a one-to-one correspondence between the occurrence of an emotion and a particular facial expression—or any other prototypical physiological or behavioral expression. But there is a large body of evidence that suggests emotions are expressed as multimodal patterns of behavior (e.g., Keltner, Sauter, Tracy, & Cowen, 2019). Thus, facial expressions continue to have value as an indicator of emotion when used in combination with other physiological and behavioral indicators, as well as information on the situational context.

2. Russell's Circumplex Model of Affect

In some ways, Russell's (1980) Circumplex Model of Affect is in marked contrast to Ekman's Basic Emotions Model. In BET, basic emotions are conceived as a relatively small set of discrete and independent emotions, each having separate neural pathways and behavioral manifestations. In comparison, the Circumplex model assumes that humans experience emotions as ambiguous and overlapping experiences without borders, much like the perception of colors on the visual spectrum (Posner, Russell, & Peterson, 2005). Moreover, emotions are not viewed as independent in Russell's model. Humans rarely describe experiencing a specific positive emotion without also reporting other positive emotions; similarly, negative emotions are likely to co-occur within individuals (Watson & Clark, 1992).

a. Synopsis of the Theory

Based on findings from studies of both verbal and nonverbal expressions, Russell (1980, p. 1162) argued that affective states are related to one another in a highly systematic manner, and that those relationships can be represented "as a circle in a two-dimensional bipolar space." The horizontal dimension in this representation is valence (unpleasant – pleasant), and the vertical dimension is arousal (deactivation – activation). According to the model, any affective state can be understood to be a linear combination of varying degrees of valence and arousal. As shown in Figure 2, emotions arranged in a circle around the neutral intersection of the two dimensions. Further, emotions directly across the circle express the opposite feelings. For example, *contented* is associated with a large positive valence value but a small negative arousal value. The polar opposite emotion *upset* has a large negative valence but a low positive arousal value.



Note: Figure adapted from Barrett and Russell (1998). The *x*-axis represents the valence dimension and the *y*-axis depicts the arousal dimension.

Figure 2. The Circumplex Model of Affect

There have been numerous variants of Russell's model. The version shown in Figure 2 shows 16 affective states that are considered *core* emotions, as described by Russell and Barrett (1999). They defined core emotions as "the most elementary consciously accessible affective feelings…that need not be directed at [a specific object]" (Russell & Barrett, 1999, p. 806). A core emotion is contrasted with a prototypical emotional episode, which is a "complex set of interrelated subevents concerned with a specific object." Thus, the emotions identified in the circumplex model are generic in that they do not apply to specific objects or episodes.

The core emotions can be described by the quadrants within which they are located in the circumplex. Starting with the upper right quadrant and proceeding clockwise are emotions defined as pleasant and activated (*alert, excited, elated, happy*), pleasant and deactivated (*contented, serene, relaxed, calm*), unpleasant and activated (*fatigued, bored, depressed, sad*), and unpleasant and deactivated (*upset, stressed, nervous, tense*).

b. Relevance to Synchronous Distance Learning

The Circumplex model is important to distance learning in that it provides a method for "computing" an affective state from values of the valence and arousal dimensions. These values may be obtained from signals emanating from two separate neurophysiological systems (Gerber et al., 2008; Posner, Russell, & Peterson, 2005; Russell & Barrett, 1999). As shown in Table 5, the mesolimbic dopamine system is seen as responsible for processing reward and pleasure (valence), whereas the reticular formation regulates arousal. These separate neurophysiological systems produce different patterns of peripheral physiological responses that are potentially measurable. For instance, the Noldus FaceReader (n.d.) applies the Circumplex model to automatically recognize affective states from analysis of facial expressions.

Table 5. Neurophysiological and Peripheral Response Systems Associated with Valence and Arousal Dimensions of Affect

Affect Dimension	Neurophysiological System	Peripheral Physiological Responses
Valence	• Mesolimbic dopamine system (MDS) • MDS begins in ventral tegmental area with dopaminergic projections to the nucleus accumbens (NA) • NA has reciprocal connections to the – Amygdala – Hippocampus – Caudate nucleus – Prefrontal cortex	• Facial expressions • Corrugator muscle activation • Zygomatic muscle activation
Arousal	• Reticular formation (RF) • RF regulates arousal through connections to the limbic system and thalamus • Stimuli relayed from thalamus to amygdala	• Increased skin conductance • Heart-rate acceleration • fMRI (functional magnetic resonance imagery) signal intensity in visual cortex • Increased EEG cerebral activation

Note: Information adapted from Posner, Russell, and Peterson (2005).

3. Watson, Clark, and Tellegen's Positive and Negative Affect Schedule

David Watson, Lee Anna Clark, and Auke Tellegen (1988) developed the Positive and Negative Affect Schedule (PANAS) as a brief self-report measure of affective state. The test has been used to diagnose anxiety and depressive disorders in clinical populations (e.g., Watson, Clark, & Carey, 1988), but it has also been employed to assess the affective states of the general population.

a. Synopsis of the Theory

The theory behind PANAS was originally based on results from reanalyses of studies on self-reported mood (Watson & Tellegen, 1985). These researchers reanalyzed the data by factor analysis using varimax rotation,[2] reporting that two orthogonal factors

[2] Varimax rotation is a method used in factor analysis to simplify the solution by reducing results to a small number of orthogonal factors.

consistently emerged, which they labeled Positive Affect and Negative Affect and defined as follows:

> …Positive Affect (PA) reflects the extent to which a person feels enthusiastic, active, and alert. High PA is a state of high energy, full concentration, and pleasurable engagement, whereas low PA is characterized by sadness and lethargy. In contrast, Negative Affect (NA) is a general dimension of subjective distress and unpleasurable engagement that subsumes a variety of aversive mood states, including anger, contempt, disgust, guilt, fear, and nervousness, with low NA being a state of calmness and serenity. (Watson et al. (1988, p. 1063))

Items for the PANAS were drawn from a factor analysis of 60 adjectives that were intended to provide broad coverage of the domain of affect (Zevon & Tellegen, 1982). Watson et al. (1988) administered the items to both college student and non-student populations, asking respondents to rate on a 5-point scale how they felt today (i.e., their current state), during the past few days, or during the past few weeks.[3] Results from the factor analysis, regardless of population or time frame, showed strong evidence of two factors corresponding to positive and negative affect. To identify conceptually pure items for the PANAS, adjectives were selected that had strong loadings on their primary factor (.50 or above) but near-zero loadings on the secondary factor.

The PANAS model has been compared to the Circumplex model because both are organized around two dimensions. A cursory analysis suggests that the two dimensions in the PANAS model (positive and negative affect) are qualitatively different than the two dimensions in the Circumplex model (valence and arousal). However, these terminology differences may not reflect a fundamental discrepancy between models. That the PANAS model can be recast into the circumplex structure and vice versa (e.g., Barrett & Russell, 1998; Watson & Tellegen, 1985) demonstrates the congruence of the models.

Table 6 lists the original version of the PANAS, comprising 20 items (10 positive and 10 negative). The table also indicates 10 items selected for short versions of the PANAS developed for older respondents (Kercher, 1992) and for international populations (Thompson, 2007).

The PANAS model has been compared to the Circumplex model because both are organized around two dimensions. A cursory analysis suggests that the two dimensions in the PANAS model (positive and negative affect) are qualitatively different than the two dimensions in the Circumplex model (valence and arousal). However, these terminology differences may not reflect a fundamental discrepancy between models. That the PANAS

---

[3] Short forms of PANAS have specified longer time frames: “in the past year” (Kercher, 1992) or how respondents feel “in general” (Thompson, 2007), both of which could be regarded more as trait measures of affect.

model can be recast into the circumplex structure and vice versa (e.g., Barrett & Russell, 1998; Watson & Tellegen, 1985) demonstrates the congruence of the models.

Table 6. Positive Affect and Negative Affect Items on Original PANAS (Watson, Clark, & Tellegen, 1988)

Positive Affect	Negative Affect
Attentive[b]	Distressed[a]
Interested	Upset[a,b]
Alert[a,b]	Hostile[b]
Enthusiastic[a]	Irritable
Excited[a]	Scared[a]
Inspired[a,b]	Afraid[a,b]
Proud	Guilty
Determined[a,b]	Ashamed[b]
Strong	Nervous[a,b]
Active[b]	Jittery

[a]Included in Kercher (1992) short form.

[b]Included in Thompson (2007) short form.

b. Relevance to Synchronous Distance Learning

Research on the PANAS demonstrates that the positive and negative affective states of learners can be reliably determined by simple and short self-report checklists. Note that some versions of PANAS are focused on more enduring affective traits of respondents, rather than their current affective states. For instance, the short forms of PANAS specified longer time frames for the assessment: "in the past year" (Kercher, 1992) or how respondents feel "in general" (Thompson, 2007). Watson et al. (1988) compared assessments using various time frames: (1) "right now"; (2) "today"; (3) "during the past few days"; (4) "during the past weeks"; (5) "during the past few weeks"; (6) "during the past year"; and (7) "in general, that is, on average." Mean scores for both PA and NA increased as the time frame lengthened. This was expected because, as the time frame increases, it was more likely that a respondent experienced a significant amount of emotion. Also, the assessments showed greater test-retest reliability (i.e., stability) for longer time frames than for shorter ones. Nevertheless, the PANAS scale exhibited substantial stability even for the assessment in the current moment.

4. Ortony, Clore, and Collins (OCC) Model of Emotions

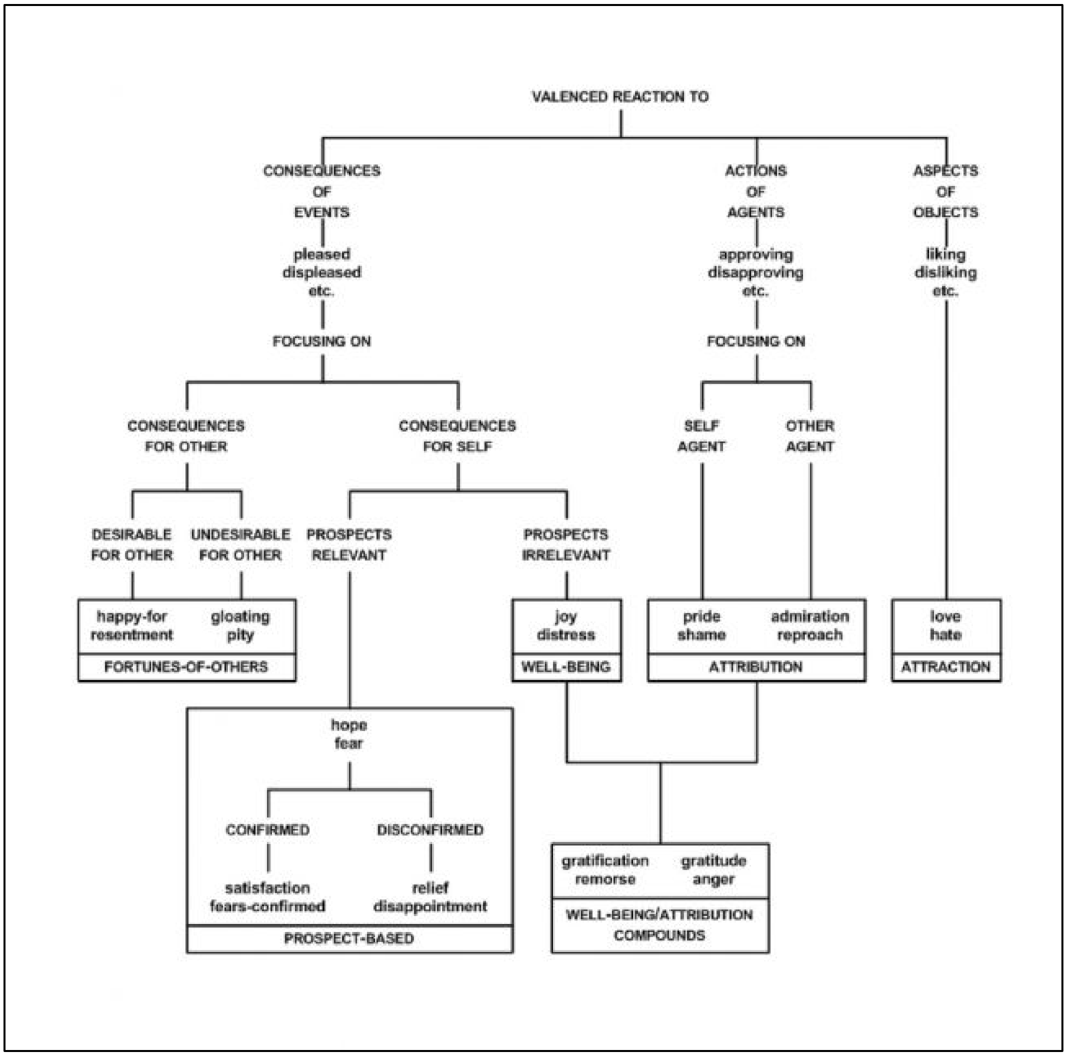
Anthony Ortony, Gerald Clore, and Allan Collins (1988) conceived their OCC model as "an account of emotion that could, in principle, be used in an Artificial Intelligence (AI) system that would, for example, be able to reason about emotion" (p. 2). This AI-based model of emotion is viewed as particularly compatible with affect-sensitive advanced learning technologies.

a. Synopsis of the Theory

According to the OCC model, emotions are the result of cognitive appraisals of psychologically significant situations. As outlined in Table 7, these appraisals are characterized on two dimensions. First, appraisals are attributed to either events, agents, or objects. Second, appraisals are focused on different aspects of the situation, are appraised on particular valence dimensions, and based on specific criteria. As shown in Figure 3, the OCC model further differentiates between emotions arising from attributions of the desirability of one's own outcomes (e.g., *hope/fear*) versus another's outcomes (e.g., *gloating/pity*). Similarly, emotions differ with respect to the appraisals of the praiseworthiness of one's own actions (e.g., *pride/shame*) versus the actions of others (e.g., *admiration/reproach*). However, attributes of objects are appraised only on appeal (e.g., *love/hate*). The resulting model identifies 11 pairs, or 22 individual emotions.

Table 7. Elements of OCC Model

Attribution Groups	Appraisal		
	Focus of Interest	**Valence Dimension**	**Criteria**
Events	Consequences of events	Desirability: Pleased/Displeased	Goals: events promote or thwart one's goals
Agents	Actions of agents	Praiseworthiness: Approving/Disapproving	Standards: agents act in accord with social, moral, and behavioral standards
Objects	Aspects of objects	Appeal: Liking/Disliking	Attitudes: objects are compatible with one's tastes and attitudes



Note: Reproduced from Ortony, Clore, and Collins (1988, 19).

Figure 3. Structure of Emotions in OCC Model

b. Relevance to Synchronous Distance Learning

The OCC model has often been used to synthesize and generate emotional reactions in intelligent agents (e.g., Bartneck, Lyons, & Saerbeck, 2008; Steunebrink, Dastani, & Meyer, 2009). It has even been incorporated into an emotional appraisal engine (GAMYGDALA) for generating emotions of nonplayer characters in games (Broekens, Hudlicka, & Bidarra, 2016).

There have been fewer applications of the OCC model to recognize emotions. This is partly due to the theory's assumption that "emotions are more readily distinguished by the situations they signify than by patterns of bodily responses" (Clore & Ortony, 2013, p. 335). Thus, the model has no provisions for interpreting physical responses, including facial expressions. However, the model implies that emotive states could be derived from an analysis of psychologically significant situations. At least two studies using the OCC model have demonstrated that contextual information can be derived from text or speech to recognize the emotional content of the discourse (Shaikh, Prendinger, & Ishizuka, 2009; Udochukwu & He, 2015).

5. Pekrun's Concept of Academic Emotions

The previously described theories pertain to emotions experienced in human life in general. Reinhard Pekrun, a professor at the University of Munich, has sought to focus his efforts on the emotions that specifically apply to academic learning in the classroom.

a. Synopsis of the Theory

As summarized in Table 8, Pekrun (2014) distinguishes among four broad types of academic emotions that affect classroom instruction: achievement, epistemic, topic, and social emotions. Topic emotions relate to the content of learning whereas social emotions pertain to the social context of the learning environment. Of these four types of academic emotions, the theory behind achievement and epistemic emotions are the most well developed and relevant to technology-based instruction. The following discussion therefore focuses on those two types.

Table 8. Four Types of Academic Emotions

Type	Description	Examples
Achievement emotions	Emotions related to the success or failure of learning activities/outcomes that are judged according to competence-related standards of quality.	*Hope* and *pride* related to success. *Anxiety* and *shame* related to failure.
Epistemic emotions	Emotions elicited by cognitive problems or solutions experienced during learning.	*Surprise, curiosity, confusion, frustration, anxiety*, and *delight*.
Topic emotions	Emotions associated with content of learning, but not directly related to learning and problem-solving.	*Empathy* for protagonists, *disgust* for medical procedures, and j*oy* for music.
Social emotions	Emotions related to feelings toward classroom teachers and peers, having potential effects on teacher-student interaction and in-group learning.	*Social anxiety*. *Love* and *sympathy* in relationships with classmates and teachers. *Compassion, admiration, envy, anger, contempt*, or *empathy* related to success or failure of others.

1) Achievement Emotions

Pekrun defines achievement emotions as those that relate to achievement activities and outcomes that are judged on competency-based standards of quality (Pekrun & Linnenbrink-Garcia, 2012). In his control-value theory of achievement emotions, Pekrun (2006) distinguished between emotions related to achievement activities (e.g., *hope* and *anxiety*) and achievement outcomes (e.g., *enjoyment* and *anger*). Pekrun further divided outcome emotions into prospective, anticipatory emotions (e.g., *hope for success*, *anxiety of failure*) and retrospective emotions for outcomes in the past (e.g., *pride* or *shame*).

In the Pekrun model, an achievement emotion is determined from two types of cognitive appraisals. The first is an appraisal of the subjective *value* of an outcome or activity. Value relates to the perceived importance of task success or failure or to the valence or appeal of an activity. The second is an appraisal of the perceived *control* that a person has over the outcome or activity. Control refers to the extent to which the learner's actions or the learner's situation will lead to a positive outcome. Subjective control and value combine to determine a unique emotion. For instance, if learners value success on an academic task over which they have a high degree of control, then they will experience *anticipatory joy*. On the other hand, if learners have a low degree of perceived control over valued tasks, they will experience *hopelessness* regardless of whether the perceived value is positive or negative. The combinations are not always linear; for instance, for a low-valued activity, learners will be *bored* if they perceive that they have either a high or a low degree of control over outcomes. Similarly, if learners perceive that they have little control over outcomes, they will experience *frustration* if they value task success or the avoidance of task failure. As shown in Table 9, this scheme identifies 14 unique achievement motivations, two of which (*hopelessness* and anger) are listed twice. *Hopelessness* results from positive or negative future outcomes over which the learner has little or no control. Furthermore, *anger* results from a negative prior outcome that was controlled by some other entity or a negative activity over which the learner has a high level of control.

Table 9. Achievement Emotions According to Pekrun's (2006) Control-Value Theory

Object Focus	Appraisals: Value	Appraisals: Control	Emotion
Outcome/prospective	Positive (success)	High	*Anticipatory joy*
		Medium	*Hope*
		Low	***Hopelessness***
	Negative (failure)	High	*Anticipatory relief*
		Medium	*Anxiety*
		Low	***Hopelessness***
Outcome/retrospective	Positive (success)	Irrelevant	*Joy*
		Self	*Pride*
		Other	*Gratitude*
	Negative (failure)	Irrelevant	*Sadness*
		Self	*Shame*
		Other	***Anger***
Activity	Positive	High	*Enjoyment*
	Negative	High	***Anger***
	Positive/Negative	Low	*Frustration*
	None	High/Low	*Boredom*

Notes: Adapted from Pekrun (2006). Note that two emotional states (hopelessness and anger) are evoked by two separate sets of goals and appraisals. Also, note that the control column refers to either level or source of control.

A unique aspect of achievement emotions, as depicted by Pekrun and associates, is that they are reciprocally related to academic outcomes (Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017). In agreement with most theories of affect and performance, they propose that achievement emotions affect academic outcomes through the allocation of cognitive resources and attention toward positively appraised tasks and away from negatively appraised ones. However, they also hypothesize that outcomes affect emotions. The reverse relationship is mediated through the effect of outcomes on students' perceived competence and level of control. In short, positive outcomes lead to heightened perceived competence and control, which in turn result in positive emotions. Likewise, negative outcomes result in reduced competence and control, leading to negative emotions.

Findings from a recent multiyear longitudinal investigation of German adolescents' mathematics achievement support the reciprocal relationship between achievement emotions and academic outcomes (Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017). Emotions were assessed using the Achievement Emotions Questionnaire-Mathematics (AEQ-M), which measured the students' emotional trait-like dispositions on seven scales: (1) *enjoyment*, (2) *pride*, (3) *anger*, (4) *anxiety*, (5) *shame*, (6) *boredom*, and

(7) *hopelessness*. Results indicated that the emotions predicted academic outcomes (end-of-year math grades and standardized test scores), and that those end-of-year outcomes predict the emotions in the next year.

2) Epistemic Emotions

Pekrun, Vogl, Muis, and Sinatra (2017) define epistemic emotions as those that relate to learning and knowledge generation. In that regard, they are much like learning-centered emotions (see Section 3.A.6 below). They differ from achievement emotions in terms of their object focus. Whereas achievement emotions (e.g., *pride* and *shame*) are centered on academic success or failure, epistemic emotions (e.g., *curiosity* and *confusion*) are focused on knowledge generation. However, the researchers also point out that some emotions can be experienced as either epistemic or achievement, depending on the object focus:

> For example, during cognitive activities, some emotions can be experienced as epistemic emotions or as achievement emotions…A student's frustration at not deriving a correct solution to a mathematics problem would be considered an epistemic emotion if the focus is on the cognitive incongruity resulting from the unsolved problem. However, if the focus is on personal failure and the inability to solve the problem, then the student's frustration would be considered an achievement emotion. (Pekrun, Vogl, et al., p. 1269)

Pekrun, Vogl, et al. (2017) developed a paper-and-pencil instrument for assessing epistemic emotions during these learning activities: reading and comprehending conflicting multiple texts on the causes and consequences of climate change. This instrument, labeled the Epistemically-Related Emotion Scales (EES), was designed to measure seven epistemic emotions: (1) *surprise*, (2) *curiosity*, (3) *enjoyment*, (4) *confusion*, (5) *anxiety*, (6) *frustration*, and (7) *boredom*. EES items are single emotional adjectives (e.g., monotonous, excited, astonished) that relate to one of the seven epistemic emotions. The full version of EES presents 3 different items for each emotion, or 21 items in all. A short seven-item version of EES was also developed that consists of only one item per epistemic emotion.

b. Relevance to Synchronous Distance Learning

Pekrun's model is important in that it provides a comprehensive list of emotions that are relevant to academic contexts. Note that as currently conceived, the model is focused on traditional live classroom education, which limits its relevance to advanced technology (i.e., computer-based or web-based) instruction.

6. Graesser and D'Mello's Learning-Centered Emotions

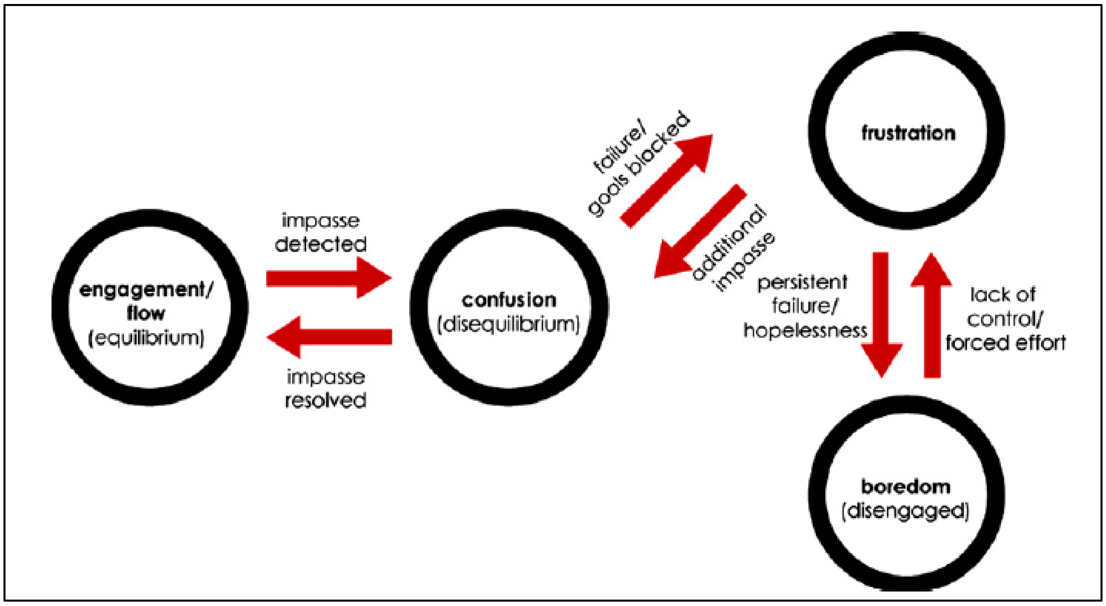
In contrast to Pekrun's treatment of broad set of emotions relating to varied academic situations, Graesser and D'Mello have focused on a smaller set of affective states that arise during short individual learning sessions implemented on advanced learning technologies.

a. Synopsis of the Theory

The learning-centered emotions model is derived from a detailed theory of complex learning (e.g., Graesser & D'Mello, 2012). Complex learning happens when people attempt to understand difficult material or solve a challenging problem. Complex learning is contrasted with shallow learning processes, which are involved in memorizing definitions or other facts. Successful complex learning, or deep learning, results from the learner's experiences of problems and solutions in coping with difficult content. Difficult materials are usually multifaceted, involving dynamic learning processes required at different points in the learning timeline. Accompanying these dynamic learning processes are characteristic emotions or affective states, such as *confusion* and *frustration* with learning problems and *delight* in discovering solutions.

Central to Graesser and D'Mello's (2012) theory of deep learning is the concept of cognitive disequilibrium. Cognitive disequilibrium occurs when the learner faces a discrepancy between the current task situation and the learner's knowledge or skill state. This discrepancy launches a trajectory of affective and cognitive states and processes to reduce the disequilibrium by either acquiring knowledge or skill relevant to the task situation or disengaging from the task. Figure 4 illustrates possible sequences of cognitive and emotive processes that may occur during complex learning. Note that *confusion* brought on by disequilibrium can lead to either *frustration* if the discrepancy is not resolved or *engagement/flow* if the discrepancy is resolved. The movement from *confusion* to *engagement* is considered to be a key mechanism for deep learning.

There are six primary learning-centered emotions that Grasser and D'Mello (2012) identified from a number of empirical studies employing different learners, topics, and technologies: *confusion, frustration, boredom, flow/engagement, delight,* and *surprise*. The first four occur more frequently and last for longer periods of time, whereas the last two are relatively infrequent and brief. These primary emotions occur across a broad range of learning tasks and situations. In contrast, two learning-centered emotions, *anxiety* and *curiosity*, are experienced in specific situations: *anxiety* when learners are faced with high-stake assessments, and *curiosity* when the learners are allowed freedom of response or when the task is intrinsically motivating. Table 10 summarizes these characteristics.



Note: Figure from Grasser and D'Mello (2012, 190)

Figure 4. Trajectory of Cognitive and Emotive Processes Triggered by Cognitive Disequilibrium

Table 10. Characteristics of Learning-Centered Emotions

Learning-Centered Emotion	Frequency	Duration	Situation
Confusion	Often	Long	General
Frustration	Often	Long	General
Boredom	Often	Long	General
Flow/engagement	Often	Long	General
Delight	Rarely	Brief	General
Surprise	Rarely	Brief	General
Anxiety	Situation-specific	Situation-specific	Learners face high-stakes assessment
Curiosity	Situation-specific	Situation-specific	Learners allowed freedom of response

A selective meta-analysis of 24 studies conducted by D'Mello (2013) focused on emotions experienced by students using advanced learning technologies in relatively short (30 to 90 minute) sessions. The findings confirmed that *flow/engagement*, *boredom*, and *confusion* were experienced relatively frequently. Also, *curiosity* and *frustration* were reliably observed but with lower frequency, as was the basic emotion of *happiness*. Notably, the other five basic emotions—*anger, fear, sadness, disgust*, and *surprise*—were not frequently observed during advanced learning technology sessions. Also, the learning-centered emotions of *delight*, *surprise*, and *anxiety* were infrequently observed in this study. As noted above, however, the theory predicts that these emotions should occur only briefly or under a specific set of circumstances.

b. Relevance to Synchronous Distance Learning

The learning-centered emotions are relevant because they are explicitly focused on those states experienced during interactions with advanced learning technologies. However, the context of those interactions is restrictive, pertaining to relatively short-term individual learning experiences with little or no human trainer intervention. In contrast, the synchronous collective/cooperative learning context in the SLC learning environment involves both student-student and student-trainer interactions. We would therefore expect a larger range of affective states to be implicated, particularly the social emotions as described by Pekrun.

7. Keltner's Consensual Taxonomy

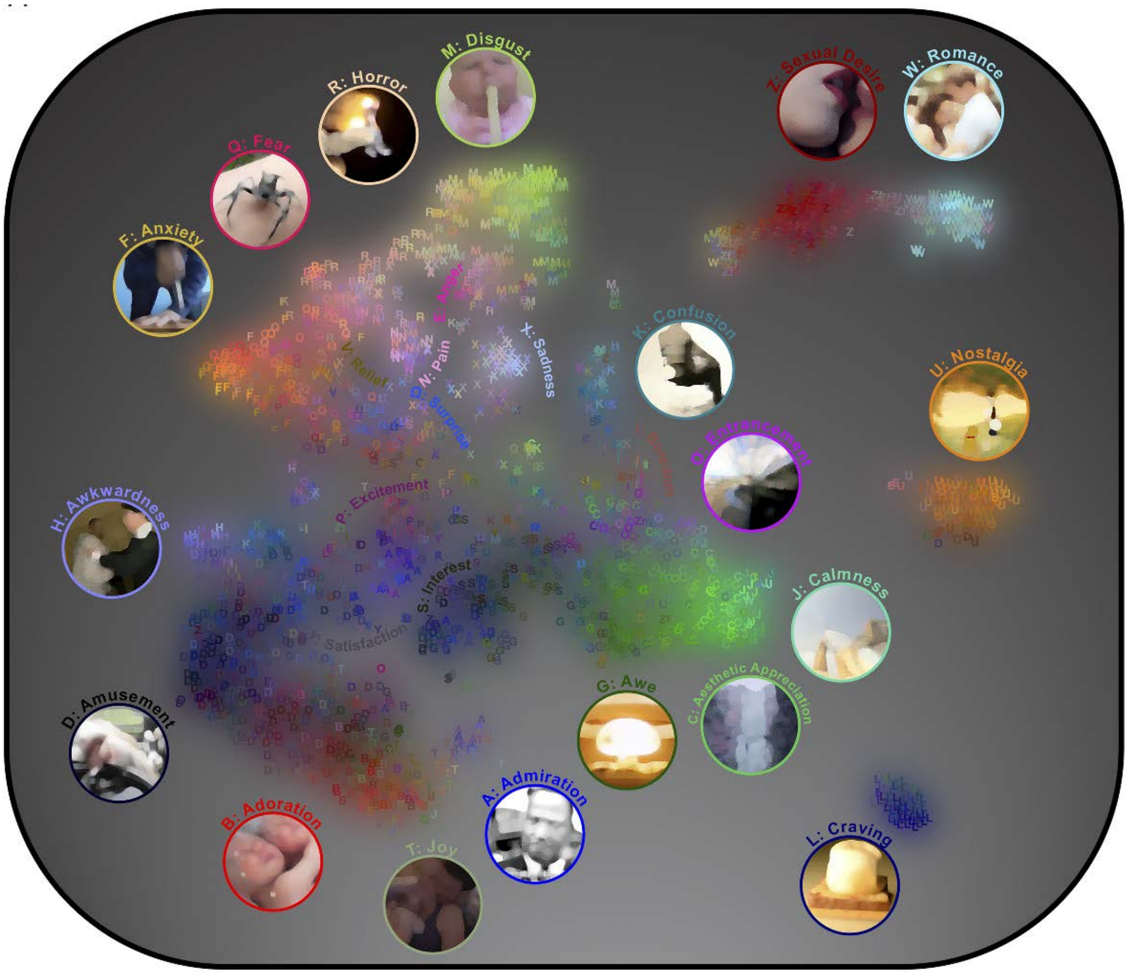
Dacher Keltner views his theory of emotion as an evolution of Ekman's notion of basic emotions (Keltner, Sauter, Tracy, & Cowen, 2019). But instead of there being a relatively small set of basic emotions, Keltner's theory holds that humans can reliably detect more than 20 emotional states.

a. Synopsis of the Theory

In a recent series of articles, Keltner, Alan Cowen, and their associates at the Berkeley Social Interaction Laboratory have demonstrated that humans are able to detect a much larger set of basic emotions than previously thought (Cowen, Elfenbein, Laukka, & Keltner, 2019; Cowen & Keltner, 2017, 2018, 2019; Cowen, Sauter, Tracy, & Keltner, 2019). Proponents of Basic Emotion Theory (BET) thought that humans are capable of recognizing fewer than 10 basic emotions, but evidence is mounting that they can reliably detect more (perhaps much more) than 20 emotional states.

In agreement with BET and appraisal theories of emotion, Keltner and colleagues view emotions as discrete affective states: "Emotions are internal states that arise following appraisals (evaluations) of interpersonal or intrapersonal events that are relevant to an individual's concerns…and promote certain patterns of response" (Cowen et al., 2019, pp. 72–73). In opposition to BET, however, these emotions are not independent, but highly interrelated, and not in a two-dimensional space as described in the Circumplex Model or in PANAS, but in a multidimensional hyperspace. The hyperspace model is derived using new big-data techniques. This model did not evolve from a fixed list of emotions based on scientists' assumptions or from standard measures based on forced-choice recognition tasks under controlled conditions. Rather, Keltner's theory is based on empirical patterns under more realistic, free-response conditions. The result is a high-dimensional semantic space that provides a comprehensive representation of the varieties of emotional expression. Some of the original work mapped subject responses to evocative videos (Cowen & Keltner, 2017; see *Note:* Figure taken from Cowen and Keltner (2017): The figure represents “a chromatic map of average emotional responses to 2,185 videos within a 27-dimensional categorical space of [self-]reported emotional experience…The resulting map reveals gradients among distinct varieties of reported emotional experiences, such as the gradients from anxiety to fear to horror to disgust” (figure 2, page E7904).

Figure 5). Similar findings have since been obtained using different emotion-evoking stimuli, such as emotional concept words, non-semantic vocal bursts, and videos of facial/bodily expressions (Cowen, Sauter, Tracy, & Keltner, 2019).



Note: Figure taken from Cowen and Keltner (2017): The figure represents “a chromatic map of average emotional responses to 2,185 videos within a 27-dimensional categorical space of [self-]reported emotional experience…The resulting map reveals gradients among distinct varieties of reported emotional experiences, such as the gradients from anxiety to fear to horror to disgust” (figure 2, page E7904).

Figure 5. Videos Mapped Along 27 Categorical Judgment Dimensions of Reported Emotional Experience

Noting the recent explosion of interest and debates in emotion research, Keltner (2019) recently proposed a "consensual" taxonomy of emotions to describe the basic points of agreement (circa 2019). Like Ekman's (1992) model of basic emotions, Keltner's consensual model holds that emotions are brief but discrete states, focused on different objects (including the self), defined by a central appraisal tendency, and possessing unique signals and underlying physiology. Unlike Ekman, Keltner maintained that there are more than six basic emotions, stating that there is "tacit inclination" among emotion researchers to recognize 20–25 different emotional states. As summarized in Table 11, his suggested list includes Ekman's original Big Six emotions, 4 emotions related to self-consciousness, and 13 emotions that describe an array of positive affective states. Keltner noted that the positive emotions are underresearched and should receive systematic attention. Keltner cautioned that the whole list was not definitive; rather, he described it as a working taxonomy that would evolve as evidence accumulates.

Table 11. Keltner's (2019) Consensual Taxonomy

Category	Emotion
Big Six emotions	Anger
	Disgust
	Fear
	Happiness
	Sadness
	Surprise
Self-conscious emotions	Embarrassment
	Shame
	Guilt
	Pride
Positive emotions	Amusement
	Awe
	Contentment
	Desire
	Ecstasy
	Gratitude
	Interest
	Joy
	Love
	Pride
	Relief
	Sympathy
	Triumph

Note that in Table 11, *pride* is listed as both a self-conscious and a positive emotion. No explanation was provided, but we speculate that he was implying that there may be different conceptual definitions of the emotion. As a self-conscious emotion, *pride* is explicitly directed toward oneself, whereas the positive emotion of *pride* may be either directed inwardly toward the self or outwardly toward other objects or people.

b. Relevance to Relevance to Synchronous Distance Learning

Current research suggest that the set of affective states that are relevant to advanced learning technologies is small. For instance, D'Mello (2013) found evidence for six emotions (*engagement/flow*, *boredom*, *confusion*, *curiosity*, *happiness*, and *frustration*) occuring with measurable frequency during individual sessions with advanced learning technologies. However, researchers are beginning to recognize that an individual interaction with an intelligent tutoring system is an emotionally impoverished situation, and this small set of emotions may not apply to the more complex group-learning situations in the SLC synchronous learning system. The SLC system may elicit more self-conscious and positive emotions as referenced in the consensual taxonomy. But as it currently stands, the consensual taxonomy is merely a listing of emotional states that can be reliably detected; the list has not been related to specific technologies or learning situations.

B. Synthesis of Models

Our review of models begins with Ekman's Basic Emotion Theory, perhaps best articulated in Ekman (1992). Keltner, Sauter, Tracy, and Cowen (2019) maintain that BET, nearly 30 years after this formulation, remains the "central narrative" in emotion research, and it provides a point of comparison in the development of competing theoretical viewpoints. Thus, we use BET as the point of reference for discussing the differences and similarities among the various models of emotion.

1. Points of Disagreement

Although BET is central to the science of emotion recognition, many researchers criticize the theory from a number of perspectives. The following are examples of criticisms that researchers have aimed at BET.

a. Emotions Are Not Discrete States

Some theorists do not regard emotions as discrete states, but rather as determined by a combination of continuous variables, such as valence and arousal (e.g., Russell, 1980; Russell & Barrett, 1999).

b. There Are Multiple Modes of Emotional Expression, Not Just Facial Movements

Researchers have also criticized the focus of BET on facial movements as the primary behavioral expression of emotions. Research indicates that humans are able to reliably recognize emotions based on diverse cues, such as complex movements of face, head, body, and hands; subtle shifts in gaze, brief touches, postural movements, and a variety of non-semantic vocal cues (Cowen, Sauter, Tracy, & Keltner, 2019).

c. There Are More than Six Basic Emotions

Researchers generally agree that there are more than six basic emotions as specified in BET. Keltner (2019) recently asserted that there is now good evidence that 20–25 states have emotion-like properties.

d. Emotions Arise from a Cognitive Appraisal Process

Any theory of emotion recognition needs to incorporate cognitive appraisal processes as explicitly stated in the OCC (Ortony, Clore, & Collins, 1988) and Pekrun's (2006) control-value theories of emotion. Appraisal processes were not considered in the original statement of BET (Ekman & Friesen, 1971), but were incorporated into the later theoretical update (Ekman, 1992).

e. The Context of the Expression is Important

In emotion recognition, the context of expression is as important, or more important, than the expression itself. This is expressed most strongly by Clore and Ortony (2013, p. 335): "emotions are more readily distinguished by the situations they signify than by patterns of bodily responses."

2. Points of Agreement

Despite the diversity of theoretical constructs in the science of emotion, researchers generally agree on some fundamental characteristics and processes of affective states. The following points of agreement were chosen because they had relevance to the technology of emotion recognition.

a. Emotional States Arise Through an Unconscious Appraisal Process

Emotions and the expression of those emotions arise spontaneously without much initial awareness by the person experiencing the emotion. That people have limited conscious control of this process implies that they have difficulty hiding their emotions. For most people, the experience of emotion is an automatic response over which they have little control. Unless a person is consciously trying to deceive others, the emotion that the person displays is the emotion that the person is experiencing.

b. Emotions Are Relatively Brief and Transitory States

Emotions are transitory experiences lasting only seconds, not minutes. The emotional experience at time *t* may not be the same as the experience at time *t* + 30 s. This implies that assessment or measurement of affect must occur in real time or be explicitly time linked.

c. Observable Expressions of Emotion Are Reliable Indications of Internal Emotional States

Any particular emotion is brought about by a distinct set of antecedent conditions or events and has unique behavioral and physiological expressions. That is, certain situations (e.g., someone cuts in front of you in a long checkout line in a grocery store) are likely to arouse certain emotions in you (*anger*). Further, you may express that anger behaviorally (loud verbal comments directed at the offender) or physiologically (release of adrenaline causing your face to redden). The implication is that there are reliable external signals that you are experiencing a certain emotion.

d. Humans Are Good at Recognizing Emotional States Experienced by Others

To extend the previous example, if I viewed you at a distance yelling at someone in line in front of you, I would guess that you were angry—even if I weren't close enough to hear what you were saying. But if I had also viewed the antecedent events (line-cutting), I would be even more certain that you were experiencing anger. Humans are quite good at correctly recognizing the emotional expression of others, and even better when they can also perceive the situation in which the expression occurs. Researchers view this ability as an important basic skill in forming social relations.

C. Synthesis of Taxonomies

The taxonomies are more difficult to synthesize. A tally of the 7 models indicates that there are 75 qualitatively different affective states that were identified by at least one of the 7 models reviewed. Table 12 crosswalks the states to each of the seven models. This analysis reveals surprisingly little agreement among the models; notably, over half of emotions (43, or 57.3%) are endorsed by only one of the seven models. The table also sorts the 75 emotions by the number of models that endorse each, from the most endorsed to the least endorsed emotions. As described below, the results are sorted in five tiers of emotions, from the most to the least level of consensus.

Table 12. Emotions Sorted into Five Tiers of Endorsement

Affective State	BET	CMA	PANAS	OCC	AE	LCE	CT
Endorsed by Five Models							
Happiness/enjoyment	+	+			+	+[a]	+
Endorsed by Four Models							
Anger	+			+	+		+
Pride			+	+	+		+
Sadness	+	+			+		+
Shame	+[b]			+	+		+
Surprise	+				+	+	+
Endorsed by Three Models							
Boredom		+			+	+	
Fear	+			+			+
Gratitude				+	+		+
Guilt	+[b]		+				+
Joy				+	+		+
Love				+	+		+
Endorsed by Two Models							
Admiration				+	+		
Alert		+	+				
Anxiety					+	+	
Awe	+[b]						+
Confusion					+	+	
Contempt	+[b]				+		
Contentment		+					+
Curiosity					+	+	
Delight					+	+	
Disgust	+						+
Distress			+	+			
Embarrassment	+[b]						+
Excited		+	+				
Frustration					+	+	
Hope				+	+		
Interest			+				+
Nervous		+	+				
Relief				+			+
Sympathy					+		+
Upset		+	+				

Affective State	BET	CMA	PANAS	OCC	AE	LCE	CT
			Endorsed by One Model				
Active			+				
Afraid			+				
Amusement							+
Anticipatory joy					+		
Anticipatory relief					+		
Ashamed			+				
Attentive			+				
Calm		+					
Compassion					+		
Depressed		+					
Desire							+
Determined			+				
Disappointment				+			
Ecstasy							+
Elated		+					
Empathy					+		
Engagement/Flow						+	
Enthusiastic			+				
Envy					+		
Fatigued		+					
Fears-confirmed				+			
Gloating				+			
Gratification				+			
Happy-for				+			
Hate				+			
Hopelessness					+		
Hostile			+				
Inspired			+				
Irritable			+				
Jittery			+				
Pity				+			
Relaxed		+					
Remorse				+			
Reproach				+			
Resentment				+			
Satisfaction				+			
Scared			+				

Affective State	BET	CMA	PANAS	OCC	AE	LCE	CT
Serene		+					
Social anxiety					+		
Stressed		+					
Strong			+				
Tense		+					
Triumph							+

Notes: BET = Basic Emotion Theory, CMA = Circumplex Model of Affect, PANAS = Positive and Negative Affect Schedule, OCC = Ortony, Clore, and Collins Model, AE = Academic Emotions, LCE = Learning Centered Emotions, CT = Consensual Taxonomy.

Despite their being endorsed by only a single theory, Tier 5 emotions in ***bold italics*** deserve further consideration due to their relevance to learning.

[a] Not in theoretical taxonomy, but D'Mello's meta-analysis (2013) shows evidence for happiness during sessions using advanced learning technologies.

[b] Outside of Big Six, but Ekman (1992) considered there to be "some evidence" that awe, contempt, embarrassment, guilt, and shame are basic emotions.

1. Endorsed by Five Models

At the very top of the list is *enjoyment/happiness*, which is endorsed by five of the seven models. This emotion was identified as one of the basic and universal emotions by Ekman (1992) and by Keltner's consensual taxonomy (2019), as a core emotion in the Circumplex model (Barrett and Russell 1998), and as an achievement emotion in Pekrun's (2006) model of academic emotions. *Happiness* is not technically in Graesser and D'Mello's (2012) model of learning-centered emotions, but D'Mello's (2013) meta-analysis shows evidence that the emotion is prevalent during interactions with advanced learning technologies.

2. Endorsed by Four Models

The next tier are the emotions endorsed by four taxonomies. This tier comprises five emotions, three of which (*anger*, *sadness*, and *surprise*) are significant in that they are Big Six emotions. The fourth affective state in this tier, *shame*, was also named by Ekman (1992) as a possible basic emotion, although it is not in the Big Six. The fifth element in this tier, *pride*, is viewed as self-directed positive emotion in the PANAS, OCC, and Pekrun's academic emotions models. Keltner's (2019) consensual model implied that *pride* could also be viewed as being outwardly directed at another person or object, as well as being an inwardly directed positive emotion.

3. Endorsed by Three Models

The third tier comprises emotions endorsed by three models. This level includes six emotional states: *boredom*, *fear*, *gratitude*, *guilt*, *joy*, and *love*. *Fear* and *guilt* are recognized as negative emotions. *Fear* is a Big Six affective state recognized in both Ekman's basic emotions theory and Keltner's consensual taxonomy and a negative affective state in the PANAS model. *Guilt* is recognized as a possible basic emotion in Ekman's model, an assessment of negative consequences for self in the OCC model, and a negative self-conscious emotion in Keltner's taxonomy. *Gratitude*, *joy*, and *love* are all positive emotions in the OCC model, Pekrun's academic emotions, and Keltner's consensual taxonomy.

Unlike the other five emotions in this tier, *boredom* is described differently in the models. *Boredom* is characterized as an unpleasant, deactivated core emotion in the Circumplex model. It is also viewed as a primary learning-centered emotion that frequently occurs in learning situations, and it is both an achievement and epistemic emotion in Pekrun's model.

4. Endorsed by Two Models

Tier 4 consists of 20 affective states, each endorsed by 2 models. As described below, emotions in this tier can be divided into four subgroups.

a. Learning-Centered and Academic Emotions

Anxiety, *confusion*, *curiosity*, *delight*, and *frustration* are all recognized as both learning-centered emotions (Graesser & D'Mello, 2012) and academic emotions (Pekrun, 2014).

b. Core Emotions

Alert, *contentment*, *excited*, *nervous*, and *upset* are classified as core emotions in the Circumplex model described by Barrett and Russell (1998). *Alert* and *excited* are pleasant and activated emotions in that scheme, whereas *contentment* is pleasant and deactivated. *Nervous* and *upset* are negative and activated core emotions. Except for *contentment*, the emotions in this category are also endorsed by the PANAS model (Watson, Clark, & Tellegen, 1988). The second endorsement for *contentment* is Keltner's (2019) consensual model, which describes it as a positive emotion.

c. Basic Emotions

Awe, *contempt*, *disgust*, and *embarrassment* are given as basic emotions in Ekman's (1992) taxonomy and Keltner's (2019) consensual list of emotions.

d. Positive and Negative Affective States

Emotions in this category are positive or negative affective states identified on the PANAS (Watson, Clark, & Tellegen, 1988). Five of the six emotions in this category (*admiration, hope, interest, relief*, and *sympathy*) are identified as positive states by PANAS and either OCC (Ortony, Clore, & Collins, 1988), academic emotions (Pekrun, 2014), or Keltner's (2019) consensual taxonomy. *Distress* is viewed as a negative affective state in the PANAS and OCC models.

5. Endorsed by One Model

The fifth tier, comprising affective states endorsed by only a single model, is by far the largest. These 43 emotions in this tier can be cast into four subgroups.

a. Continued Consideration

Despite being cited only by a single taxonomy, 11 emotions should continue to be under consideration because they have clear implications for learning and instruction. They can be sorted into four subsets:

- *Positive Emotions:* Keltner (2019) cites a number of positive emotions that have received increasing research interest stimulated by Fredrickson's (2004) broaden-and-build theory. Nine of those positive emotions have been noted in higher tiers (*awe, contentment, gratitude, interest, joy, love, pride, relief*, and *sympathy*), but four have not (*amusement, desire, ecstasy,* and *triumph*).
- *Social emotions:* These affective states may come into play in the cooperative-learning approach implemented in the virtual classroom paradigm. In that regard, Pekrun (2014) identifies four social emotions that relate to the success and failure of others: *compassion*, *empathy*, *envy*, and *social anxiety*.
- *Negative Polar Opposites:* Two negative emotions are the polar opposites of important positive affective states. One is *hate*, which is the antipode of *love* in the OCC model (Ortony, Clore, & Collins, 1988). The other is *hopelessness*, which is portrayed as the antithesis of *hope* in Pekrun's (2014) academic model.
- *Engagement/Flow:* Only one taxonomy, the learning-centered model (Graesser & D'Mello, 2012), endorses *engagement/flow* as a discrete and distinguishable emotion. This is partly because engagement is viewed differently from other emotions—it is usually seen as less of a diagnostic clue of an ongoing learning process (i.e., an independent variable) and more of a desired end state of the process (a dependent variable). Regardless, engagement is an important emotional state to consider in relation to learning and memory processes.

b. Moods

Some of the affective states in this tier are more accurately described as moods rather than emotions. Moods are pervasive and sustained affective states having no unique nonverbal expression. Emotions, in contrast, are short-lived feelings that can be recognized by overt nonverbal expressions. The following affective states in Tier 4 are regarded as moods: *active, attentive, depressed, determined, enthusiastic, fatigued, hostile, inspired, irritable, jittery, relaxed, stressed, strong,* and *tense*.

c. Idiosyncratic Affects

These emotions in this subgrouping are overly specific, applying to a specific theory. *Anticipatory joy* and *anticipatory relief* are prospective emotions defined by Pekrun's (2006) control value theory of achievement motivation. The remaining emotions in this subgroup are all derived from the OCC model of emotions (Ortony, Clore, & Collins, 1988): *disappointment, fears-confirmed, gloating, gratification, happy-for, pity, remorse, reproach, resentment,* and *satisfaction*.

d. Similar Concepts

Six emotions in Tier 5 appear to be synonyms or closely related to emotions in higher tiers: *afraid* and *scared* (cf. *fear* in Tier 3), *calm* and *serene* (cf. *contentment* in Tier 4), *ashamed* (cf. *shame* in Tier 2), and *elated* (cf. *enjoyment/happiness* in Tier 1).

6. Summary

Our analysis of the taxonomies of the seven prominent theories of emotion failed to identify a small set of commonly accepted affective states. Instead, we identified a relatively large set of emotive states that differed in their endorsements from the theories. In the end, we identified 43 affective states that could be considered as candidate emotions for instructional feedback, including 32 that were endorsed by two or more models and 11 endorsed by only a single model yet were deemed relevant to learning. Not all these emotions could or should be monitored for any educational application. Whether an emotion should actually be monitored depends on (1) the accuracy of the technology that is used to detect the affective state and (2) the effect that the affective state has on learning processes and outcomes.

4. Determine Learning Intervention

Even if an instructor or learning system were able to reliably detect the affective state of students, the question remains about what sort of instructional intervention should be performed. Should the intervention induce another certain affective state, enhance or diminish the current state, or is an intervention even necessary? These questions are the focus of this chapter.

D'Mello and Graesser (2015) proposed that affect-sensitive learning systems are designed to intervene either proactively or reactively. Proactive systems are those that are designed before the fact to induce positive emotional states or impede negative states. In comparison, reactive systems are those that detect and respond to affective states as they arise.

Aleven, McLaughlin, Glenn, and Koedinger (2016) described a similar distinction in terms of three general approaches to adapting instruction to learner affect and motivation.[4] These three adaptation strategies differ in the focus of the adaptations and in terms of their time scale. The slowest is *design-loop adaptivity*, which refers to adjustments to course design that are made before course implementation or between course iterations. Design-loop adaptivity is essentially equivalent to D'Mello and Graesser's concept of proactive systems. Aleven et al. divided reactive systems into two subtypes. The slower type of reactive system applies *task-loop adaptivity*, which focuses on selection of learning tasks or problems that are appropriate to the individual learner's states or traits. The faster version employs *step-loop adaptivity*, which pertains to changes within a task or learning activity based on the learner's momentary state.

The following sections describe three types of interventions to adapt instruction to the affective states of learners: (1) proactive design-loop adaptation; (2) reactive, task-loop adaptation; and (3) reactive, step-loop adaptation. For each type, we present examples from the literature and summarize evidence of their effectiveness. Although this review is not exhaustive, it is representative of the methods used and results from these three approaches.

---

[4] Aleven et al. (2016) also examined adaptation to other learner characteristics, including knowledge level, learning paths, learning strategies, and learning styles; however, this chapter focuses on learner affect and motivation.

A. Proactive Design-Loop Adaptation

Aleven et al. (2016) pointed out that the proactive or design-loop approach differs from the reactive methods in how training adapts to variability in student affect. In the reactive approach, instruction is adjusted to individual differences in affective states, both between different learners and within the same learner over time. In contrast, proactive systems adapt to learner similarities in affect. The intent of the proactive approach is to design (or redesign) instruction to maximize positive affective states and minimize negative states that are experienced by many or most students.

1. Examples

In this section, we describe three examples of the proactive design-loop approach. In the first two examples (Crystal Island and Virtual Schoolhouse), the systems incorporate game-like features that are explicitly designed to induce emotional states that promote engagement and positive learning outcomes. In the third example (ConfusionTutor), the system is designed to induce confusion, which is commonly regarded as a negative emotional state. Ironically, however, this negative state produces positive learning outcomes.

a. Crystal Island

The Crystal Island educational game was designed by James Lester and his colleagues at the Center for Educational Informatics at North Carolina State University. Crystal Island is a narrative-centered learning environment built on Valve Software's Source engine, the 3D game platform for Half-Life 2 (Sabourin & Lester, 2014). The game, which is a science mystery set on a volcanic island, is designed around the North Carolina eighth-grade microbiology curriculum. The program incorporates a narrative instructional approach that is designed to promote student engagement and deep learning. The developers hypothesized that "by enabling learners to be co-constructors of narratives, narrative-centered learning environments can promote the deep, connection-building meaning-making activities that define constructivist learning" (Mott, Callaway, Zettlemoyer, Lee, & Lester, 1999, p. 79).

There is evidence that the emotion-eliciting strategy designed into Crystal Island is effective in promoting learning. For instance, Rowe, Shores, Mott, and Lester (2011) studied middle school students interacting with the game and found significant gains in pre- to post-experiment knowledge and problem-solving tests. They also found larger gains in learning outcomes for students scoring higher in engagement as measured by several metrics. Also, Sabourin and Lester (2014) investigated the relationship between affect, engagement, and learning outcomes. They found positive emotions were associated with increased learning, increased interest, and more on-task behavior (i.e., less disengagement).

b. Virtual Schoolhouse

The Virtual Schoolhouse (VSH) is based on virtual-world technology and currently is serving as the synchronous distance-learning environment for the SLC. In their initial assessment of the system, Aten and DiRenzo (2014) maintained that the game-like features and interactive capabilities of the VSH potentially lead to "increased engagement compared with other distance education tools, and better learning outcomes resulting from collaborative training activities" (p. 1). On the other hand, the researchers also recognized that there are unnatural aspects of virtual-world environments, particularly the lack of face-to-face (FTF) contact that requires students to adapt their behaviors. The adaptations can be cognitively demanding and therefore result in reduced student engagement and enjoyment. To test these competing notions, Aten and DiRenzo trained sailors on a technical subject either in the VSH virtual-world environment or a traditional FTF classroom setting.

The training content was drawn from a segment of the AN/SQQ-89 Ops Course. The AN/SQQ-89 is a naval anti-submarine warfare (ASW) system for surface warships. Both initial entry and fleet returnee sailors are assigned to this introductory course, where they acquire basic knowledge of the system in the classroom and learn to operate the system in a laboratory environment. Three sequential cohorts of IE and fleet returnee sailors were individually assigned to either VSH or FTF conditions. The overall results indicated that although students assigned to the VSH condition scored marginally better than the FTF students on the classroom knowledge test and laboratory practical exam, VSH students rated the course less engaging and less satisfying than their FTF counterparts.[5] But further analysis revealed that Cohort 1 comprised only initial entry sailors, whereas Cohorts 2 and 3 were mixes of initial entries and fleet returnees. Also, Cohort 1 performance scores were lower than the two subsequent cohorts. Finally, initial entries rated the VSH less favorably than did fleet returnees. When Cohort 1 was dropped from the analysis, the relationship between the two conditions reversed: VSH scored higher than FTF in both engagement and satisfaction (Aten & DiRenzo, 2014).

In this test of VSH, student experience and instructor experience were confounded, which leads to two likely explanations of the findings, both of which may be true. First, VSH was more engaging and effective for fleet returnees than initial entries because returnees had a better understanding of the tactical context and purpose of the AN/SQQ-89 system through their fleet experience. Second, the VSH was more engaging and effective for Cohorts 2 and 3 than for Cohort 1 because instructors had become better at using the VSH. Aten and DiRenzo (2014) did not stipulate whether instructors practiced using the system before Cohort 1.

---

5 Aten and DiRenzo (2014) provide no results from statistical tests, so it is not clear whether reported differences were or were not statistically significant.

c. ConfusionTutor

The design-loop approach to adaptivity was also implemented in ConfusionTutor. As its name implies, this intelligent tutoring system was deliberately designed to induce confusion (D'Mello, Lehman, Pekrun, & Graesser, 2014). Confusion can lead to deep learning if there is discrepancy (disequilibrium) between the learner's current competency level and the skill or knowledge state required to solve the problem. To benefit from confusion, however, the learner has to recognize this discrepancy and have the informational means to correct it.

To test the notion that confusion can stimulate deep learning, D'Mello et al. (2014) developed learning sessions on concepts in critical reasoning and scientific inquiry (e.g., construct validity, experimenter bias) and tested the system using college students. Confusion was induced by introducing false or contradictory information by one or both of the agents. Learning outcomes were measured by knowledge tests administered before and after learning sessions. Results on the post-test were compared between items from experimental trials where either or both agents provided false or contradictory information and items from control trials where both agents provided true information. Results showed that students were not confused by every trial providing false or contradictory information, as measured by self-reports or by forced-choice questions occuring after the contradictory information. However, for the cases where the students were demonstrably confused, they performed better on post-test items from trials that provided false or contradictory information than trials based on the control cases that provided true information from both agents.

2. Application to Synchronous Distance Learning

The examples indicate that the proactive or design-loop approach can be effective for adapting instruction to student emotional states. Furthermore, the study by Aten and DiRenzo (2014) shows how the proactive approach can be successfully applied to the current SLC synchronous distance-learning system with two qualifications: (1) having some prior knowledge or experiential understanding of the training subject can enhance virtual instruction, and (2) instructors become increasingly effective over repeated iterations of a synchronous distance training program.

As advised by Aleven et al. (2016), the application of the design-loop approach requires data on the emotive states of learners that is time-linked to events that occur during synchronous training sessions. If the learning environment does not automatically monitor those states, there are two obvious ways to obtain those data. First, students could be polled periodically to provide self-reports on their momentary affective state. Although self-reports are often used to detect emotions in research, third-party human judgments using systematic methods, such as BROMP, are considered more valid and reliable. BROMP provides a standardized method for collecting data on emotional states, which can be linked to specific learning events (Ocumpaugh, Baker, & Rodrigo, 2015). However, BROMP should be modified for collecting data within a distance-learning environment.

As mentioned earlier, Aleven et al. (2016) pointed out that the analysis of emotion data collected in a proactive or design-loop system should focus on the similarities, not the differences between students. For example, if student affect data identify course elements that are particularly confusing or frustrating to most or many students, the course would then be redesigned to remove those sources of confusion. Similarly, if elements associated with positive emotions (e.g., enjoyment, engagement) are identified, the course could be designed to enhance those elements or recreate similar elements at different points in the course. Emotions should continue to be monitored in the redesigned course and used to adjust subsequent training in future iterations.

B. Reactive Task-Loop Adaptation

In a task-loop version of the reactive approach, the instructor or learning system detects and responds to individual learner states or traits by selecting a particular learning activity, problem, or task that is appropriate to that learner. For instance, the Cognitive Tutor is designed to make inferences about individual student knowledge state from the student's responses to the tutor, and it assigns problems intended to advance the student's knowledge level (Corbett, McLaughlin, & Scarpinatto, 2000). Aleven et al. (2016) commented that the task-loop approach to adaptivity has become a standard in commercially available intelligent tutoring systems.

1. Examples

The following are two examples of reactive task-loop adaptivity based on student emotive states. The intent of the first is to promote positive emotional states; the intent of the second is to discourage counterproductive behavior provoked by negative emotions.

a. Personal Interests

Aleven et al. (2016) cited a series of studies showing that presenting problems personalized to individual student's interests has a positive impact on academic outcomes. Presumably, adapting instruction to a student's interests promotes positive emotions, such as engagement and enjoyment. An example of these studies is the research by Walkington (2013), who developed experimental Cognitive Tutor Algebra (CTA) session problem scenarios that match students' out-of-school interests in areas such as sports, music, and movies as determined by an interest survey. Ninth-grade students were randomly assigned either to a condition where the story scenarios for problems were personalized to students' individual interest (experimental) or to a condition that presented the same problems but in the standard CTA course impersonalized scenarios (control). Results showed the students in the experimental personalized condition solved the same problems faster and more accurately than students in the control condition. The advantage to personalized problems was most pronounced for one particular skill—writing symbolic equations from the story scenarios. Even when the personalization was withdrawn after the intervention, the experimental students retained the speed and accuracy advantage in this skill, suggesting that personalization had a robust effect on student learning that persisted over time and transferred to new learning tasks.

b. Gaming the System

Baker et al. (2006) developed a component for the Cognitive Tutor that is designed to discourage gaming the system. This particular Cognitive Tutor lesson was developed for middle school mathematics education. The anti-gaming component of the tutor was enacted by an animated agent named "Scooter the Tutor." This agent, which was developed using graphics from the Microsoft Personal Assistant, reacted when the component detected that the student was gaming the system. Gaming was formally defined as "attempting to succeed in an educational environment by exploiting properties of the system rather than by learning the material and trying to use that knowledge to answer correctly" (p. 392). Previous machine-learning analyses had determined features of student performance that were predictive of gaming the system as determined by human coders: "(1) Several quick actions in a row; (2) A high percentage of errors on skills that involve popup menus (i.e., multiple choice); [and] (3) quick actions on problem steps that need a numerical answer" (Baker, Corbett, Koedinger, & Roll, 2005, p. 223). Scooter reacted to these gaming features in two ways: (1) he showed displeasure to discourage further gaming, and (2) if students got the right answer by gaming, he gave them additional exercises intended to provide another chance to cover the material that the student bypassed by gaming.

Baker et al. (2006) tested their approach on a math lesson on scatterplots. Middle school students were randomly assigned to either a lesson that employed Scooter (experimental) or to an equivalent lesson that did not employ Scooter (control). All students were tested before and after the implementation. Comparisons of experimental and control groups showed that the implementation of Scooter resulted in only a marginal reduction in gaming and no improvement in learning. But it was noted that gaming was observed in only a small percentage of the students. Students in the experimental group were then divided into those receiving either more or fewer additional problems, an indication of the level of gaming in that group. It was found that students assigned more additional problems (presumably those who were gaming the system) started at a lower level of performance on the pretest than those assigned fewer problems. However, the students having more additional problems exhibited larger pre-post gains than did the students assigned fewer problems, effectively equaling the students who were not gaming on the post-test.

2. Application to Synchronous Distance Learning

Results from the two example studies suggest that the task-loop approach—that is, adjusting learning tasks and problems to emotional states—can have a positive effect on learning outcomes. The problem is that task-loop adaptivity calls for learning adjustments at the individual level and thus has limited applicability to the collective training situation in a synchronous training environment. However, it could be possible to make small adjustments if the instructor has the means to determine the emotional states of individual students in real time. For example, let's suppose that an instructor has reason to believe that students are likely to be confused at a certain point in the lesson. To determine whether or not they are confused, the instructor could poll students and ask for self-reports. But research on the ConfusionTutor suggests that a more reliable and valid approach would be to ask questions that probe potential points of confusion or inattention (D'Mello, Lehman, Pekrun, & Graesser, 2014). Ideally, these probes would be prepared ahead of time and require only short answers that could be summarized and displayed immediately on an instructor dashboard without attribution to particular students. In the case where a significant proportion of students answer incorrectly, the instructor could choose to repeat the relevant section or elaborate on the topic. If only a few students answered incorrectly, the instructor could explain why those particular answers were incorrect, thereby removing the source of confusion, and move on. If those few students consistently respond incorrectly, the instructor may prescribe individualized instruction outside the collective synchronous learning environment. If the errors were more widespread, however, the instructor could opt to redesign that part of the course, which would constitute a more proactive or design-loop approach.

This somewhat contrived example illustrates how aspects of task-loop adaptivity could be applied to synchronous distance learning. However, there are at least two major obstacles to making that happen: First, there is the technical challenge of providing real-time information on the affective states of individual students (see Chapter 3). Furthermore, this information must be provided in a form that can be understood and acted on by the instructor. Second, instructors must decide whether and how to react to these states. In other words, they must be prepared to make changes to their lesson plans on the fly. We doubt that there are many instructors who would be willing or able to spontaneously adapt training to the students' affective states, which are in perpetual flux.

C. Reactive Step-Loop Adaptation

Step-loop adaptivity refer to instructional adjustments based on moment-to-moment changes in student state within a learning task or activity, as inferred by learner actions or states. These inferences are often based on conversations between learning system agents and students in natural language. To make rapid inferences based on incomplete and imperfect information requires state-of-the-art computing technology. As a result, systems that employ step-loop adaptivity have only been tested in controlled laboratory settings.

1. Examples

The following are three examples of individual learning systems that react to students based on their momentary affective state.

a. Affective AutoTutor

The first example of reactive step-loop adaptivity is provided by D'Mello et al. (2010). In this study, a version of the AutoTutor was designed to detect and react to student affective states of boredom, confusion, and frustration. The system monitored multiple cues: conversations between student and the tutor agent, gross body language, and facial features. The animated tutor agent reacted to cues to encourage and maintain student engagement, boost student self-confidence, and pique student interest. The agent's face could display emotions, such as skepticism, approval, disapproval, enthusiasm, surprise, empathy, as well as no particular emotion (neutral). The tutor responses were tailored to the student's current and previous affective state, the conceptual quality of the student's immediate response, and the student's general ability level determined over the entire session. For instance, if the AutoTutor system determined that a student (1) had been performing generally well, but less so on a current problem, and (2) had previously been frustrated but is currently showing signs of boredom, then the tutor might respond, "Maybe this topic is getting old. I'll help you finish so we can try something new."

For the D'Mello et al. (2010) study, a 60-minute session on computer literacy was developed for the AutoTutor system. Eighty-four college students were assigned to either a session that provided emotional support as described above (experimental) or a session that did not include emotional support (control). Both groups were given knowledge pretests and post-tests before and after two sessions on different topics in computer literacy. Students in both groups were defined as having either high or low prior knowledge determined by a median split on their pretest knowledge scores. The two major findings were (1) the affect-sensitive version of AutoTutor was more effective than the regular tutor for low-domain knowledge students in the second session, but not in the first session; and (2) the affect-sensitive version never had a positive effect on high prior-knowledge students but was actually detrimental in the second session compared with the regular version used in the control group.

b. Gaze Tutor

In the second example, D'Mello, Olney, Williams, and Hays (2012) developed an intelligent tutoring system capable of dynamically detecting and responding to student boredom and disengagement. The capability was provided by a commercial eye tracker built into the Guru intelligent tutor, a system that explicitly models expert human tutor behaviors and student-tutor dialogues. The criterion for disengagement was not looking at the tutor agent or the instructional image for more than 5 seconds. When the system detected student disengagement, the tutor agent reacted by immediately responding with statements such as “please pay attention.” The Guru tutor was programmed to deliver four sessions on biology topics: Golgi body, cytoskeleton, phases of mitosis, and ecological succession.

For the D’Mello et al. (2012) study, 48 college students received instruction on 2 of these topics with the gaze-reactive capability activated (experimental condition) and two topics using the standard Guru tutor without the gaze-reactive capability (control). All students were tested before and after training using questions to assess students’ knowledge of shallow facts and deeper questions that required causal reasoning and inferences. Results indicated that fully one-third of the students (16) never reached the disengagement criterion in the experimental condition, and they were not included in the analyses. Analyses of the remaining 32 students indicated that the gaze-sensitive dialogues were effective in reorienting students’ attention toward the important parts of the screen presentation. For the content learned in the experimental condition, there were larger gains in those questions assessing deep learning. The findings also showed that the experimental gaze-reactive capability had a larger effect on students with higher academic aptitude, as measured by self-reports of ACT or SAT scores, than on students with average or lower ability scores. Results showed that high-aptitude students were more effective in reorienting their gaze when instructed to do so, and these same students showed greater knowledge gains under experimental conditions. D’Mello et al. (2012) speculated that high-aptitude students were more skilled at reallocating attentional resources than lower aptitude students.

c. TC3Sim Serious Game

DeFalco et al. (2018) conducted a multi-experiment study employing TC3Sim, a serious game used to train U.S. Army personnel on providing combat casualty care. The game was integrated with the Generalized Intelligent Framework for Tutoring (GIFT) to detect affective states and provide appropriate feedback to trainees. By examining the entire process of developing an affect-sensitive system, the DeFalco study provides “a template for future research to examine the sensitivity of instructional methods and to conduct validation studies or to examine the impact of various adaptive instructional tools or methods” (Sottilare, Baker, Graesser, & Lester, 2018).

The study was conducted as three sequential experiments using cadets at the U.S. Army Military Academy as subjects. Experiment 1 provided the baseline for the subsequent investigations. Two types of posture sensors and interactions with the game were used to detect the affective states of boredom, confusion, engaged concentration, frustration, and surprise as determined by human observers using the BROMP technique.

Results indicated that the best results were obtained for detecting frustration from student interactions with the TC3Sim system. These interactions included conditions and behaviors such as whether the student was under fire, taking cover, or separated from his or her unit. Another important finding from Experiment 1 was that observed frustration was negatively correlated with performance on the TC3Sim game, confirming its relevance to instruction.

In Experiment 2 (DeFalco et al., 2018), the frustration-detection system triggered three different types of feedback to students, appealing to (1) the perceived value of the learning activity, (2) the student's social identity, or (3) the student's self-efficacy to achieve a learning goal through effort. Comparisons of knowledge tests administered before and after game training indicated that all three motivational methods produced larger learning gains than the neutral feedback and no-feedback control groups. Of the three types of motivational feedback messages, however, the self-efficacy message produced the largest gains. Also, cadets scoring lower on the Short Grit Scale, a measure of self-motivation, showed large learning gains when provided motivational feedback. In contrast, those scoring higher on the Short Grit Scale showed negative, but non-significant declines in learning when given motivational feedback.

In Experiment 3, DeFalco et al. (2018) tested two different frustration-detection systems: one based on interactions developed in Experiment 1 and a posture-sensing system based on the Microsoft Kinect product. When these sensing systems detected frustration, the feedback system delivered a self-efficacy message. These two frustration-sensing systems were compared to a control condition where self-efficacy messages were sent to the student on a regular schedule, regardless of the student's affective state. The results showed significant learning gains in all three groups, but no differences in those gains between groups. A post hoc analysis revealed that the Kinect sensor failed to detect any frustration and consequently delivered no feedback, effectively making this a second no-feedback control group. Comparisons between groups showed no differences between the group provided feedback contingent on their interactions in the game and the two other de facto control conditions, which was a finding in contradiction to Experiment 2.

2. Application to Synchronous Distance Learning

Findings from research provide some evidence for the effectiveness of reactive step-loop adaptivity, but the effectiveness is highly conditioned on other factors. The results of D'Mello et al. (2012) suggest that there is some advantage to step-loop system if adaptations are based on a single relevant emotional state (engagement) as determined by a relatively unambiguous cue (gaze). However, the advantage may accrue only to those high-aptitude learners who are more capable of controlling their attention. In contrast, if step-loop systems are based on multiple emotions that are ambiguously or inaccurately detected, frequent adaptive feedback may be less helpful and even annoying to high-knowledge students (D'Mello et al., 2010) or for students more motivated to persevere (DeFalco et al., 2018).

Apart from questions about its effectiveness, we think that the step-loop approach is fundamentally incompatible with synchronous distance learning. The limitations of the task-loop approach, discussed previously, apply here. In addition, the step-loop requirement to adapt to the momentary actions of individual students is practically impossible to implement in a collaborative distance-learning environment.

5. Conclusions and Recommendations

In this chapter, we summarize our main conclusions and provide practical recommendations for applying emotion-recognition science and technology to synchronous distance learning. Our conclusions and recommendations are based on our review of the research on affect-sensitive instruction and on our interviews of experts in the field. Appendix A provides a summary of those interviews.

The chapter is divided into two parts. The first presents conclusions concerning the theories and technologies of emotion recognition and affect-sensitive instruction. The second describes specific recommendations for applying emotion-recognition technologies to the synchronous distance-learning environment used, or planned to be used, at the SLC.

A. Conclusions

1. Successful Implementation Requires Solving Three Separate Problems

Our expert interviewees were remarkably consistent in their comments that the development of affect-sensitive instruction can be divided into three separate but interrelated problem goals: (1) assess the affective state, (2) understand the affective state, and (3) determine the learning intervention. The effects of the intervention are then assessed, which restarts the process. The three goals can be viewed as an iterative loop as depicted in Figure 6. To close the loop of affective computing, the instructional developer needs to have a solution for all three goals. Each of these problems is described below:

- *Goal 1: Assess affective state.* Perhaps the most fundamental problem, the instructor or the learning system must be able to detect a student's affective state. The focus of research has been on learning-centered emotions (e.g., frustration or confusion). Experts agreed that research and development efforts have been aimed at solving this problem, and considerable progress has been made. For example, as documented in Chapter 2, much progress has been made using eye-tracking as a unimodal signal for affect and facial features paired with contextual cues (e.g., dialogue) as multimodal cues.
- *Goal 2: Understand the affective state.* This problem requires modeling the emotional processes and their effects on learning. Experts regard this as the most difficult problem of the three. Chapter 3 documents some points of agreement among models on the basic functions and processes of emotions, but there is a

dearth of knowledge concerning the impact that these functions and processes have on learning.

- *Goal 3: Determine the learning intervention*. Once the affective state is detected and its effects on learning are understood, the instructor or instructional system selects an appropriate instructional response intended to support training. Some progress has been made on this problem, but it is limited by the lack of understanding of affective states (Goal 2).

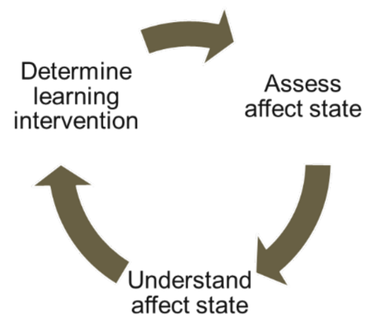


Figure 6. Affect-Sensitive Instructional Loop

2. Emotion Recognition is Based on Well-Established Science Though Some Theoretical Issues Remain Contentious

Our review of seven theories of emotions span over 50 years of research and reveal a variety of scientific concepts and metaphors. Nevertheless, Ekman's Basic Emotion Theory (BET, 1992) remains the "central narrative" of the science of emotion (Keltner, Sauter, Tracy, & Cowen, 2019), providing a point against which other theories are compared. Some of the theoretical deviations from that central narrative highlight some of the major points of contention:

- Emotions are not discrete states as they are depicted by Ekman; rather, emotions are determined by a combination of continuous characteristics, such as valence and arousal.
- There are more, perhaps many more, than the six basic emotions enumerated by Ekman.
- In recognizing and understanding an affective state, the context of an emotional expression is as important, or more important, than the expression itself.

Despite these points of disagreement, researchers concur on some basic capabilities of emotions and the functions that they serve. In particular, four points of agreement have implications for emotion-recognition technology:

- Emotions arise through an unconscious appraisal process.

- Emotional experiences are brief and transitory.
- There are distinct and detectable cues for determining the emotions that someone is experiencing.
- Humans are quite good at recognizing and interpreting the emotional cues experienced from other people.

Thus, despite some points of theoretical contention, emotion-recognition technology is built on a relatively broad and deep scientific base.

3. The Evidence is Inconclusive Whether Affect-Sensitive Instruction Improves Learning Outcomes

There is some evidence that affect-sensitive instruction can improve learning outcomes, but there is also evidence that it has no effect and even a negative effect on learning. The inconsistency in findings appears to be due to a number of moderating factors, some of which are discussed below.

a. The Relation Between Affect and Learning is Indirect

D'Mello, Blanchard, Baker, Ocumpaugh, and Brawner (2014) point out that it is unlikely that there are direct links between affect and learning outcomes, but instead, affect indirectly relates to learning by modulating cognitive processes. As an example, they argue that anxiety is unlikely to directly cause poorer learning; rather, anxiety negatively affects cognition in that anxiety-related thoughts, such as fear of failure, are consuming working-memory resources, reducing the amount available for ongoing cognitive processes If researchers failed to find a relationship between anxiety and learning outcomes, the results could be because there is no effect of anxiety on working memory, there is no effect of working memory on learning, or both. In general, adding links to the causal chain decreases the probability of detecting a significant relationship between anxiety and learning.

b. Affect Interacts with Individual Differences

Our brief review in Chapter 4 indicated that the effects of affect-sensitive interventions may vary for people with different dispositions and abilities. For instance, results from DeFalco et al. (2018) suggested that motivating messages had a positive effect on students scoring low on resiliency (also known as grit). But the same messages had a deleterious effect on students with high resilience. If such individual differences are not accounted for, they can lessen, or even, nullify the overall effects of an intervention.

c. This is a Fundamentally Difficult Problem

As argued in the previous conclusion, the problem of improving learning outcomes through affect-sensitive instruction comprises actually three separate but interrelated problems. All three problems are somewhat ill-defined in that they have no algorithmic solutions. This situation has some characteristics of a “wicked problem” that defies traditional solutions.

Thus, demonstrating the training effectiveness of an affect-sensitive intervention is no mean feat. However, some solace and lessons learned can be gleaned from the handful of published studies demonstrating that affect-sensitive instruction can indeed enhance learning outcomes under certain circumstances.

4. Student Disengagement Is Not Necessarily Counterproductive

Increased learner engagement is often cited as evidence for the effectiveness of an affect-sensitive learning program—many demonstrations show that engagement leads to positive learning outcomes. While agreeing that engagement is related to learning outcomes, James Lester (one of our expert interviewees) pointed out that students cannot remain engaged indefinitely. Lester’s argument alluded to the Yerkes-Dodson Law. As represented in Figure 7, this law stipulates that increased arousal improves performance up to a point, but past that point increases can actually hurt performance. Although this analogy to the effects of arousal is intuitively appealing, this relationship has not been demonstrated for learner engagement.

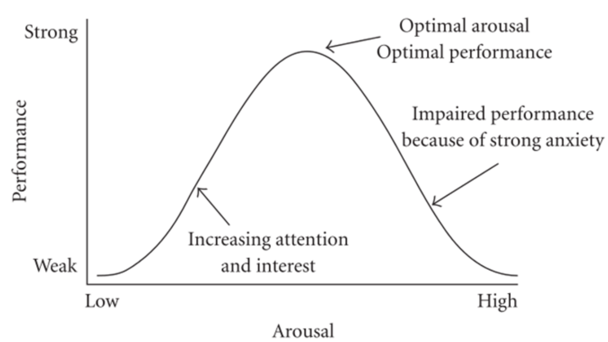


Figure 7. Yerkes-Dodson Law

The larger point Lester was making was that student disengagement is not necessarily counterproductive. It may be the student’s method to prevent or dissipate over-arousal. Lester refers to this strategic disconnection with learning as “constructive disengagement.” Although the theoretical explanation of constructive disengagement is speculative, the implication is clear: distance learning must schedule breaks within learning sessions and allow students to temporarily disengage from the learning task.

5. Emotional Traits, as Well as States, Are Also Relevant

Affect-sensitive training systems are designed to react to the momentary emotional states of learners. These systems should also take into consideration the learner's emotional traits. These traits are relatively stable predispositions toward a particular emotional experience. When we make comments that "Joe is an angry man" or "Linda is a happy child," we are implying that they possess certain emotional traits.

Emotional traits can influence how a learner may react to an affect-sensitive intervention. An example of a technology that is reactive to affective states and traits is the TC3Sim Serious Game (DeFalco et al., 2018), which is described in more detail in Chapter 4. In the DeFalco et al. (2018) study, the research team administered the Short Grit Scale before the study. This personality scale measures a person's tendency to persevere on tasks and focus on long-term goals. Subjects were assigned to conditions such that they were or were not provided motivational messages when the system detected that the subject was frustrated. The findings indicated that learners scoring low on grit showed learning gains when provided motivational messages. In contrast, learners having high grit scores showed learning decrements when provided motivational feedback. The researchers speculated that the high-grit learners regarded the messages as unnecessary and even annoying, and the negative reaction may have led to frustration and disengagement.

Such findings suggest that if learning system developers know the dominant emotional traits of their student population, they can better predict the students' reaction to affect-sensitive interventions. Thus, it would be useful to collect emotional trait data along with data on affective states.

B. Recommendations

1. Use Multiple Modes of Affect Detection

The probability of accurately detecting an emotion using a single mode of detection is low, but with advances in technology, unimodal detection is not out of the question. (Multimodal detection can provide a stronger signal, however.) Further, there are differences in accuracy between individual modes. For instance, eye tracking, mouse tracking, and language (speech and text) are the strongest signals. Also, unimodal research on eye-tracking and natural-language processing show promise. On the other hand, most researchers agree that facial movements provide a noisy and unreliable signal of emotional state, despite decades of research on this detection mode.

Researchers are turning to combining multiple recognition modes to increase detection probability. Widely disparate modes provide complementary information sources. For instance, facial recognition combined with student data and text dialogue can provide a strong multichannel source. To be effective, however, mode combinations should include student-specific data, like log data and exams for accurate detection. The ways different signals are combined are widely varied across the academic literature. Outside a lab setting, obvious constraints for multimodal detection include appropriateness of implementing signals in a classroom (i.e., feasibility, distraction), and cost.

In addition, the type of signal one uses (sensor free vs. sensor based) contributes to the strength of the signal. For example, facial recognition (sensor based) is a noisy signal, but combining facial recognition with dialogue (sensor-free) could provide a substantial increase in affect detection. Log and dialogue data are particularly strong sensor-free signals that, when combined with sensor-based signals (e.g., eye tracking or mouse tracking), provide one of the strongest measures of affect and attention.

In summary, we recommend using more than one mode to detect emotions. At the same time, we note that collecting and fusing signals from different input modes is technically challenging. Thus, we temper the recommendation to use multiple modes with the phrase "to the extent practical or feasible."

2. Focus on Most Relevant and Detectible Emotions

Our review of theories and taxonomies identified 43 affective states that are potentially related to learning in general. This list was derived mostly from theoretical considerations, not empirical data. Also, the list does not consider important technical factors, such as frequency of occurrence, detectability, and specific implications for instructional interventions. Thus, this list likely includes emotions that are irrelevant to learning.

In contrast, Graesser, D'Mello, and colleagues have focused their efforts on those emotions that have direct relevance to instruction in intelligent tutoring systems, the so-called learning-centered emotions. For instance, D'Mello's (2013) meta-analysis of studies employing individual advanced learning technologies reveals that only six states are detected with any frequency: *engagement/flow*, *boredom*, *confusion*, *curiosity*, *happiness*, and *frustration.* However, this focus on individual instruction omits consideration of social emotions (e.g., compassion, empathy, envy, social anxiety) that potentially pertain to collective or collaborative training employed in many synchronous distance-learning situations. Thus, this list may not include emotions relevant to the SLC learning environment.

There is no agreed-on set of emotions that affect-sensitive learning systems should detect and respond to. Emotions of interest depend largely on the capability of the system and goals of the learning content. Nevertheless, the choice of emotions should follow two general guidelines:

- The emotion should be detectible by the learning system. "The learning system" includes the human instructor in synchronous distance learning.

- The emotion has some direct or indirect effect on learning outcomes.

3. Use Student Audio and Video Feeds to Detect Emotions

Many virtual classrooms provide video and audio input of students as they interact with the instructor and each other. A number of our expert interviewees suggested that the video and audio feeds from students provide the least expensive and most practical solution for affect-sensitive instruction in a synchronous distance-learning environment. Nevertheless, the solution requires that developers address two separate challenges.

The first challenge is that instructors must learn to recognize the affective states of students as they occur. As discussed in Chapter 2, we all have some innate capability to recognize emotions in others, but not in a standardized way and using common language. To train instructors to accurately and reliably recognize emotions, our interviewees suggested that instructors adopt the Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP, Ocumpaugh, Baker, & Rodrigo, 2015), which is regarded as "gold standard," to validate other emotion-recognition technologies. The system is sensitive to multiple emotion-recognition modes, including movement cues like body posture and position on chair, in addition to facial and verbal cues. However, it is not known the extent to which such signals are recognizable through the typical audio and video feeds employed in videoconferencing.

The second challenge is that instructors need to know whether or how to intervene once they detect the emotional states of their students. There are production rules (if-then statements) for responding to affective states experienced in AutoTutor (e.g., D'Mello et al., 2010). Note, however, that these rules were explicitly developed for individual advanced learning technology systems. Research is needed to determine the extent to which these rules and other algorithms apply to synchronous distance learning. This research would provide the foundation for instructional intervention training. For maximal effectiveness, intervention training should be integrated with training to recognize emotional states.

In short, the audio-video feed from students in virtual classrooms provides a straightforward method of capturing emotions and engagement during learning. But to make use of this data source, instructors must be trained to reliably recognize emotional states and respond appropriately.

4. Use Proactive Approach to Design Courses That Induce Positive Emotions and/or Impede Negative Emotions

D'Mello and Graesser (2015) proposed that affect-sensitive learning systems are designed to intervene either proactively or reactively. Proactive systems are those that are designed before the fact to induce positive emotional states or impede negative states. In contrast, reactive systems are designed to detect and respond to affective states as they arise. The problem with reactive systems is that it is practically impossible to adapt instruction to the momentary actions of individual students in a collaborative distance-learning environment.

In Chapter 4, we reviewed research on proactively designed affect-sensitive learning systems that have been demonstrated to enhance engagement and learning outcomes in individual tutorials (D'Mello, Lehman, Pekrun, & Graesser, 2014) and game-based instruction for classrooms (Rowe, Shores, Mott, & Lester, 2011; Sabourin & Lester, 2014). In addition, Aten and DiRenzo (2014) contended that the SLC Virtual Schoolhouse was proactively designed to promote student engagement and learning, but their data did not conclusively support that conclusion.

Although proactive systems do not use data on student affect to react in real time, proactive programs typically do use affect data to design or redesign a course before implementation. Aleven et al. (2016) suggest that training developers collect affect data while students use the learning system to identify points where students are confused or frustrated. Developers then use the data to redesign the course to ameliorate those problems. Affect data should continue to be collected after the redesign is implemented to assess whether the changes work as intended.

To design an affect-sensitive capability for a synchronous distance-learning environment, we suggest a proactive approach where features that induce positive emotional states or impede negative states are designed into the system before the fact. Although such systems do not react to student affect in real time, proactive systems should be designed and assessed in accordance with student affect data.

5. Incorporate Instructional Features to Enhance Engagement

Several of our expert interviewees recommended adding minor instructional changes or tweaks that are designed to build in moments of engagement into courses. Such relatively inexpensive course modifications are viewed as an alternative to complete course redesign as discussed in the previous recommendation. Example features include the following:

- **Periodic questioning.** Instructors can use brief questioning periods to poll students on key concepts. The intent is to quickly gauge learning and engagement from students. Questions should require some thinking, but not be too difficult. The aggregated and anonymous responses should be immediately available to the instructor in real time so that the instructor can make an assessment in real time and adapt instruction if needed.
- **Small-group discussions.** Students should be broken into smaller groups to discuss the subject material. Such small-group interactions enhance student

engagement. Instructors should listen into these discussions to assess student engagement and learning.

- **Shorter lectures and longer practice periods.** Consistent with a previous conclusion, some of our interviewees mentioned that students cannot focus on difficult material for longer than 20 minutes in lectures. On the other hand, periods where students actively practice problem-solving skills can enhance engagement. Thus, lessons should be designed with shorter lectures and longer practice periods.

In short, our advice is to take advantage of such cost-effective tweaks to increase student engagement in synchronous distance learning.

6. Machine-Learning Models Show Promise, but Have Caveats

Ongoing research indicates that machine-learning methods can be used to accurately detect emotional states from multiple data streams. Among the strongest signals are audio, text, eye tracking, and mouse tracking.

Although promising, machine-learning methods are complex to develop (see Figure 8). Machine-learning systems also require lots of data for training and validation, and crucially, the data need to be highly specific to the application context. Much of these data are from human coders, who need to be appropriately trained. Selecting and training coders is costly and time consuming. While the machine-learning pipeline is the approach that most researchers implement to collect, annotate, and analyze data, for SLC to implement such a method would require the help of a subject-matter expert to implement this method. But the success of the method is not guaranteed.

Perhaps the most limiting feature is that machine-learning models are typically developed based on data from single learners, and those individual models don't generalize to others very well. In fact, our research and expert interviews highlighted that the best machine-learning approach is creating student-specific models. There are significant limitations to this approach, but there may be some hope in methods under development by Ben Nye at the Institute for Creative Technology. Nye seeks to derive machine-learning models based on groups of individuals that potentially can be generalized to others within those groups.

Though promising, machine-learning approaches are time intense, are costly (human coders, signal-detection instruments), and require significant ground truth data. They are currently employed primarily in laboratory settings and are not ready for classroom applications.

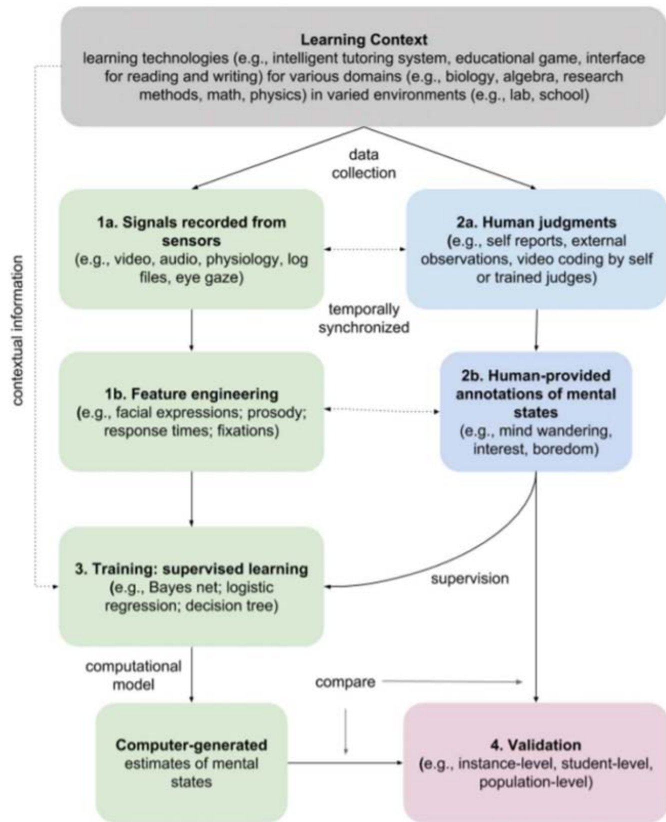


Figure 8. Steps Involved in Building a Machine-Learning Affect-Recognition System

Appendix A.
Expert Interviews

The IDA team interviewed 12 subject-matter experts who actively work in the field of affective computing. The experts we interviewed included professors, research scientists, and research psychologists working in a variety of fields (e.g., computer science, human factors, signal processing, and cognitive science). Table A-1 lists the people who were interviewed.

Conducted in the fall of 2019, the interviews included visits to the University of Colorado, Boulder, the Signal Analysis and Interpretation Laboratory (SAIL) at the University of Southern California (USC), The Institute for Creative Technologies (a University Affiliated Research Center located in Los Angeles), and I/ITSEC (Interservice/Industry Training, Simulation and Education Conference) in Orlando, Florida. The information gathered from the conducted interviews have been incorporated into this report, and here we summarize the interviews.

SME interviews focused on the IDA team gaining a deeper, more nuanced understanding of the state of the art in terms of affective computing and learning as it relates to SLC’s unique problem space. In addition, we were also interested in learning about current trends in the field, in particular, specific avenues of interest the SMEs were focusing on in their research, and gaining an overall better understanding of the field at hand (e.g., what’s working, what’s failing).

Overall, almost every SME we interviewed asserted that knowing a student’s affective state does not equate to learning gains and is only one piece of the affective-loop (step one). Furthermore, depending on the dynamics of the student population, affective state might not have any correlation with learning gains at all. For example, adult learners might be sad or angry while learning, but still engaged or focused and able to perform, but the opposite might be true for children. In addition, certain student populations (e.g., those in the military) might not be particularly effusive to begin with and therefore detecting affective state is not particularly fruitful. Along the same lines, displaying affect is contextually dependent, and people deceive and hide their affective states all the time. For example, a student who is frustrated or confused may not display this to avoid social stigma in classrooms. One researcher strongly supported this view and noted that research should not look at facial expressions as an indicator of affect since humans can control and deceive others with facial expressions.

Table A-1. List of Experts Interviewed for IDA Project

Name	Title	Area of Expertise	Location
Sidney D'Mello	Associate Professor	Affective and attentional computing, multimodal interaction	Institute of Cognitive Science, Computer Science, Psychology and Neuroscience, UC Boulder
Shri Narayanan	Professor	Electrical engineering, signal processing	SAIL, USC
Jonathan Gratch	Director of Virtual Humans Research; Research Professor	Virtual humans, computational models of emotion, cognition and emotion	Institute for Creative Technologies (USC-ICT)
Mohammad Soleyami	Research assistant professor	Computer science, behavior understanding, psychological signals	USC-ICT
Benjamin Nye	Director of Learning Science	Intelligent tutoring systems, intelligent agents, engineering	USC-ICT
Keith Brawner	Senior Researcher and Project Manager	Learning systems, AI, real-time algorithms, intelligent tutoring	U.S. Army Futures Command-Simulation and Training Technology Center
Jeanine DeFalco	Research Scientist	Artificial intelligence, education, intelligent tutoring systems	U.S. Army Futures Command-Simulation and Training Technology Center
Lauren Reinerman-Jones	(former) Director of Prodigy, Associate Professor	Cognitive Psychology	N/A
Randall Spain	Research Psychologist	Psychology, education, advanced training technologies	Department of Computer Science, Center for Educational Informatics, NCSU
Bradford Mott	Senior Research Scientist	Computer science, AI, game-based learning	Department of Computer Science, Center for Educational Informatics, NCSU
Jonathan Rowe	Research Scientist	Computer science, human-computer interaction, AI, game-based learning	Department of Computer Science, Center for Educational Informatics, NCSU
James Lester	Director of Center for Educational Informatics	Education, AI-augmented learning	Department of Computer Science, Center for Educational Informatics, NCSU

With regard to detecting emotions, some researchers noted that detecting valence (positive or negative) or arousal (calm or excited/agitated) is much easier than detecting specific affective states. For this reason, it was noted that focusing efforts on certain emotions like confusion or frustration is more fruitful than focusing on the full suite of learning-centered emotions, for two reasons: confusion and frustration are thought to be the most important emotions related to learning, and they are also high on the valence/arousal spectrum, making them easier to detect. In general, because detecting emotions is difficult, and there is a lot of noise in recorded data (especially biometric; e.g., facial expressions are particularly difficult), simplifying the problem space is recommended.

Some researchers strongly supported specific coding systems, like BROMP, as a method of collecting ground-truth emotions of students and using these human-generated data to feed machine-learning models. While data for automatic affect-detection models need to be generated, other researchers suggested that student-specific data be collected to move toward creating single-student models. Student-specific data include class performance (e.g., quizzes, attendance, exam grades); single-student level models are more accurate and combine student performance and physiological data to predict performance and learning gains. Any coding of data should be more student-specific.

The type of coding done to acquire data is no doubt also related to the types of technologies most affective-computing researchers use. In particular, researchers employed various techniques to collect ground-truth data. Techniques include BROMP, FACS, peer-to-peer coding (i.e., one peer codes another peer), self-report, emote aloud, among others. These data are then used in machine-learning models to develop an automatic affect sensor. Machine-learning algorithms were adopted, depending on the particular problem set of researchers (e.g., classifies and support-vector machines, being used to distinguish/classify different affective states).

In terms of psychological approach, the researchers with psychology backgrounds focused mostly on learning-centered emotions and the role of frustration and confusion in learning (e.g., cognitive disequilibrium being the most widely adopted approach). Those with engineering backgrounds didn't adhere as strongly to certain psychological theories (i.e., if a theory fit or helped an engineering problem, it was adopted). Those with engineering and psychology backgrounds saw psychological theory playing a central role in effective affective-algorithm development. In fact, it's well supported in the literature that any successful affective-computing approach needs to view the human and computer as one interacting system (Calvo & D'Mello, 2010).

Turning to the more effective signals, almost unanimously, researchers agreed that eye-tracking was the best measure of attention and therefore engagement; where eyes are attending to is a direct measure of what they're comprehending. Of course, there are individual difference (e.g., looking up and to the right while thinking can be interpreted as not paying attention), but overall, this is the gold star for engagement and attention. Mouse-tracking was the second best signal suggested to measure attention and engagement and is a good proxy for eye-tracking; that is, how someone is interacting with content on a computer screen is indicative of where they are attending to. Finally, speech or text is the third best signal, also being a good signal for measuring affect (e.g., detecting frustration via prosodic features or confusion via text responses). Text alone has been a signal used throughout the AutoTutor automatic affect-detection enterprise as either the single source of affect or part of a larger suite of affect signals. Of course, all SMEs interviewed used other biometric sensors in their research (e.g., wearables like FitBit,).

References

Ai, H., Litman, D. Forbes-Riley, K. Rotaru, M. Tetreault, J., and A. Purandare (Eds.) (2006). *Using system and user performance features to improve emotion detection in spoken tutoring dialogs*. University of Pittsburgh.

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. In R. E. Mayer, & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (2nd ed., p. 522). New York: Routledge.

Arroyo, Ivon, Burleson Cooper, B. Woolf, K. Muldner, & Christopherson, R. (2009). Emotion sensors go to school. In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A. Graesser (Eds.), *Artificial intelligence in education: building learning systems that care - from knowledge representation to affective modeling* (pp. 17-24). Washington, DC: IOS Press.

Aten, K., & DiRenzo, M. (2014). *Assessing the potential of virtual worlds for Navy training: Cogntive learning processes and outcomes in the Virtual School House* (*NPS-GSBPP-15-002*). Monterey, California: Naval Postgraduate School.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Roll, I. (2005). Detecting when students game the system, across tutor subjects and classroom contexts. In L. Ardissono, P. Brna, & A. Mitrovic (Eds.), *User Modeling* (*UM*) *2005 Proceedings* (Vol. 2538, pp. 220-224). Berlin, Heidelberg, Germany: Springer.

Barrett, L. B., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology*, *74*(4), 967–984.

Baker, R. S., Corbett, A., Koedinger, K. R., Evenson, S., Rolf, I., Wagner, A. Z., . . . Beck, J. E. (2006). Adapting to when students game an intelligent tutoring system. In M. Ikeda, K. D. Ashley, & T. W. Chan (Eds.), *Intelligent Tutoring Systems* (*ITS*) *2006*. *Lecture Notes in Computer Science* (Vol. 4053, pp. 392-401). Berlilne, Heidelberg, Germany: Springer.

Baker, Ryan S. J. d., & Jaclyn Ocumpaugh. (2015). Interaction-based affect detection in educational software. In R. A. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford handbook of affective computing*. New York: Oxford University Press. Retrieved from: https://doi.org/10.1093/oxfordhb/9780199942237.013.009

Baker, R. S. J. d., D'Mello, S. K., Rodrigo, 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. https://doi.org/10.1016/j.ijhcs.2009.12.003

Barrett, L. F., Adophs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest, 20*(1), 1–68. doi:10.1177%2F1529100619832930

Bartneck, C., Lyons, M., & Saerbeck, M. (2008). The relationship between emotion models and artificial intelligence. *Proceedings of the Workshop on the Role of Emotion in Adaptive Behaviour and Cognitive Robotics in affiliation with the 10th International Conference on Simulation of Adaptive Behavior: From Animals to Animates* (*SAB 2008*), (pp. 16-27). Osaka, Japan.

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., & Zhao, W. (2015). Automatic detection of learning-centered affective states in the wild. In *Proceedings of the 20th International Conference on Intelligent Interfaces*, 379–88. https://doi.org/10.1145/2678025.2701397

Broekens, J., Hudlicka, E., & Bidarra, R. (2016). Emotional appraisal engines for games. In K. Karpouzis, & G. N. Yannakakis (Eds.), *Emotion in games. Socio-affective computing, Volume 4* (pp. 215-232). Springer. doi:10.1007/978-3-319-41316-7

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. https://doi.org/10.1109/T-AFFC.2010.1

Calvo, R., A., D'Mello, S., Gratch, J., & Kappas, A. (Eds.) (2015). *The Oxford handbook of affective computing*.

Clore, G. L., & Ortony, A. (2013). Psychological construction in the OCC model of emotion. *Emotion Review, 5*(4), 335–343. doi:10.1177/1754073913489751

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures.* Washington, DC: The National Academies Press. https://doi.org/10.17226/24783

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. https://doi.org/10.1007/s11257-009-9062-8

Corbett, A., McLaughlin, M., & Scarpinatto, K. C. (2000). Modeling student knowledge: Cognitive tutors in high school and college. *User Modeling and User-Adapted Interaction, 10*, 81-108.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences, 114*(38), E7900-E7909. doi:10.1073/pnas.1702247114

Cowen, A. S., & Keltner, D. (2018). Clarifying the conceptualization, dimensionality, and structure of emotion: A reply to Barrett and colleagues. *Trends in Cognitive Science, 22*(4). doi:10.1016/j.tics.2018.02.003

Cowen, A. S., & Keltner, D. (2019). What the face displays: Mapping 28 emotions conveyed by naturalistic expression. *American Psychologist*. doi:http://dx.doi.org/10.1037/amp0000488

Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist, 74*(6), 698-712. doi:DOI:10.1037/amp0000399

Cowen, A., Sauter, D., Tracy, J. L., & Keltner, D. (2019). Mapping the passions: Toward a high-dimensional taxonomy of emotional experience and expression. *Psychological Science in the Public Interest, 20*(1), 69-90. doi:10.1177/1529100619850176

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), 777–88. https://doi.org/10.1080/02699930701516759

D'Mello, S., & Graesser, A.. (2010a). Mining bodily patterns of affective experience during learning. In R. S. J. de Baker, A. Merceron, & P. I. Pavlik Jr. (Eds.) *Educational Data Mining 2010, The 3rd International Conference on Educational Data Mining, Pittsburgh, PA, USA, June 11-13, 2010. Proceedings* (pp. 31–40).

D'Mello, S., & Graesser, A.. (2010b). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction, 20*(2), 147–187. https://doi.org/10.1007/s11257-010-9074-4

D'Mello, S., & Graesser, A. (2012). AutoTutor and Affective Autotutor. *ACM Transactions on Interactive Intelligent Systems*, *2*(4), 1–39. https://doi.org/10.1145/2395123.2395128

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze Tutor: A Gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies, 70*(5), 377–398. https://doi.org/10.1016/j.ijhcs.2012.01.004

D'Mello, S., Kappas, K., & Gratch, J. (2018). The affective computing approach to affect measurement. *Emotion Review, 10*(2), 174–183.

D'Mello, S., Dieterle, E., & Duckworth, A. (2017.) Advanced, Analytic, Automated (Aaa) measurement of engagement during learning. *Educational Psychologist, 52*(2), 104–123. https://doi.org/10.1080/00461520.2017.1281747

D'Mello, S., Picard, R. W., & Graesser, A. (2007). "Toward an affect-sensitive autotutor." *IEEE Intelligent Systems, 22* (4), 53–61. doi: 10.1109/MIS.2007.79

DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., . . . Lester, J. C. (2018). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education, 28*, 152-193. doi:10.1007/s40593-017-0152-1

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 104*(4), 1082-1089. doi:10.1037/a0032674

D'Mello, S., & Graesser, A. C. (2015). Feeling, thinking, and computing with affect-aware learning technologies. In R. Calvo, S. D'Mello, J. Gratch, & A. Kappas

(Eds.), *The Oxford handbook of affective computing* (pp. 419-434). New York: Oxford University Press.

D'Mello, S., Blanchard, N., Baker, R., Ocumpaugh, J., & Brawner, K. (2014). I feel your pain: A selective review of affect-sensitive instructional strategies. In R. A. Sottilare, A. Graesser, X. Hu, & B. Goldberg (Eds.), *Recommendations for intelligent tutoring systems: Volume 2-Instructional management.* Orlando, Florida: U.S. Army Research Laboratory.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153-170. doi:10.1016/j.learninstruc.2012.05.003

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., . . . Graesser, A. (2010). A time for emoting: Wehn affect-sensitivity is and isn't effective at promoting deep learning. *10th International Conference on Intelligent Tutoring Systems* (pp. 245–254). Pittsburgh, Pennsylvania: Springer.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies, 70*, 377-398. doi:10.1016/j.ijhcs.2012.01.004

Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. In H. Wagner, & A. Manstead (Eds.), *Handbook of social psychophysiology* (pp. 143-164). Chichester, England: Wiley.

Ekman, P. (1992). An Argument for basic emotions. *Cognition and Emotion, 6*(3/4), 169-200.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotions. *Journal of Personality and Social Psychology, 17*(1), 124-129.

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A techniique for the measurement of facial movement.* Palo Alto, California: Consulting Psychologists Press.

EyeSee. (2019). Facial coding. Retrieved from EyeSee Research Web site: https://eyesee-research.com/facial-coding/

Fredrickson, B. L. (2004). The broaden-and-build theory of positive emotions. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359*(1449), 1367–1377. doi:10.1098/rstb.2004.1512

Gerber, A. J., Posner, J., Gorman, D., Colibazzi, T., Yu, S., Wang, Z., . . . Peterson, B. S. (2008). An affective circumplex model of neural systems subserving valence, arousal, and cognitive overlay during appraisal of emotional faces. *Neuropsychologia, 46*, 2129-2139. doi:10.1016/j.neuropsychologia.2008.02.032

Graesser, A. C., & D'Mello, S. (2012). Emotions during the learning of difficult material. In B. Ross (Ed.), *The psychology of learning and motivations* (pp. 183–225). Burlington, MA: Elsevier.

Graesser, A., & D'Mello, S. (2019). Affect-sensitive intelligent tutoring system. US Patent 10,388,178 B2, issued 2019.

Graesser, A., Chipman, P., Hayes, B., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions in Education, 48,* 614–618.

Graesser, A., S. Lu, G. Jackson, Mitchell, H., Ventura, M., & Olney, A. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments & Computers, 36,* 180–193.

Hamm, J., Kohler, C. G., Gur, R. C., & Verma, R. (2011). Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of Neuroscience Methods, 200*(2), 237-256. doi:10.1016/j.jneumeth.2011.06.023

Hjortsjo, C.-H. (1970). *Man's face and mimic language.* Lund, Sweden: Studentlitteratur. Retrieved from http://diglib.uibk.ac.at/ulbtirol/content/titleinfo/782346

Khan, Farman, Graf, Sabine, Weippl, Edgar, & Tjoa, A Min. (2009). An approach for identifying affective states through behavioral patterns in web-based learning management systems. iiWAS2009 – The 11th International Conference on Information Integration and Web-based Applications and Services, 431-435. doi: 10.1145/1806338.1806418

Keltner, D. (2019). Toward a consensual taxonomy of emotions. *Cognition and Emotion, 33*(1), 14–19. doi:10.1080/02699931.2019.1574397

Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior, 43*, 133–160. doi:10.1007/s10919-019-00293-3

Kercher, K. (1992). Assessing subjective well-being in the old-old: The PANAS as a measure of orthogoal dimensions of positive and negative affect. *Research on Aging, 14*(2), 131–168.

Lee, C.-C., Kiml J., Metallinou, A., Busso, C., Lee, S., & Narayanan, S. S. (2015). Speech in affective computing." In R. A. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford handbook of affective computing*. Oxford; New York: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199942237.013.021

Lee, C. M., & Narayanan, S. S. (2005). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing, 13*(2), 293–303. https://doi.org/10.1109/TSA.2004.838534

Lester, J. C., Rowe, J. P., & Mott, B. W. (2013). Narrative-centered learning environments: A story-centric approach to educational games. In C. Mouza, & N. Lavigne (Eds.), *Emerging technologies for the classroom* (pp. 223-237). New York: Springer.

Mott, B. W., Callaway, C. B., Zettlemoyer, L. S., Lee, S. Y., & Lester, J. C. (1999). Towards narrative-centered learning environments. *Proceedings of the 1999 AAAI Fall Symposium on Narrative Intelligence* (pp. 78-82). Menlo Park, California: Association for the Advancement of Artificial Intelligence.

Noldus. (n.d.). "FaceReader." Retrieved from https://www.noldus.com/facereader

Nye, B., Graesser, A., & Xu, X. (2014). AutoTutor and Family: A review of 17 years of natural language tutoring. *Int J Artif Intell Educ, 24,* 427–469. doi: https://doi.org/10.1007/s40593-014-0029-5

Ocumpaugh, J., Baker, R. S., & Rodrigo, M. T. (2015). *Baker Rodrigo Ocumpaugh Monitoring Protocol* (*BROMP*) *2.0 Technical and Training Manual.* New York, NY: Teachers College, Columbia University.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions.* Cambridge, England: Cambridge University Press.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Education Psychology Review, 18,* 315–341. doi:10.1007/s10648-006-9029-9

Pekrun, R. (2014). *Emotions and learning* (*IBE/2014/ST/EP24*). Belley, France: International Academy of Education/International Bureau of Education. Retrieved from http://www.ibe.unesco.org/

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 259-282). New York, New York: Springer. doi:10.1007/978-1-4614-2018-7_12

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*, 36–48. doi:10.1016/j.cedpsych.2010.10.002

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development, 88*(5), 1653–1670. doi:10.1111/cdev.12704

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The Epistemically-Related Emotion Scales. *Cognition and Emotion, 31*(6), 1268–1276. doi:10.1080/02699931.2016.1204989

Picard, R. W. (1997). *Affective Computing*. Cambridge, MA: MIT Press.

Poria, Soujanya, Erik Cambria, Rajiv Bajpai, and Amir Hussain. (2017). "A Review of Affective Computing: From Unimodal Analysis to Multimodal Fusion." *Information Fusion* 37: 98–125. https://doi.org/10.1016/j.inffus.2017.02.003

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17*(3), 715–734.

Rodrigues, M., Fdez-Riverola, F. & Novais, P. (2011). Moodle and affective computing - knowing who's on the other side. ECEL-2011 - 10th European Conference on e-Learning, (University of Brighton, Brighton, UK 10-11 November 2011), ISBN: 978-1-908272-22-5, 678–685.

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments.

International Journal of Artificial Intelligence in Education, 21(1–2), 115–133. doi:10.3233/JAI-2011-019

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76*(5), 805–819.

Russell, J. A., & Fehr, B. (1994). Fuzzy concepts in a fuzzy hierarchy: Varieties of fear. *Journal of Personality and Social Psychology, 67*(2), 186–205.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1): 145–172. https://doi.org/10.1037/0033-295x.110.1.145

Sabourin, J. L., & Lester, J. C. (2014). Affect and engagement in game-based learning environments. *IEEE Transactions on Affective Computing, 5*(1), 45–56. doi:10.1109/T-AFFC.2013.27

Shaikh, M. A., Prendinger, H., & Ishizuka, M. (2009). A linguistic interpretation of the OCC emotion model for affect sensing from text. In J. Tao, & T. Tan (Eds.), *Affective information processing* (pp. 45–73). London: Springer. doi:10.1007/978-1-84800-306-4

Sottilare, R. A., Baker, R. S., Graesser, A. C., & Lester, J. C. (2018). Special issue on the Generalized Intelligent Framework for Tutoring (GIFT): Creating a stable and flexible platform for innovations in AIED research. *International Journal of Artificial Intelligence in Education, 28*, 139–151. doi:10.1007/s40593-017-0149-9

Steunebrink, B. R., Dastani, M., & Meyer, J.-J. C. (2009). The OCC model revisited. In D. Reichardt (Ed.), *Proceedings of the 4th Workshop on Emotion and Computing.* Paderborn, Germany: Association for the Advancement of Artificial Intelligence.

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology, 38*(2), 227–242. doi:10.1177/0022022106297301

Udochukwu, O., & He, Y. (2015). A rule-based approach to implicit emotion detection in text. *International Conference on Applications of Natural Language to Information Systems* (pp. 197-203). Passau, Germany: Springer.

VanLehn, K. (2016). Regulative loops, step loops and task loops. *International Journal of Artificial Intelligence in Education, 26*, 107–112. doi:10.1007/s40593-015-0056-x

Walkington, C. A. (2013). Using adaptive learning technologies to personalize instruction to atudent interests: The impact of eelevant xontexts on performance and learning outcomes. *Journal of Educational Psychology, 105*(4), 932–945. doi:10.1037/a0031882

Watson, D., & Clark, L. A. (1992). On traits and temperament: General and specific factors of emotional experience and their relation to the five-factor model. *Journal of Personality, 60*(2), 441–475.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Pscyhological Bulletin, 98*(2), 219–235.

Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology, 97*(3), 346–353.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and vlidation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070.

Whitehill, J., Bartlett, M., & Movellan, J. R. (2008). Automatic facial expression recognition for intelligent tutoring systems. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Anchorage, AK, 1–6. doi: 10.1109/CVPRW.2008.4563182

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. W..(2009). Affect-aware tutors: Recognising and responding to student affect. *Int. J. Learning Technology 4*(3/4), 129–163.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43*(1), 111–122.

Abbreviations

ASW	anti-submarine warfare
AU	action unit
BET	Basic Emotion Theory
BROMP	Baker-Rodrigo Observation Monitoring Protocol
CTA	Cognitive Tutor Algebra
EES	Epistemically-Related Emotion Scales
FACS	Facial Action Coding System
fMRI	functional magnetic resonance imagery
FTF	face-to-face
I/ITSEC	Interservice/Industry Training, Simulation and Education Conference
LMS	learning management system
MDS	mesolimbic dopamine system
NA	nucleus accumbens
NETC	Navy Education and Training Command
OCC	Ortony, Clore, and Collins
PANAS	Positive and Negative Affect Schedule
RF	reticular formation
SAIL	Signal Analysis and Interpretation Laboratory
SLC	Submarine Learning Center
SME	subject-matter expert
USC	University of Southern California
VSH	Virtual Schoolhouse

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE September 2020	2. REPORT TYPE FINAL	3. DATES COVERED *(From–To)*

4. TITLE AND SUBTITLE

Affect-Sensitive Instructional Strategies for Synchronous Distance Learning

5a. CONTRACT NUMBER
HQ0034-14-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Morrison, John E.
Fedele, Emily A.

5d. PROJECT NUMBER
AI-2-4637

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Defense Analyses
Systems and Analyses Center
4850 Mark Center Drive
Alexandria, VA 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER

IDA Document D-14245

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Training and Education Command
N516 (Planning, Analysis, and Research & Development)
Pensacola, FL 32508

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited (29 June 2021).

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Synchronous distance learning does not provide nonverbal student feedback to the instructor indicating the student's emotional state. Nonverbal emotive cues provide the instructor valuable information to adjust and adapt the pace and content of instruction to the students' affective and cognitive states. The emerging technology of automated affect recognition provides an innovative approach to providing nonverbal instructional feedback. However, to take full advantage of this technology, an instructional system must not only detect the affective states of students but also respond appropriately to those states. Thus, the development of an affect-sensitive learning system must address three separate problems: (1) dynamically collect cognitive and affective information from the learner to assess affective state, (2) understand and model the implications that those affective states have on instruction, and (3) choose an appropriate instructional intervention for individual students and contexts. Once the intervention is deployed, student affect is reassessed and the cycle restarts. The objective of this report is to examine and assess the maturity of the science and technology behind the three problems (assess state, understand state, and determine intervention) and suggest how best affect-sensitive learning technologies can be deployed to enhance synchronous distance learning.

15. SUBJECT TERMS

adaptive learning strategies; affect-sensitive instruction; emotion recognition; synchronous distance learning

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Petro, James
Uncl.	Uncl.	Uncl.	SAR	95	19b. TELEPHONE NUMBER *(include area code)* (571) 372-6434

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18