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

Ryan S.J.d. Baker[a],[*],[1], Sidney K. D'Mello[b], Ma.Mercedes T. Rodrigo[c], Arthur C. Graesser[b]

[a]Worcester Polytechnic Institute, Department of Social Science and Policy Studies, 100 Institute Road, Worcester, MA 01609, USA
[b]Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152, USA
[c]Department of Information Systems and Computer Science, Ateneo de Manila University, Katipunan Avenue, Loyola Heights, Quezon City 1108, Philippines

Abstract

We study the incidence (rate of occurrence), persistence (rate of reoccurrence immediately after occurrence), and impact (effect on behavior) of students' cognitive–affective states during their use of three different computer-based learning environments. Students' cognitive–affective states are studied using different populations (Philippines, USA), different methods (quantitative field observation, self-report), and different types of learning environments (dialogue tutor, problem-solving game, and problem-solving-based Intelligent Tutoring System). By varying the studies along these multiple factors, we can have greater confidence that findings which generalize across studies are robust. The incidence, persistence, and impact of boredom, frustration, confusion, engaged concentration, delight, and surprise were compared. We found that boredom was very persistent across learning environments and was associated with poorer learning and problem behaviors, such as gaming the system. Despite prior hypothesis to the contrary, frustration was less persistent, less associated with poorer learning, and did not appear to be an antecedent to gaming the system. Confusion and engaged concentration were the most common states within all three learning environments. Experiences of delight and surprise were rare. These findings suggest that significant effort should be put into detecting and responding to boredom and confusion, with a particular emphasis on developing pedagogical interventions to disrupt the "vicious cycles" which occur when a student becomes bored and remains bored for long periods of time.

1. Introduction

The field of interface development was radically transformed when design decisions began to be informed by users' physical limitations and cognitive constraints in addition to the technical concerns that initially dominated issues in system development (Carroll, 1997). Stemming from the human factors movement in the early fifties and the cognitive revolution of the sixties and seventies, the impetus of human–computer interaction began to gradually shift away from the computer and more towards the human.

A second key shift occurred with the change in emphasis from focusing primarily on the cognitive constraints of the user (e.g. working memory load, information overload, split attention, etc.), to focusing on users' affective experiences (emotions, moods, feelings) and how affect influences other aspects of the broader human–computer interaction. Since Picard's influential 1997 book *Affective Computing*, there has been a burst of research that focuses on creating technologies that can monitor and appropriately respond to the affective states of the user (Picard, 1997). Such

*Corresponding author. Tel.: + 1 508 831 5355; fax: + 1 508 831 5896.
E-mail addresses: ryan@educationaldatamining.org,
rsbaker@andrew.cmu.edu, rsbaker@wpi.edu (R.S. Baker).
sdmello@memphis.edu (S.K. D'Mello), mrodrigo@ateneo.edu
(M.T. Rodrigo), a-graesser@memphis.edu (A.C. Graesser).
[1]Tel.: + 1 412 268 9690; fax: + 1 412 268 1266.

systems attempt to bridge the communicative gap between emotionally expressive humans and generally socially deficient computers. Even 5 years ago, it could be argued that affect was being consistently de-emphasized within HCI research (McNeese, 2003; Picard et al., 2004). However, in the last few years, there has been considerable research aspiring to incorporate the affective states of a user into the decision cycle of the interface in an attempt to develop more effective, user-friendly applications (Hudlicka and McNeese, 2002; Klein et al., 2002; Mandryk and Atkins, 2007; Marinier and Laird, 2006, 2007; Norman, 2004; Prendinger and Ishizuka, 2005; Whang et al., 2003).

The inclusion of emotions into the decision cycle of computer interfaces is motivated by the hypothesis that there is a complex interplay between cognition and emotion (Mandler, 1984). Simply put, emotions are systematically affected by the knowledge and goals of the user, and vice versa (Mandler, 1984, 1999; Stein and Levine, 1991). Cognitive activities such as causal reasoning, deliberation, goal appraisal, and planning processes operate continually throughout the experience of emotion. Given the complex relationships between affect and cognition, some key user states that are highly relevant to students' experiences, such as confusion and engaged concentration, can be considered a blend of affect and cognition. Within this paper, we refer to these states as *cognitive–affective states.* Due to the interrelationships between affect and cognition, an interface that is sensitive to a user's affective states as well as their cognitive states is likely to be more usable, useful, socially appropriate, enjoyable—all factors that may lead to wider use and acceptance.

1.1. Affect in intelligent tutoring systems

Intelligent Tutoring Systems (ITSs) are hypothesized to be a particularly good candidate for improvement by addressing the connections between affect, cognition, motivation, and learning (e.g. Arnold, 1999; Bower, 1992; Sylwester, 1994). ITSs are a type of educational software that offer guided learning support to students engaged in problem-solving. Existing intelligent tutors tailor their support of students' needs in a variety of ways, including identifying and correcting student errors (Anderson et al., 1995; VanLehn, 1990) and promoting mastery learning through assessments of the probability that the student knows each skill relevant to the system (Corbett and Anderson, 1995). ITSs have emerged as valuable systems to promote active learning, with learning gains associated with sophisticated ITSs at around a 1.0 SD improvement (about a letter grade) when compared to controls (Dodds and Fletcher, 2004; Koedinger and Corbett, 2006; Van-Lehn et al., 2005). These learning gains are higher than those achieved by inexperienced human tutors (~ 0.4 SD, see Cohen et al., 1982) but are not quite as good as learning gains achieved by expert human tutors (~ 2.0 SD, see Bloom, 1984).

Over the last few years there has been work towards incorporating assessments of the learner's affect into intelligent tutoring systems. Kort et al. (2001) proposed a comprehensive four-quadrant model that explicitly links learning and affective states. This model was used in the MIT group's work on their Affective Learning Companion, a fully automated computer program that recognizes a learner's affect by monitoring facial features, posture patterns, and onscreen keyboard/mouse behaviors (Burleson, 2006). de Vicente and Pain (2002) developed a system that could track several motivational and emotional states during a learning session with an ITS. The system was trained on judgments by expert coders. Conati (2002) developed a probabilistic system that can reliably track multiple affective states (including joy and distress) of the learner during interactions with an educational game, and use these assessments to drive the behavior of an intelligent pedagogical agent (Conati and Zhou, 2004). Litman and Silliman's (2004) work with the ITSPOKE (2004) conceptual physics ITS has used a combination of discourse markers and acoustic–prosodic cues to detect and respond to a learner's affective states.

Though there have already been multiple attempts to detect affect in learning environments, there is still relatively little understanding of the impact of affect on students' behavior and learning during tutorial sessions (Graesser et al., 2009). Our knowledge on the natural dynamics of affect during learning with software is similarly impoverished. For instance, some research has focused on reducing users' frustration (Hone, 2006; Klein et al., 2002; McQuiggan et al., 2007), but it is not even clear that frustration is always a negative experience in all types of human–computer interaction. In particular, Gee (2004) has proposed that frustration to a certain degree may actually enhance the enjoyability of computer games. In general, it is not yet clear how harmful or persistent the affective states currently thought to be negative actually are. For example, confusion, a cognitive–affective state that is often perceived as being negative, has been shown to be positively correlated with learning (Craig et al., 2004; Graesser et al., 2007).

This paper investigates the cognitive–affective states that occur during interactions with learning environments, the persistence of these states over time, and the extent to which these states correlate with student behaviors associated with poorer learning. Since differences among learning contexts play an important role in students' affective and cognitive experiences, these issues are studied using different populations (Philippines, USA), different methods (quantitative field observation, self-report), and different types of learning environment (dialogue tutor, problem-solving game, and problem-solving-based ITS). By varying the studies along multiple factors, we can have greater confidence that findings which replicate across studies are general (though our ability to draw conclusions about differences among systems is correspondingly reduced).

1.2. Gender, age, and cultural differences in experiences and expressions of affect

One challenge to the field of affective computing has been determining the generalizability of the results. There is considerable diversity in potential users and characteristics of interactive systems, and the display of emotion differs in striking ways between cultures. It is possible for people to recognize some emotions of individuals in other cultures when given appropriate cues (Elfenbein and Ambady, 2002a, 2002b; Ekman and Friesen, 1971, 1978; Hess et al., 1988), but there is also evidence that it is easier to recognize the emotions of people from the same culture (Russell, 1994). Differences among cultures and individuals present a challenge for affect recognition systems that are initially designed using data from a small number of users in a single culture.

Differences in the age, gender, and personality of users are also a matter of some concern. For example, young children are generally more expressive than teens and adults. They are less likely to suppress and disguise their emotions due to societal pressures (Ekman and Friesen, 1969). Therefore, systems that recognize affect may have problems generalizing across different age groups. Furthermore, systems may need to invoke different responses in order to appropriately manage the emotions of the full diversity of users. Boys and girls often show different patterns of engagement to the same classroom activity; similarly, boys and girls who are disengaged within learning software may require different responses by the system (Peterson and Fennema, 1985).

Individual differences are also a critical factor when designing affect-sensitive interfaces. For example, learners have different preferences that must be taken into account. Within the context of learning environments, some results suggest that "adventuresome" learners prefer to be challenged with difficult tasks, worrying little about negative emotion if they fail, while other "cautious" learners prefer easier tasks and try to avoid failure and its resulting negative emotions (Clifford, 1988, 1991; Meyer and Turner, 2006). The literature on performance goals and learning goals (Dweck, 2002) also suggests that learners' goals may alter their affective responses to successes and failures in learning tasks. Achievement-oriented students focus most on being perceived to have performed well, whereas mastery-oriented students focus more on understanding the subject matter (Dweck, 2000).

Different types of learning environments and content domains may also result in very different affective profiles. For example, an immersive game-like learning environment may evoke a different profile of affective experiences than a classical computer-assisted instruction (CAI) system (Lepper and Cordova, 1992). It is reasonable to expect higher engagement in a game environment than a traditional CAI system (Gee, 2004; Prensky, 2007). Students' engagement undoubtedly varies across subject matter, with some students more interested in history and others more interested in algebra. The context of use presumably also alters students' motivation and emotions. For example, a mathematics tutoring system which prepares students for high stakes testing may evoke different affective experiences, depending on whether it is being used 6 months or 2 days before the test.

1.3. Measurement issues in affective computing

There are several ways that affect within computerized systems could be measured for study. One possibility is automatic detection of cognitive–affective states by computers (D'Mello et al., 2007; Kapoor et al., 2007; Prendinger and Ishizuka, 2005). Pantic and Rothkrantz (2003) have surveyed the considerable recent progress in real time affect detection. Paiva et al. (2007) discuss progress on affect detection through body movement and gestures (Castellano et al., 2007), acoustic–prosodic cues (Grimm et al., 2007), lexical features (Wagner et al., 2007), and physiological features (Komatsu et al., 2007). Despite these inroads, accuracy rates in real-world environments are not yet sufficient to use as dependent measures in some research applications, particularly within in-vivo classroom learning settings. There is also the practical challenge of deploying some of the affective sensors in sufficient quantities, as in the cases of physiological measurement equipment, pressure sensitive chairs, and eye-trackers.

Thus, the majority of affect research still relies on humans to measure cognitive–affective states, an approach we have adopted within the studies presented in this paper. However, the selection of human judgment for measuring these states still leaves open a number of methodological possibilities. There are open questions about the use of self-reports versus external observers, about whether observations should occur concurrently with the interaction or later, and about where the judgments should be conducted (i.e. in the same room as the participant or elsewhere, perhaps from a room equipped with a one-way mirror or remote laboratory). As with every design decision, these alternatives have tradeoffs that need to be carefully evaluated and tested. However, any effect that is obtained only in a very specific research setting may be contingent on the method of measurement rather than being truly generalizable. In order to understand how generalizable our findings are, we have chosen to apply different methods to address the same research questions (described below), both within this paper and in our prior research. Within the studies presented here, our goal is to determine which aspects of the affective experience generalize above and beyond differences in the user, the system, and the methodology of study. An interactive system designed in accordance with consistent findings permits a system designer to have higher confidence about the generalizability of the system. What are the cognitive–affective factors that are general across populations, contexts, and

methods? Which of these factors are important to the overall interaction experience? These questions are explored in this paper.

1.4. Research questions

This paper focuses on three research questions related to students' cognitive–affective states during interactions with learning software:

1. What cognitive–affective states do students experience more often during learning sessions with computerized learning environments?
2. Are there differences in how these states persist over time? That is, are some states persistent and others ephemeral?
3. What is the impact of cognitive–affective states on students' choices of how to interact with an interactive learning environment?

1.4.1. Cognitive–affective states during learning

Much of affect research in psychology has focused on Ekman and Friesen's (1978) six basic emotions, which are hypothesized to be ubiquitous in everyday experience: fear, anger, happiness, sadness, disgust, and surprise. However, researchers have increasingly called into question the relevance of these basic emotions to the learning process (D'Mello et al., 2007; Graesser et al., 2007; Lehman et al., 2008a, 2008b; Kort et al., 2001). For example, it is fairly unlikely that a student consistently experiences sadness or fear while interacting with a tutoring system. Although it is conceivable that these emotions may play a role when the learning task occurs over a larger temporal bandwidth (e.g. completing a dissertation), it is highly unlikely that they routinely occur in a typical learning session of 30 min to 2 h.

As an alternative to the basic emotions, we focus on a set of cognitive–affective states that several researchers have hypothesized to influence cognition and deep learning. These include boredom (Csikszentmihalyi, 1990; Miserandino, 1996), confusion (Craig et al., 2004; Graesser et al., 2008; Kort et al., 2001), delight (Fredrickson and Branigan, 2005; Silvia and Abele, 2002), engaged concentration (cf. Csikszentmihalyi, 1990), frustration (Kort et al., 2001; Patrick et al., 1993), and surprise (Schutzwohl and Borgstedt, 2005). The definition of most of these terms is well-known and needs no further explication here. "Engaged concentration" needs some clarification, however. Engaged concentration is a cognitive–affective state that sometimes has a short time span, but at other times forms part of Csikszentmihalyi's conception of flow. Engaged concentration is a state of engagement with a task such that concentration is intense, attention is focused, and involvement is complete. However, it need not involve some of the task-related aspects which Csikszentmihalyi associates with flow, such as clear goals, balanced

challenge, or direct and immediate feedback. It also may not involve some of the aspects of Csikszentmihalyi's conceptualization which refer to extreme intensity, such as time distortion or loss of self-consciousness.

It should be noted that some researchers may view some of these states as pure cognitive states, whereas most researchers would classify them as either emotions, affect states, or blends of cognition and affect (see Barrett, 2006; Stein et al., 2008; Meyer and Turner, 2006). We adopt the position of identifying them as cognitive–affective states because they have significant cognitive and affective components in the context of learning.

Our set of learning-centered cognitive–affective states can be situated within a broader perspective of emotion, in particular Russell's (2003) Core Affect framework. This perspective holds that an affective state is composed of two integrated components: *valence* (pleasure to displeasure) and *arousal* (activation to deactivation). These components can be depicted graphically with valence represented on the *X*-axis and arousal on the *Y*-axis (see Fig. 1). Moving from left to right along the *X*-axis (valence) would correspond to increasing feelings of pleasure. Moving upward along the *Y*-axis (arousal) would correspond to increasing feelings of activation and energy.

Fig. 1 depicts the mapping of the learning-centered cognitive–affective states on Russell's core-affect framework (2003). *Boredom* has a negative valence and low level of arousal. *Confusion* has a negative valence and a moderate level of arousal. *Frustration* has a high negative valence and a high level of arousal. *Delight* has a positive valence and a high level of arousal, whereas *surprise* has high arousal but can have either positive or negative valence. *Engaged concentration* has a positive valence. In terms of arousal, there is not yet consensus about engaged concentration—hence, we have tentatively listed it as neutral in the circumplex. When engaged concentration is stimulated by novel input, we can infer a slight increase in arousal, whereas there is a decrease in arousal to the extent that the person experiences uninterrupted, organized cognition or action (Mandler, 1984).

1.4.2. Are there differences in how these states persist over time? Are some states persistent and others ephemeral?

This question pertains to the manner in which a student's cognitive–affective state persists over time in learning environments that do not explicitly attempt to monitor and alter affect. Understanding the persistence of cognitive–affective states in these learning environments will be useful to researchers in many fashions. In particular, a better understanding of the persistence of cognitive–affective states will help us set goals for the design of affect-sensitive learning environments, i.e. systems that incorporate assessments of learners' cognitive–affective states into their decision cycles. In deciding which cognitive–affective states a learning system should respond to, it is important to know if a student's state is likely to shift naturally, and how it may shift. Some cognitive–affective
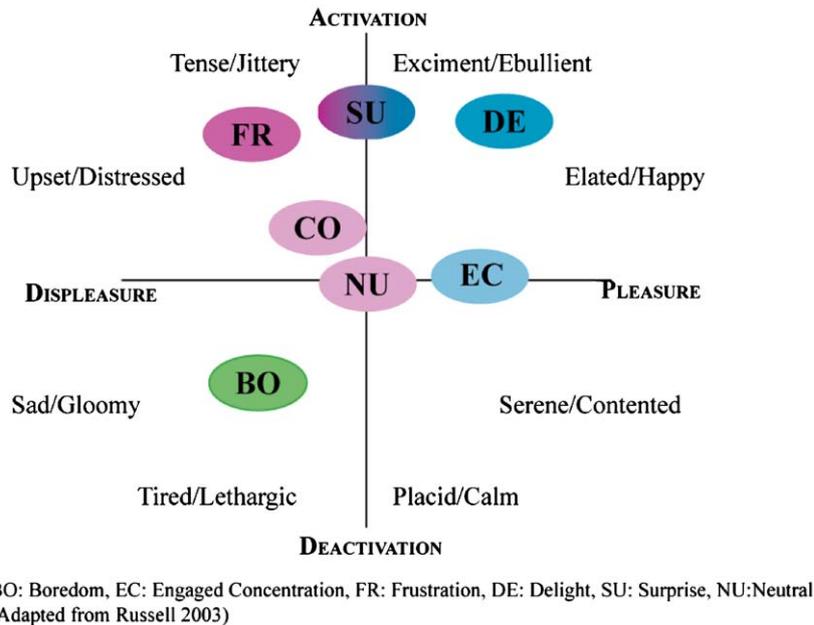
BO: Boredom, EC: Engaged Concentration, FR: Frustration, DE: Delight, SU: Surprise, NU:Neutral
(Adapted from Russell 2003)

Fig. 1. Learning-centered cognitive–affective states mapped onto Russell's Core Affect Framework (2003). Recreated and modified from Russell (2003).

states, once entered, may be quite persistent, and therefore may merit a response by the learning environment, particularly when those states have negative valence. Other states may be more transitory and therefore may not warrant a response. Of course, even transitory states may be problematic if they commonly transition to other negative states, and especially so if two or more negative cognitive–affective states form a ''vicious cycle'' (cf. D'Mello et al., 2007).

Past work has not produced conclusive evidence on which cognitive–affective states persist, relative to other cognitive–affective states. D'Mello et al. (2007) propose that vicious cycles exist, but do not concretely provide evidence as to which cycles pre-dominate, or which states persist over time. Csikszentmihalyi (1990) indicates that flow (which, as previously mentioned, is a more complex construct that incorporates, but is not limited to, engaged concentration) can be quite persistent, but it is not clear whether either flow or engaged concentration is more persistent than other states such as boredom. It seems reasonable to hypothesize that surprise will be fairly transitory (how long can one stay surprised?), and that engaged concentration, boredom, confusion, and frustration will be more persistent. It is not clear, however, from prior empirical research, which states will be more and less persistent within each of these broader categories.

There is similarly a lack of evidence on which affective transitions occur naturally. Kort et al. (2001) propose a model of affect over time during learning, that hypothesizes that it will be common to see transitions from confusion to frustration, and confusion to boredom (but that the reverse order will be less likely). However, Kort et al. did not provide empirical evidence for this model. Perkins and Hill (1985), proposed that frustration leads to boredom, but this hypothesis was based on data showing

that the two states were associated rather than evidence that they are causally or temporally related (boredom preceded by frustration).

Understanding and modeling affective persistence can provide a baseline for analyses that explore the impact of systems designed to influence affect. If a persistent cognitive–affective state becomes less persistent, for example, one might infer that the environment has influenced students' affect or cognition. Yet another virtue of a model of affective persistence is that it enables better estimation of the base rate of cognitive–affective states. For instance, boredom may be more common after boredom than its overall frequency might suggest. Understanding affective persistence may therefore make it possible to develop more successful and accurate detectors of affect by integrating better baseline information with sensor and/or keystroke data.

1.4.3. What is the impact of cognitive–affective states on students' choices of how to interact with an interactive learning environment?

The third research question investigated in this paper is the impact of affect on students' choices of how to interact with an interactive learning environment. That is, how often does a given cognitive–affective state precede usage choices which are known to be associated with reduced or enhanced learning? In particular, consider the phenomenon of *gaming the system,* which consists of attempting to succeed in an interactive learning environment by exploiting properties of the system rather than by learning the material. Gaming the system is associated with poorer learning (Baker et al., 2004; Baker, 2005; Walonoski and Heffernan, 2006). Walonoski and Heffernan reported that the same students who game the system also experience frustration. Given this evidence, it is reasonable to hypothesize that frustration leads a student to game the

system. Baker et al. (2009) reported that user interface features designed to increase interest also decrease gaming. Hence, it may also be reasonable to hypothesize that boredom may lead a student to game the system.

1.5. Research plan

This paper investigates, across three studies, the affective profiles and persistence of cognitive–affective states during learners' use of educational software. In order to study the generalizability of our findings, we study learning situations and learners that vary simultaneously on several dimensions. Hence, findings that replicate across studies are more likely to be generalizable. By contrast, a more conservative strategy would vary only one contextual variable at a time; this approach would take more time in effort to the extent there are a large number of potential contextual variables to consider, with many potentially being inconsequential. To this end, we conducted studies with learners from different cultures, at different ages, and both genders. We examined affect within learning environments in different domains (Computer Literacy, Algebra, Concrete Logic Puzzles) and with different underlying pedagogical strategies (dialogue tutor, traditional work-book-style tutor, and a simulation game).

As with every design decision, allowing method to vary has both positive and negative consequences. On the positive side, varying methodology allows us to ascertain whether any patterns observed in the data generalize across studies. This can be accomplished by analyzing each study independently, identifying the reliable patterns, and assessing whether these patterns replicate across studies. However, the fact that more than one factor was varied across studies considerably reduces our ability to make causal inferences on the impact of any given difference among studies. In our viewpoint, this limitation is acceptable because the primary goal of this paper is to investigate generalizability of affective patterns, and not to make claims about the causes of differences among studies. Any differences observed among studies can instead become testable hypotheses for further research.

The paper is organized as follows: we begin by describing the three learning environments studied, and the protocol for the three studies that attempted to measure affect during learning sessions. The results section investigates the occurrence of cognitive–affective states that occurred within each system, the temporal persistence of the states, and how the states influence students' behavior with the learning environments (e.g. whether the student games the system). We conclude by indicating how our findings can be used to scaffold the development of affect-sensitive learning environments.

2. Descriptions of learning environments

Cognitive–affective state data was gathered from participants who used three different interactive learning environments: AutoTutor, the Incredible Machine, and Aplusix. These learning environments differ not only in subject matter domain and interface issues, but also in the underlying pedagogical principles they embody.

2.1. AutoTutor

AutoTutor is a fully automated computer tutor that simulates human tutors and holds conversations with students in natural language (Graesser et al., 2001, 2005). AutoTutor attempts to comprehend students' natural language contributions and then responds to the students' typed input with adaptive dialogue moves similar to human tutors. AutoTutor helps students learn by presenting challenging problems (or questions) and engaging in a mixed-initiative dialogue while the learner constructs an answer. Fig. 2 presents a screen shot of the AutoTutor interface.

AutoTutor has different classes of dialogue moves that manage the interaction systematically. AutoTutor provides feedback on what the student types in (positive, neutral, or negative feedback), pumps the student for more information ("What else?"), prompts the student to fill in missing words, gives hints, fills in missing information with assertions, identifies and corrects misconceptions and erroneous ideas, answers the student's questions, and summarizes topics. A full answer to a question is eventually constructed during this dialogue, which normally takes between 30 and 100 turns between the student and tutor for one particular problem or main question.

The efficacy of AutoTutor in imparting conceptual knowledge has been validated in six experiments in the domain of physics in which learning gains were evaluated on approximately 500 college students (VanLehn et al., 2007). The subject matter in these experiments was introductory qualitative Newtonian physics. Tests of AutoTutor have produced gains of 0.4 to 1.5 sigma (a mean of 0.8, about a letter grade), depending on the learning measure, the comparison condition, the subject matter, and version of AutoTutor. Similar results were reported in a version of AutoTutor on the subject matter of computer literacy (Graesser et al., 2004).

2.2. The Incredible Machine (TIM)

The Incredible Machine: Even More Contraptions (Sierra Online Inc., 2001) is a simulation environment where students complete a series of logical "Rube Goldberg" puzzles. In each puzzle, the student is given (a) objects with limited interactivity, including mechanical tools like gears, pulleys, and scissors, (b) more active objects such as electrical generators and vacuums, and (c) animals. The student must combine these objects in a creative fashion to accomplish each puzzle's goal. Goals range from relatively straightforward goals, such as lighting a candle, to more complex goals, such as making a mouse run. If a student is stuck, he or she can ask for a hint; hint messages display
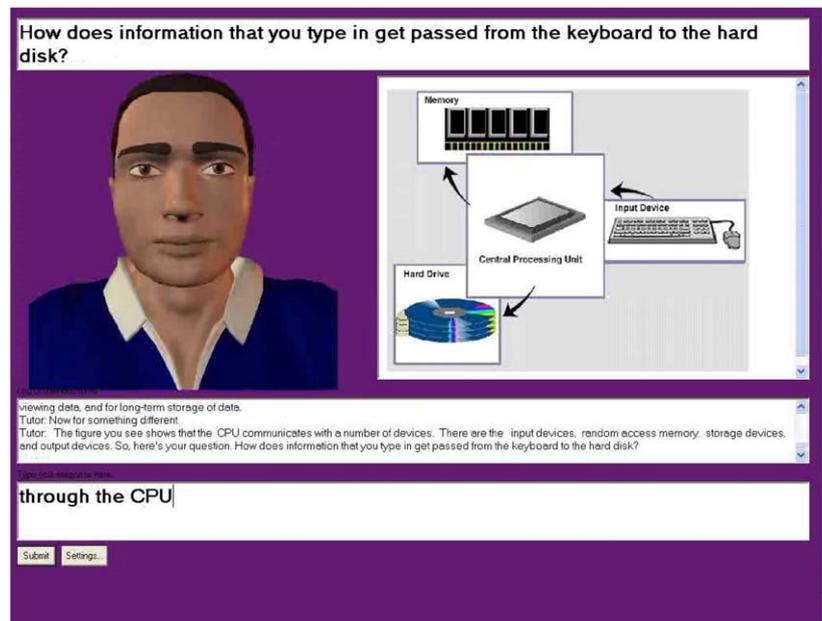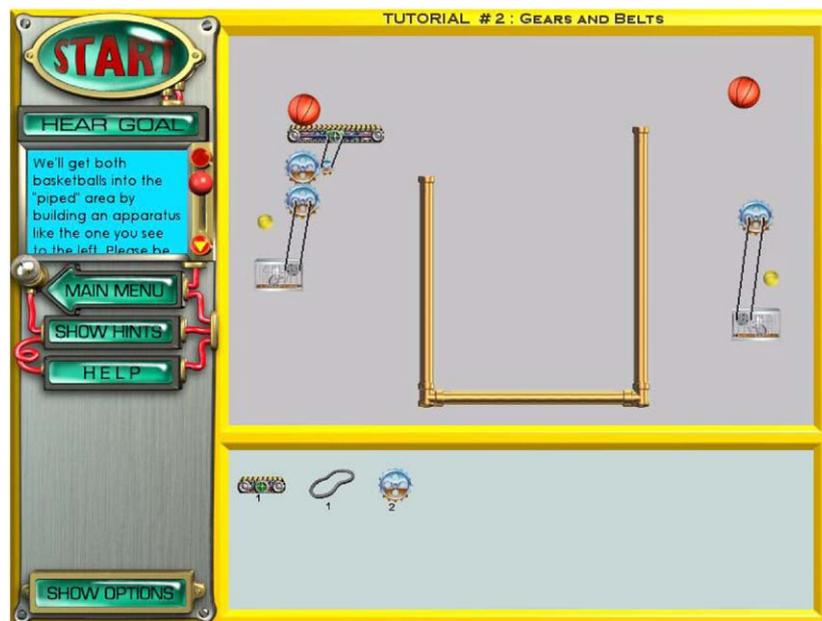
Fig. 2. The AutoTutor interface.

Fig. 3. A screen shot from The Incredible Machine: even more contraptions. Using the objects at the bottom of the screen, the player must build a device that will propel the basketball on the right side of the screen into the U-shaped area at the center of the screen.

where items should be located in a correct solution to the current problem (without displaying which items should be placed in each location). A screen shot from The Incredible Machine is shown in Fig. 3.

2.3. Aplusix

Aplusix II: Algebra Learning Assistant (Nicaud et al., 2007) is an intelligent tutoring system for mathematics. Topics are grouped into six categories (numerical calcula-tion, expansion and simplification, factorization, solving equations, solving inequations, and solving systems), with four to nine levels of difficulty each. Aplusix presents the student with an arithmetic or algebraic problem from a problem set chosen by the student. Students then solve the problem one step at a time. At each step, Aplusix displays equivalence feedback: two black parallel bars mean that the current step is equivalent to the previous step, two red parallel bars with an X mean that the current step is not equivalent to the previous step (see Fig. 4). Aplusix does

not indicate which part of the current step requires further editing. A student can end the exercise when they believe they have completed the problem. Aplusix then tells the student whether errors still exist along the solution path or whether the solution is not in its simplest form yet. The student also has the option of looking at the solution at any point.

Since 2002, thousands of students in grades 8, 9, and 10 from several countries, including France, Brazil, India, Vietnam, and the Philippines, have used Aplusix. Studies have shown statistically significant improvements in learning on mathematics problem-solving tests (Nicaud et al., 2004).

3. Methods

All three environments were studied using human judgments about the same set of six cognitive–affective states (and a seventh state, the neutral state). However, the
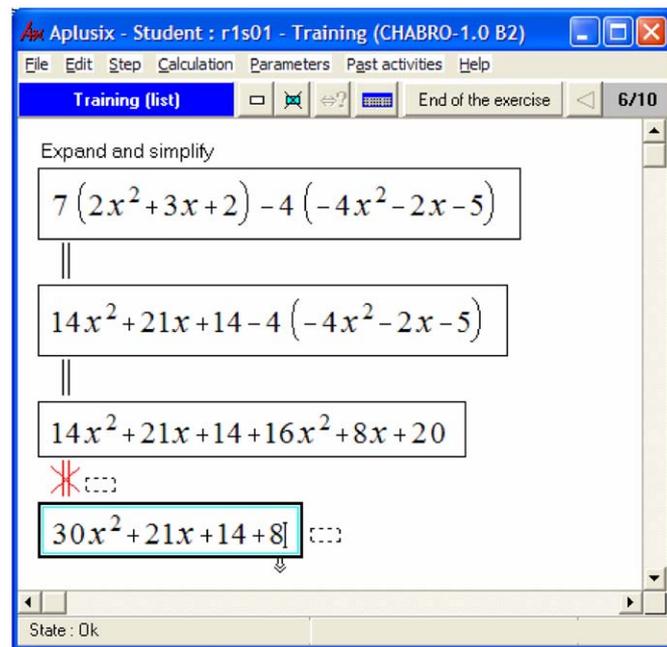

Fig. 4. A screen shot from Aplusix: algebra learning assistant.

three environments were studied in different contexts, with different populations, and using different methods. By varying the studies along multiple factors, we can have greater confidence that findings which generalize across studies are robust. Of course, this approach has the disadvantage of making it difficult to interpret when two environments have different results. For this reason, differences in the pattern of cognitive–affective states between learning environments will not be explicitly analyzed within this paper. A summary of the methods used in the three studies appears in Table 1.

3.1. Study 1—AutoTutor

3.1.1. Participants
The participants were 28 undergraduate students from a university in the mid-south of the USA, who participated in this study for extra course credit.

3.1.2. Interaction procedure
A standard pre-test–intervention–post-test design was utilized. After completing the pretest, participants used the AutoTutor system for 32 min on one of three randomly assigned topics in computer literacy (Hardware, Internet, Operating Systems). During the tutoring session, a video of the participants' faces, their posture patterns (see D'Mello et al., 2007), and a video of the content of their computer screen were recorded. Lastly, after completing the tutoring session, the participants completed a 36-item posttest assessment on the topics of computer literacy studied.

3.1.3. Cognitive–affective state judgment procedure
The judging of students' cognitive–affective states proceeded by synchronizing and displaying video streams of both the computer screen and the learner's face, both of which were captured during the AutoTutor session. Posture data was not utilized during judging. Each participant made judgments on what cognitive–affective states they had been experiencing at every 20-s interval (i.e. at the end of each interval the video automatically paused),

Table 1
Summary of different studies.

Dimension	Factor	Study 1	Study 2	Study 3
User characteristics	Participants	28	36	140
	Age	College students	14–19	12–15
	Gender	23F + 5M	17F + 19M	83F + 57M
System characteristics	System	AutoTutor	Incredible Machine	Aplusix
	Domain	Computer literacy	Logic puzzles	Algebra
	Pedagogical strategy	Dialogue-based ITS	Serious game	Computer tutor
Methodological characteristics	Interaction time (min)	32	10	45
	Affect judgment	Offline	Online	Online
	Affect judge	Self-judgments	Trained judges	Trained judges
	Sampling rate (s)	Continuous	60	200

as well as any other states they observed in between these intervals.

A list of the cognitive–affective states and definitions was provided for all participants. The states were frustration, confusion, engaged concentration, delight, surprise, boredom, and neutral. Frustration was defined (for participants) as dissatisfaction or annoyance. Confusion was defined as a noticeable lack of understanding, whereas engaged concentration was a state of interest that results from involvement in an activity. Delight was defined as a high degree of satisfaction. Surprise was defined as wonder or amazement, especially from the unexpected. Boredom was defined as being weary or restless due to lack of interest. Participants were given the option of making a neutral judgment to indicate a lack of distinguishable affect. Neutral was defined as no apparent emotion or feeling.

3.2. Study 2—The Incredible Machine

3.2.1. Participants

The participants for the Incredible Machine study were students in a private high school in Quezon City, the Philippines. Student ages ranged from 14 to 19, with an average age of 16. Thirty-six students participated in this study (17 female, 19 male).

3.2.2. Interaction procedure

Students used The Incredible Machine for 10 min, and each student was observed several times as they used the system. During the laboratory sessions in which the data was gathered, it was not possible for the entire class to use the software at the same time, due to the size of the school computer laboratory. Students therefore used the software in groups of nine (one student per computer) during their class time.

3.2.3. Cognitive–affective state and behavior judgment procedure

The observations were carried out by a team of six observers, working in pairs. Each pair was assigned three students per observation period. The observers were graduate students in Education or Computer Science, and all but one had prior teaching experience.

Each observation lasted 20 s, and was conducted using peripheral vision. That is, the observers stood diagonally behind or in front of the student being observed and avoided looking at the student directly (cf. Baker et al., 2004), in order to make it less clear when an observation was occurring. This method of observing using peripheral vision was previously found to be highly successful for assessing student behavior, achieving good inter-rater reliability (Baker et al., 2006), and forming the basis for the construction of highly accurate automated detectors of student behavior (Baker, 2007; Baker et al., 2008), which captured the relevant constructs sufficiently well to be able

to drive automated interventions which significantly improved student learning (Baker et al., 2006).

If two distinct cognitive–affective states were seen during an observation, only the first state observed was coded; similarly, if two distinct behaviors were seen during an observation, only the first behavior observed was coded. Any behavior or (evidence of a cognitive–affective state) by a student other than the student currently being observed was not coded. Each pair of observers was assigned to three students and alternated among them. Since each observation lasted 20 s, each student was observed once per minute.

The observers based their judgment of a student's state or behavior on the student's work context, actions, utterances, facial expressions, body language, and interactions with teachers or fellow students. These are, broadly, the same types of information used in previous methods for coding affect (e.g. Bartel and Saavedra, 2000), and in line with Planalp et al.'s (1996) descriptive research on how humans generally identify affect, using multiple cues in concert for maximum accuracy rather than attempting to select individual cues. The judgments of behavior were based on a coding scheme developed by Baker et al. (2004). The judgments of cognitive–affective state were based on a coding guide developed by the third author and her students. Prior to sessions where qualitative observations were conducted, the observers discussed the coding categories and how to classify specific student behaviors with reference to these categories. The observers also practiced coding during a pilot observation period prior to this study. After the pilot observations, we checked for cases where inter-rater agreement was unacceptably low, with Cohen's (1960)κ below 0.50. For each of those cases, we examined the confusion matrix to determine which cognitive–affective states were involved when raters disagreed. Raters were then debriefed about the different characteristics that they used to distinguish one cognitive–affective state from the other. The discussion continued until the raters reached a consensus. By the second observation session, agreement between raters improved dramatically. It is worth noting, from the behaviors included in the guide, that students were demonstrative about their affect during system usage. In learning settings where students are less demonstrative (such as in other cultures), researchers may need to infer affect based on behaviors different from those used here, and may find it more difficult to assess affect through peripheral vision.

Inter-rater reliability from coding conducted with the guide was acceptably high. In total 706 observations were collected, for an average of 19.6 observations per student. Inter-rater reliability was acceptably high across all observations: Cohen's (1960)$\kappa = 0.71$ for usage observations, $\kappa = 0.63$ for observations of cognitive–affective states. This is in line with past reports of high agreement in affect coding between raters (Bartel and Saavedra, 2000). Although the kappa score for the states was lower

than the kappa for usage observations, it is on par with kappas reported by other researchers who have assessed the reliability of detecting naturally occurring emotional expressions (Ang et al., 2002; Grimm et al., 2006; Litman and Forbes-Riley, 2004; Shafran et al., 2003).

Within an observation, each observer coded one of the six cognitive–affective states (and neutral), as in the AutoTutor study:

1. **Boredom**—behaviors included in the coding guide included slouching, resting the chin on his/her palm; statements such as "Can we do something else?" or "This is boring!"
2. **Confusion**—behaviors included in the coding guide included scratching his/her head, repeatedly looking at the same interface elements; consulting with a fellow student or the teacher; looking at another student's work to determine what to do next; statements like, "I'm confused!" or "Why didn't it work?"
3. **Delight**—behaviors included in the coding guide included clapping hands; laughing with pleasure; statements such as, "Yes!" or "I got it!"
4. **Engaged concentration**—behaviors included in the coding guide included immersion, focus, and concentration on the system, with the appearance of positive engagement (as opposed to frustration); leaning towards the computer; mouthing solutions; pointing to parts of screen
5. **Frustration**—behaviors included in the coding guide included banging on the keyboard or mouse; pulling his/her hair; deep sighing; statements such as, "What's going on?!"
6. **Surprise**—behaviors included in the coding guide included sudden jerking or gasping; statement such as "Huh?" or "Oh, no!"
7. **Neutral**—behaviors included in the coding guide included coded when the student did not appear to be displaying any of the other cognitive–affective states or when the student's affect could not be determined for certain

Behavior categories were also coded, using the following coding scheme developed by Baker et al. (2004):

1. **On-task**—working within The Incredible Machine.
2. **On-task conversation**—talking to the teacher or another student about The Incredible Machine, or its puzzles.
3. **Off-task conversation**—talking about any other subject.
4. **Off-task solitary behavior**—any behavior that did not involve The Incredible Machine or another individual (such as reading a magazine or surfing the web).
5. **Inactivity**—instead of interacting with other students or the software, the student instead stares into space or puts his/her head down on the desk.
6. **Gaming the system**—sustained and/or systematic guessing, such as arranging objects haphazardly or trying an object in every conceivable place. Also, repeated and

rapid help requests used to iterate to a solution, without reflection were coded as gaming.

3.3. Study 3—Aplusix

3.3.1. Participants

The participants in the Aplusix study were first and second year high school students from four schools within Metro Manila and one school in Cavite, a province south of Manila. Students' age ranged from 12 to 15 with an average age of 13.5. In total 140 students completed the study (83 female, 57 male).

3.3.2. Interaction procedure

Students used Aplusix in groups of ten, one student per computer. Each student used Aplusix for 45 min.

3.3.3. Cognitive–affective state and behavior judgment procedure

The observers for Aplusix were taken from the same pool of observers that collected the data for the Incredible Machine. They followed a process almost identical to the one detailed in the previous section. The only differences were as follows: one pair of observers observed each group of 10 students. They observed each student for 20 s before proceeding to the next. There were 180 s between each observation of a single student.

Thirteen pairs of observations were collected per student, totaling 3640 observations in all. Inter-rater reliability was acceptably high: Cohen's $\kappa = 0.78$ for usage observations, $\kappa = 0.63$ for observations of cognitive–affective states.

4. Results and discussion

4.1. Incidence of cognitive–affective states

In this section, we report the incidence of each cognitive–affective state across the three learning environments. A summary of the results is shown in Fig. 5 and Table 2.

4.1.1. Incidence of states within each study

We conducted three separate repeated measures ANOVAs to investigate the cognitive–affective states that were prominent in each of the three learning environments (AutoTutor, the Incredible Machine, and Aplusix). We found that each of the ANOVAs were statistically significant: for AutoTutor, $F(6, 162) = 10.81$, $MSe = 0.023$, $p < 0.001$, partial $\eta^2 = 0.286$; for the Incredible Machine—$F(6, 210) = 63.94$, $MSe = 0.025$, $p < 0.001$, partial $\eta^2 = 0.646$; for Aplusix—$F(6, 834) = 953.10$, $MSe = 0.009$, $p < 0.001$, partial $\eta^2 = 0.873$.

Bonferroni posthoc tests yielded the following patterns in the data at the 0.05 significance level. For AutoTutor, the pattern was: (Surprise = Delight) < (Boredom = Confusion = Engaged Concentration = Frustration = Neutral). Therefore, learners interacting with AutoTutor are
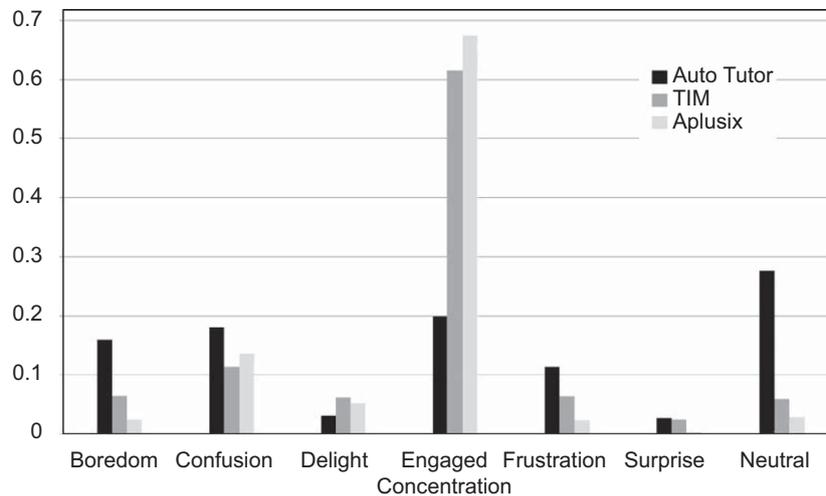
Fig. 5. The frequency of each affective state, in each learning environment.

Table 2
Descriptive statistics on proportions of each cognitive–affective state observed in each learning environment and aggregated across studies.

State	AT	TIM	A6	Averaged
Boredom	0.16 (0.026)	0.07 (0.025)	0.03 (0.006)	0.05
Confusion	0.18 (0.024)	0.11 (0.021)	0.13 (0.009)	0.13
Delight	0.03 (0.007)	0.06 (0.015)	0.05 (0.006)	0.05
Engaged concentration	0.20 (0.030)	0.62 (0.046)	0.68 (0.014)	0.60
Frustration	0.11 (0.020)	0.06 (0.018)	0.02 (0.004)	0.04
Surprise	0.03 (0.005)	0.03 (0.020)	0.003 (0.001)	0.01
Neutral	0.29 (0.248)	0.05 (0.121)	0.01 (0.029)	0.06

Standard errors given in parentheses.
Note: AT—AutoTutor, A6 Aplusix, TIM—Incredible Machine.

less likely to experience delight and surprise than the other states.

The pattern for the Incredible Machine was: (Boredom = Confusion = Delight = Frustration = Neutral) < Engaged Concentration. Experiences of surprise were on par with boredom, delight, frustration, and neutral but less than engaged concentration and confusion. Quite clearly, engaged concentration dominates when learners interact with this game-like learning environment.

The pattern for Aplusix was: (Boredom = Neutral = Frustration) < Delight < Confusion < Engaged Concentration. Experiences of surprise were on par with neutral but lower than the other states. Therefore, in a fashion similar to that observed with the Incredible Machine, engaged concentration dominates interactions with Aplusix.

4.1.2. Incidence of states, aggregated across studies

To give a better idea of the overall frequency of each state, we can also look at the average frequency of each state across environments. In doing so, we weight data from each student equally in order to statistically test for differences in the proportional occurrence of the cognitive–affective states. It should be noted, however, that weighting data from each student equally biases

estimates in favor of the larger Aplusix study. So the aggregated analyses are intended to complement, but not replace, the earlier analysis that investigated patterns within each study.

A repeated measures ANOVA indicated that there was a significant difference in the proportional distribution of the various states, $F(6, 1218) = 448.28$, $Mse = 0.020$, $p < 0.001$, partial $\eta^2 = 0.688$. Bonferroni posthoc tests revealed the following pattern in the data, (Surprise) < (Boredom = Frustration = Neutral = Delight) < Confusion < Engaged concentration.

Across the three learning environments, engaged concentration was the most common state, occupying an average of 60% of student time. Engaged concentration is hypothesized to be associated with positive affect and is also one of the components of flow (Csikszentmihalyi, 1990). Furthermore, engaged concentration is a cognitive–affective state that is positively correlated with learning (Craig et al., 2004; Graesser et al., 2007), so it is a positive sign that the experience of engaged concentration dominates across learning environments.

Confusion was the second most common state, with an occurrence of 13%. Confusion occurs when learners experience impasses that provide opportunities to learn, since students can resolve their confusion by

self-explanation. While episodes of confusion that go unresolved over longer periods of time have no pedagogical value, shorter-term instances of confusion are associated with learning (Craig et al., 2004; Graesser et al., 2007) as would be predicted by impasse-driven theories of learning (e.g. VanLehn et al., 2003). However, unmotivated or low-domain knowledge students' can also alleviate their confusion by avoiding activities that require deep thought, via (for example) gaming the system (cf. Rodrigo et al., 2007).

Boredom, frustration, neutral, and delight were each observed an average of 4–6% of the time. Surprise was by far the rarest state, occurring an average of 1% of the time across environments.

It is worth noting that the proportion of engaged concentration and neutral were unstable across environments. It is not possible to determine whether this was due to methodological differences (self-judging versus external observers, for instance), population differences, or differences between the learning environments. For this reason, we do not interpret this result within this paper, but simply note it as being worthy of further investigation. This type of finding would be better studied through a different form of research design than the one used here, a more standard design where every study design element is identical except for the system studied or judging methodology. Further discussion of these issues is given is Section 5.3.

4.2. Persistence of cognitive–affective states

In this section, we will focus our analysis on the persistence of cognitive–affective states. Persistence is operationally defined as the student being in the same state for two successive observations. In this analysis, we consider how a student's state at a given time influences their state in the successive observation. Within the study conducted with The Incredible Machine, successive observations of a single student were 40 s apart: 20 s observing, 40 s not observing, 20 s observing, and so on. Within the study conducted in Aplusix, successive observations of a single student were 180 s apart: 20 s observing, 180 s not observing, 20 s observing, and so on. Within the study conducted with AutoTutor, there was no delay between observations: 20 s observing, 20 more s observing, and so on.

We investigate the following question: Do different states have different overall degrees of persistence? Specifically, do some cognitive–affective states persist substantially longer than other states, across studies and learning environments? And, conversely, are some states always short-lived? Understanding the general persistence of each cognitive–affective state will be useful for focusing future research on affect in learning software.

4.2.1. Metrics
In each of the three studies, we analyze the persistence of a cognitive–affective state using a transition likelihood

metric, L. L provides an indication of the probability of a transition above and beyond the base rate of each cognitive–affective category. For instance, engaged concentration was the most common state in The Incredible Machine and Aplusix, whereas neutral was the most common state in AutoTutor; therefore, these states are likely to be the most common state that follows any other cognitive–affective state in these environments.

L explicitly accounts for the base rate of each state when assessing how likely a transition is, giving the probability that a transition between two states occurs, and given the base frequency of the destination state. L is computed as shown below

$$L = \frac{\Pr(Next \,|\, Prev) - \Pr(Next)}{1 - \Pr(Next)} \qquad (1)$$

A value of 1 means that the transition will always occur, whereas a value of 0 means that the transition's likelihood is exactly what it would be given only the base frequency of the destination state. Values above 0 signify that the transition is more likely than it could be expected (i.e. greater than the base frequency of the destination state), and values under 0 signify that the transition is less likely (i.e. less than the base frequency of the destination state).

For a given transition, we calculate a value for L for each student and then calculate the mean and standard error across students. We can then determine if a given transition is significantly more likely than chance (chance = 0) using the two-tailed t-test for one sample. The number of degrees of freedom for the two-tailed t-test is the number of students who were ever in the state minus one (df = $N - 1$). Students that never entered the state give no evidence on whether the state is persistent. As a consequence, the number of degrees of freedom varies among cognitive–affective states within each learning environment.

4.2.2. Results
The cognitive–affective state of boredom was the most persistent state. Evidence for this state's persistence was obtained in all three learning environments. The mean value of the transition likelihood metric L was significantly or marginally significantly above chance (chance $L = 0$) for all three learning environments: for AutoTutor, mean $L = 0.130$, $t(27) = 4.17$, $p < 0.001$, for The Incredible Machine (marginally), mean $L = 0.261$, $t(7) = 2.27$, $p = 0.06$, and for Aplusix, mean $L = 0.212$, $t(38) = 3.69$, $p < 0.01$. Hence, boredom appears to be a persistent state across all three learning environments.

Four other states appeared to be persistent across at least two of the three learning environments. Confusion was persistent within AutoTutor, mean $L = 0.087$, $t(27) = 2.71$, $p = 0.01$, and within The Incredible Machine (marginally), mean $L = 0.097$, $t(19) = 1.69$, $p = 0.11$. Confusion's persistence was not significantly different from chance in Aplusix, mean $L = 0.0076$, $t(119) = 0.32$, $p = 0.75$.

Engaged concentration was persistent within AutoTutor, mean $L = 0.100$, t(27) = 2.69, p = 0.01, and within The Incredible Machine (marginally), mean $L = 0.202$, t(33) = 1.75, p = 0.09. Engaged concentration's persistence was not significantly different from chance in Aplusix, mean $L = 0.062$, t(138) = 1.03, p = 0.30.

Frustration was marginally persistent within Aplusix, mean $L = 0.071$, t(35) = 1.95, p = 0.06, and within Auto-Tutor, mean $L = 0.059$, t(27) = 1.87, p = 0.07. Frustration's persistence was not significantly different from chance in The Incredible Machine, mean $L = 0.12$, t(12) = 1.56, p = 0.14.

Delight was persistent within The Incredible Machine (marginally), mean $L = 0.096$, t(16) = 1.71, p = 0.11, and within Aplusix (marginally), mean $L = 0.047$, t(68) = 1.82, p = 0.07. Delight's persistence was not significantly different from chance in AutoTutor, mean $L = -0.002$, t(27) = -0.20, p = 0.39.

A fifth state was considerably less persistent across environments than the other cognitive–affective states: surprise. Surprise was not persistent within any environment, which is what could be expected, as it is difficult to image a learner sustaining a state of surprise for more than a few seconds. Within AutoTutor, the relationship actually pointed in the opposite direction, towards non-persistence, to a statistically significant degree, mean $L = -0.026$, t(27) = -7.33, p < 0.01. Within Aplusix, the relationship also pointed in the opposite direction, but was so rare across students that it was not possible to calculate statistical significance, mean $L = -0.004$. Within The Incredible Machine, the relationship was in the direction of persistence, but was not statistically significant, mean $L = 0.031$, t(7) = 0.54, p = 0.61.

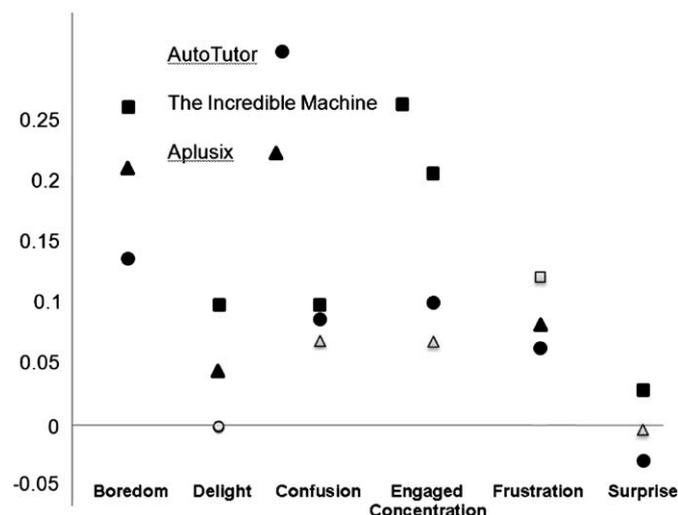The overall pattern of persistence of different cognitive–affective states is shown in graphical form in Fig. 6.

Fig. 6. The persistence of affective states (mean L) for each environment and each affective state. Solid dots signify a *p*-level of 0.11 or lower; gray dots signify lack of statistical significance.

4.2.3. Implications

The pattern of results suggests that, boredom is the most persistent state, and perhaps the closest to being a "mood". The probability that a bored learner will stay bored is significant in all three systems and is much higher than the average for the other cognitive–affective states. Only in one case across all of the other 15 mean values of L across environments is there a value as high as the lowest of boredom's mean L values (see Fig. 6).

On the other hand, surprise is the most transitory of the states. If anything, there seems to be evidence that surprise is non-persistent. Surprise had mean L values below 0 in two of three cases (with one case statistically significant).

4.3. Cognitive–affective states and student behavior: evidence from The Incredible Machine and Aplusix

One important question when studying student affect within interactive learning environments is how a student's affect impacts their learning experience. Negative affect is in many ways a problem in itself—so we would prefer not to develop learning environments in which students are continually experiencing negative affect, such as boredom or frustration. Nevertheless, the primary goal of interactive learning environments is learning. Frustration and confusion may be a natural and unavoidable part of the experience of learning when difficult material is encountered. Thus, the goal of a system may not be to entirely eliminate negative affect, especially if negative affect is sometimes a byproduct of positive learning experiences.

However, there is the possibility that negative affect may lead students to use (or fail to use) learning environments in ways that reduce their learning. For example, Baker and his colleagues have found that students sometimes engage in gaming the system, systematically guessing or abusing hint features in order to perform well in an interactive learning environment without learning the material. (Baker et al., 2004). Baker et al. (2008) found that students who report frustration on a questionnaire game significantly more than students who do not report frustration. Gaming the system has been found in multiple studies to be associated with poorer learning (Baker et al., 2004; Baker, 2005; Cocea et al., 2009; Walonoski and Heffernan, 2006). If negative affect leads to significantly more gaming, as Baker et al. (2008) seems to suggest, negative affect may indirectly reduce students' learning.

As we have data on gaming behavior only from The Incredible Machine and Aplusix, our analyses will be restricted to data from these two systems.

4.3.1. Metrics

To determine whether any cognitive–affective state is associated with increased gaming behavior, we use essentially the same approach as used in the previous section to determine persistence. We compute the transition likelihood metric L, to determine how likely gaming the system is to follow a given state. This approach takes

into account gaming the system's overall relative frequency, compared to other usage behaviors. Once we have computed the value of L, for each transition and student, we will as before determine if a given transition is significantly more likely than chance (chance = 0), given the base frequency of gaming, using the two-tailed t-test for one sample.

$$L = \frac{\Pr(GAMING^{timeN+1}|AFFECT^{timeN}) - \Pr(GAMING^{timeN+1})}{(1 - \Pr(GAMING^{timeN+1}))}$$

(2)

Results: Gaming the system occurred in both Aplusix and The Incredible Machine in sufficient quantity to analyze these behaviors. While there was the appearance of a difference in frequency between the environments, this comparison is out of the scope of this paper. Boredom was significantly more likely than chance to lead to gaming the system within Aplusix, mean $L = 0.129$, $t(38) = 2.38$, two-tailed p = 0.02. Boredom was not significantly more likely than chance to lead to gaming the system within The Incredible Machine, mean $L = 0.037$, $t(7) = 0.67$, two-tailed p = 0.53. However, there appeared to be some evidence that boredom may have impacted some students differently than others. Students who were bored over a third of the time in the Incredible Machine gamed the system an average of 26% of the time after being bored, which was marginally significantly more frequent than chance, despite being an extremely small sample, mean $L = 0.195$, $t(2) = 3.38$, two-tailed p = 0.08. By contrast, the 5 students who were bored less than a third of the time never gamed the system after being bored. The difference in gaming frequency between the less frequently bored and more frequently bored students was statistically significant, $t(7) = 5.55$, two-tailed p < 0.001.

Delight was not significantly more likely than chance to be followed by gaming within Aplusix, mean $L = 0.012$, $t(68) = 0.94$, two-tailed $p = 0.35$. Delight was never followed by gaming among any of the 16 students who were delighted at least once in The Incredible Machine.

Confusion was never followed by gaming among any of the 121 students who were confused at least once in Aplusix. In addition, confusion was not significantly more likely than chance to be followed by gaming within The Incredible Machine, mean $L = 0.032$, $t(19) = 0.63$, two-tailed $p = 0.54$.

Engaged concentration was not significantly more likely than chance to be followed by gaming within Aplusix, mean $L = -0.004$, $t(138) = -0.90$, two-tailed $p = 0.37$. In addition, engaged concentration was not significantly more likely than chance to be followed by gaming within The Incredible Machine, mean $L = -0.022$, $t(33) = -0.65$, two-tailed $p = 0.52$.

Frustration was never followed by gaming among any of the 39 students who were frustrated at least once in Aplusix. In addition, frustration was not significantly more likely than chance to be followed by gaming within The
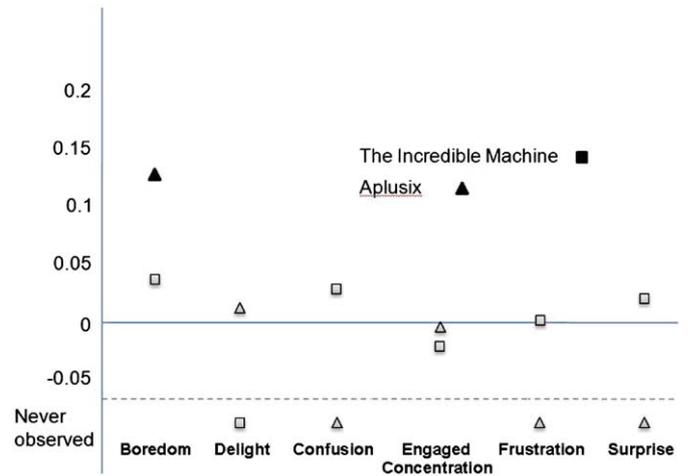

Fig. 7. The likelihood of observing gaming the system (mean L) after each affective state, within The Incredible Machine and Aplusix. Solid dots signify a *p*-level of 0.05 or lower; gray dots signify lack of statistical significance.

Incredible Machine, mean $L = 0.000$, $t(33) = 0.01$, two-tailed $p = 0.99$.

Surprise was never followed by gaming among any of the 10 students who were frustrated at least once in Aplusix. In addition, surprise was not significantly more likely than chance to be followed by gaming within The Incredible Machine, mean $L = 0.020$, $t(8) = 0.18$, two-tailed $p = 0.86$.

The overall pattern of how likely each cognitive–affective state is to be followed by gaming the system is shown in graphical form in Fig. 7.

This pattern of results suggests that, of the set of states studied, boredom is the only state that leads students to game the system (or at least, the only state for which there is evidence for this conclusion). Gaming the system is known to be associated with poorer learning in some learning environments (Baker et al., 2004). Hence, boredom may reduce learning more than other cognitive–affective states by leading students to engage in gaming behaviors which are associated with poorer learning.

5. General discussion

5.1. Summary of findings, and implications

In this paper, we have examined data on students' cognitive–affective states as they use three educational environments: AutoTutor, a dialogue-based tutor on computer literacy, Aplusix, a problem-solving-based tutor on mathematics, and The Incredible Machine, a game based on solving logic puzzles.

We analyzed the prevalence of a set of six cognitive–affective states (and neutral) in the three environments. Engaged concentration was the most common state when the data was aggregated across the three environments, showing that engagement is the norm in learning with technology (at least over short periods of study). Confusion

was the second most common state across environments, thereby substantiating the significant role of confusion (also referred to as perplexity) in complex learning (Craig et al., 2004; Festinger, 1957; Graesser et al., 2005; Guhe et al., 2004). When the learner is confused, they are in the state of cognitive disequilibrium, heightened physiological arousal, and more intense thought. Other states were considerably less common.

A second set of analyses presented within this paper addressed the persistence of different cognitive–affective states within the three environments. Within all three environments, the most persistent state was boredom, whereas the least persistent state was surprise. The other four states studied (frustration, engaged concentration, confusion, and delight) were less persistent than boredom in all environments, but were still significantly more persistent than could be expected by chance in at least some environments. The pattern of results suggests that, of the set of cognitive–affective states studied, boredom is the closest to being a non-transitory "mood". Once a student is bored, it appears to be difficult to transition out of boredom—suggesting that it is important to prevent boredom before it ever occurs.

Finally, we examined the relationships between the cognitive–affective states and the choice to game the system, a behavior known to be associated with poorer learning (Baker et al., 2004). Boredom was found to significantly increase the chance that a student will game the system on the next observation. In contrast, none of the other five cognitive–affective states were found to be associated with gaming.

Our findings suggest that boredom is the primary cognitive–affective state which interactive learning environments should focus on detecting and quickly responding to. In all three systems, boredom was the most persistent state. Boredom also was uniquely associated with gaming the system, a behavior known to lead to significantly poorer learning. Furthermore, research with AutoTutor (Craig et al., 2004; Graesser et al., 2008) has reported that boredom is significantly negatively correlated with learning.

Our recommendation to focus on boredom is not aligned with past research in the human–computer interaction community, which has focused on detecting and responding to frustration more than other cognitive–affective states (cf. Hone, 2006; Klein et al., 2002). Our results suggest, by contrast, that boredom should receive greater research attention than frustration, and that in many cases frustration may not need remediation. Mentis (2007) also argues that frustration does not always require remediation. In Mentis's perspective, frustration among users of information systems is only of concern if it is associated with events that are outside of the user's locus of control, such as a program bug. A frustrating event of this nature interrupts the user's cognitive flow, leading to a negative cognitive loop that causes the uses to keep selecting the same erroneous interface options over and over again. Sometimes frustration is not attributable to an external event. If frustration is a natural part of a cognitive processing activity, it might not require external intervention.

5.2. Generalizability of results to other domains and interfaces

The three studies investigated a set of six cognitive–affective states in learning environments using human judgments, but differed in many other fashions. The three learning environments studied had different interface qualities and pedagogical principles (game, dialogue tutor, problem-solving tutor), different material (concrete manipulation, computer literacy, mathematics), were used by different age groups (university and high school), in different settings (laboratory and classroom), and in different countries (the USA and the Philippines) (see Table 1). Despite these differences, there were many similarities in the profiles of cognitive–affective states.

The diversity of research contexts makes interpretation of differences among environments difficult. However, it enables us to have more confidence about the generality of the results that the environments had in common. Any result that is the same in three different studies, in such radically different contexts, is reasonably robust. Specifically, the evidence for the persistence of boredom is strong. Not only is this effect statistically significant in three different studies, it is consistent across type of learning environment, material, age group, setting, and country. The evidence for the positive relationship between gaming the system and boredom is also strong, persisting across the two learning environments where it could be assessed. The low incidence of surprise across environments is noteworthy, as were the high incidences of confusion and engaged concentration.

Given the considerable variation between these three studies, the commonalities found are more interpretable than they might be if the three studies were more similar. While it is clearly not appropriate to claim that these patterns will apply across all contexts, they do apply in fairly divergent situations. It will be valuable, in future research, to see under what conditions these patterns are not seen, or can be disrupted. In particular, research on environments where boredom is less persistent may support new designs that respond effectively to student boredom. We discuss this topic further in Section 5.4.

5.3. Self-report versus observer judgments

Although the primary focus of this paper was on similarities across environments, there were also some differences between the affective profiles in the three environments. These differences have not been a focus of this paper, because it is difficult to determine whether these differences emerged from genuine differences between the environments, or from differences in the populations (i.e.

age, culture, and computer exposure). It is also likely that methodological factors led to some of the differences in the affective profiles across systems.

One methodological difference of particular importance is that self-reports were utilized as the measure of a student's cognitive–affective state in the AutoTutor study whereas trained judges provided the judgments for the Incredible Machine and Aplusix studies. It is likely that learner self-reports are more sensitive to particular certain cognitive–affective states, while a different set of states are likely to be particularly salient to trained judges. Although a systematic investigation of the cognitive–affective profiles obtained from self-reports versus other judges is beyond the scope of this paper, we briefly describe another recent study conducted with AutoTutor, to inform our readers' future use of these methods.

In this study (Graesser et al., 2006), peers and trained judges provided judgments of a learner's cognitive–affective state during use of AutoTutor (Graesser et al., 2006). By comparing the affective profiles coming from self-report data, peer judges, and trained judges, it was possible to investigate what impact the coding method has on the findings. The results of this study indicated that there was no difference in the occurrence of boredom, engaged concentration, neutral, delight and surprise between the self and other judges. There were differences in the proportion of confusion (less self-report of confusion than in the external judges) and frustration (more self-report of frustration). This suggests that the methods are comparable, except in some confounding of these two specific fairly similar cognitive–affective states. Trained judges (the type of judgment used in the Aplusix and The Incredible Machine studies in this paper) had significantly better agreement with self-report than the peer judges did. However, it is not entirely clear whether trained judges or self-report should be considered the gold standard for assessment of cognitive–affective states. Importantly, actor–observer biases may play an important role in the judgment of ill-defined constructs such as the cognitive–affective states that arise during learning (Jones and Nisbett, 1971). When possible, using both forms of coding is probably the most defensible approach. In practice, there are many situations where one approach or the other is not feasible. For example, continual self-report is too disruptive to use in many educational tasks, but trained observers are not available in all situations and do not scale well.

5.4. Future work

Two lines of future work emerge from the three studies presented here. The first potential direction involves studying how cognitive–affective states vary between environments. We have reported results that address factors that appear not to vary between environments, but the high-variation-between-studies method used here sheds little light on the factors that do vary between environments. Future work will need to explicitly contrast different types of learning environments, research methods, and populations.

The second line of future work is to develop learning systems that can detect and respond to boredom and confusion. This work will have two components: detecting these states, and responding to them in a manner that promotes positive affect, learning, and engagement. We have already made some substantial advances in automatically detecting these cognitive–affective states by monitoring facial features, gross body language, and conversational cues (D'Mello et al., 2007). Therefore, the next challenge is to devise pedagogical and motivational strategies to respond to these negative cognitive–affective states in order to maximize learning. One method to respond to boredom would be to engage the learner in an activity that increases interest. These might include options of choice, increased challenge, or embedded games. The systems need to somehow shift users out of frustration to more positive affect states (Hone, 2006; Klein et al., 2002; McQuiggan et al., 2007). However, given boredom's persistence and the low arousal that students become "stuck in", it may be more difficult to shift students to more positive cognitive–affective states. Nonetheless, the integration of these two components into educational software that responds effectively to differences in boredom seems achievable.

Since confusion is a cognitive–affective state that accompanies deep learning and is linked to learning gains, it is important for learning environments to manage the learner's confusion productively. Some learners tend to give up when they are confused because they attribute their confusion to having low ability in general (Dweck, 2002; Meyer and Turner, 2006); these learners need to be encouraged and also informed that working on the problem will be fruitful and that confusion is a sign of progress. Other learners become motivated when they are confused because it is a signal that they are being challenged, and they have confidence in their ability to conquer the challenge. Although the optimal pedagogical strategy to help learners' regulate their confusion is unclear, mechanisms will need to be sensitive to cognitive and motivational characteristics of the learners in addition to their emotional states.

We eventually hope to create affect-sensitive learning environments that respond constructively and effectively to boredom and confusion. When we do, we will have made significant progress towards improving students' learning experiences, reducing problem behaviors such as gaming the system, managing students' frustration and confusion in the face of impasses, and ultimately improving students' learning.

Acknowledgements

References

Anderson, J.R., Corbett, A.T., Koedinger, K.R., Pelletier, R., 1995. Cognitive tutors: lessons learned. The Journal of the Learning Sciences 4, 167–207.

Ang, J., Dhillon, R., Krupski, A., Shriberg, E., Stolcke, A., 2002. Prosody-based automatic detection of annoyance and frustration, Human–Computer Dialog. Proceedings of the International Conference on Spoken Language Processing, Denver, CO, pp. 2037–2039.

Arnold, J., 1999. Affect in Language Learning. Cambridge University Press, Cambridge, UK.

Baker, R.S., 2005. Designing intelligent tutors that adapt to when students game the system. Doctoral Dissertation. CMU Technical Report CMU-HCII-05-104.

Baker, R.S.J.d., 2007. Modeling and understanding students' off-task behavior in intelligent tutoring systems, Proceedings of ACM CHI 2007: Computer–Human Interaction, pp. 1059–1068.

Baker, R.S.J.d, Corbett, A.T., Aleven, V., Koedinger, K.R., de Carvalho, A.M.J.A., Raspat, J., 2009. Educational software features that encourage and discourage "Gaming the System", Proceedings of the 14th International Conference on Artificial Intelligence in Education, pp. 475–482.

Baker, R.S.J.d, Corbett, A.T., Koedinger, K.R., Evenson, S.E., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., Beck, J., 2006.

Adapting to when students game an intelligent tutoring system, Proceedings of the Eighth International Conference on Intelligent Tutoring Systems, pp. 392–401.

Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z., 2004. Off-task behavior in the cognitive tutor classroom: when students "Game the System", Proceedings of ACM CHI 2004: Computer–Human Interaction, pp. 383–390.

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

Baker, R.S.J.d, Corbett, A.T., Wagner, A.Z., 2006. Human classification of low-fidelity replays of student actions, Proceedings of the Educational Data Mining Workshop at the Eighth International Conference on Intelligent Tutoring Systems, pp. 29–36.

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., Koedinger, K., 2008. Why students engage in "Gaming the System" behavior in interactive learning environments. Journal of Interactive Learning Research 19 (2), 185–224.

Barrett, L.F., 2006. Are emotions natural kinds?. Perspectives on Psychological Science 1, 28–58.

Bartel, C.A., Saavedra, R., 2000. The collective construction of work group moods. Administrative Science Quarterly 45 (2), 197–231.

Bloom, B.S., 1984. The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher 13, 3–16.

Bower, G., 1992. How might emotions affect learning. In: Christianson, S.A. (Ed.), The Handbook of Emotion and Memory: Research and Theory. Erlbaum, Hillsdale, NJ, pp. 3–31.

Burleson, W., 2006. Affective learning companions: strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive and meta-affective approaches to learning, motivation, and perseverance. Doctoral Thesis, Massachusetts Institute of Technology.

Carroll, J.M., 1997. Human–computer interaction: psychology as a science of design. Annual Review of Psychology 48, 61–83.

Castellano, G., Villalba, S.D., Camurri, A., 2007. Recognizing human emotions from body movement and gesture dynamics. In: Paiva, A., Prada, R., Picard, R.W. (Eds.), Affective Computing and Intelligent Interaction. Springer-Verlag, Berlin, Heidelberg, pp. 71–82.

Clifford, M.M., 1988. Failure tolerance and academic risk-taking in ten- to twelve-year-old students. British Journal of Educational Psychology 58, 15–27.

Clifford, M.M., 1991. Risk taking: theoretical, empirical, and educational considerations. Educational Psychologist 26, 263–298.

Cocea, M., Hershkovitz, A., Baker, R.S.J.d., 2009. The impact of off-task and gaming behaviors on learning: immediate or aggregate?, Proceedings of the 14th International Conference on Artificial Intelligence in Education, pp. 507–514.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20, 37–46.

Cohen, P.A., Kulik, J.A., Kulik, C.C., 1982. Educational outcomes of tutoring: a meta-analysis of findings. American Educational Research Journal 19, 237–248.

Conati, C., 2002. Probabilistic assessment of user's emotions in educational games. Journal of Applied Artificial Intelligence 16, 555–575.

Conati, C., Zhao, X., 2004. Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game, Proceedings of the Ninth International Conference on Intelligent User Interface, pp. 6–13.

Corbett, A.T., Anderson, J.R., 1995. Knowledge tracing: modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction 4 (4), 253–278.

Craig, S.D., Graesser, A.C., Sullins, J., Gholson, B., 2004. Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. Journal of Educational Media 29 (3), 241–250.

Csikszentmihalyi, M., 1990. Flow: The Psychology of Optimal Experience. Harper-Row, New York.

de Vicente, A., Pain, H., 2002. Informing the detection of the students' motivational state: an empirical study, Proceedings of the Sixth International Conference on Intelligent Tutoring Systems, pp. 933–943.

D'Mello, S.K., Chipman, P., Graesser, A.C., 2007. Posture as a predictor of learner's affective engagement, Proceedings of the 29th Annual Cognitive Science Society, pp. 905–910.

D'Mello, S.K., Taylor, R., Graesser, A.C., 2007. Monitoring affective trajectories during complex learning, Proceedings of the 29th Annual Cognitive Science Society, pp. 203–208.

Dodds, P., Fletcher, J.D., 2004. Opportunities for new "smart" learning environments enabled by next-generation web capabilities. Journal of Educational Multimedia and Hypermedia 13 (4), 391–404.

Dweck, C.S., 2000. Self-Theories: Their Role in Motivation, Personality, and Development. Psychology Press, Philadelphia, PA.

Dweck, C.S., 2002. Messages that motivate: how praise molds students' beliefs, motivation, and performance (in surprising ways). In: Aronson, J. (Ed.), Improving Academic Achievement: Impact of Psychological Factors on Education. Academic Press, Orlando, FL, pp. 61–87.

Ekman, P., Friesen, W.V., 1969. Nonverbal leakage and clues to deception. Psychiatry 32, 88–105.

Ekman, P., Friesen, W.V., 1971. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology 17, 124–129.

Ekman, P., Friesen, W.V., 1978. The Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, CA.

Elfenbein, H.A., Ambady, N., 2002a. On the universality and cultural specificity of emotion recognition: a meta-analysis. Psychological Bulletin 128, 203–235.

Elfenbein, H.A., Ambady, N., 2002b. Is there an ingroup advantage in emotion recognition?. Psychological Bulletin 128, 243–249.

Festinger, L., 1957. A Theory of Cognitive Dissonance. Row, Perterson & Company, Evanston, IL.

Fredrickson, B.L., Branigan, C., 2005. Positive emotions broaden the scope of attention and thought–action repertoires. Cognition and Emotion 19, 313–332.

Gee, J.P., 2004. Situated Language and Learning: A Critique of Traditional Schooling. Routledge, Taylor & Francis, London, UK.

Graesser, A.C., D'Mello, S.K., Chipman, P., King, B., McDaniel, B., 2007. Exploring relationships between affect and learning with AutoTutor. In: Luckin, R., Koedinger, K., Greer, J. (Eds.), Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work. IOS Press, Amsterdam, pp. 16–23.

Graesser, A.C., Rus, V., D'Mello, S., Jackson, G.T., 2008. AutoTutor: learning through natural language dialogue that adapts to the cognitive and affective states of the learner. In: Robinson, D.H., Schraw, G. (Eds.), Current Perspectives on Cognition, Learning and Instruction: Recent Innovations in Educational Technology that Facilitate Student Learning. Information Age Publishing, pp. 95–125.

Graesser, A.C., D'Mello, S., Person, N.K., 2009. Metaknowledge in tutoring. In: Hacker, D., Donlosky, J., Graesser, A.C. (Eds.), Handbook of Metacognition in Education. Taylor & Francis, Mahwah, NJ.

Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., Louwerse, M.M., 2004. AutoTutor:a tutor with dialogue in natural language. Behavior Research Methods, Instruments, and Computers 36, 180–193.

Graesser, A.C., Lu, S., Olde, B.A., Cooper-Pye, E., Whitten, S., 2005. Question asking and eye tracking during cognitive disequilibrium: comprehending illustrated texts on devices when the devices break down. Memory and Cognition 33, 1235–1247.

Graesser, A.C., Person, N., Harter, D., Tutoring Research Group, 2001. Teaching tactics and dialogue in AutoTutor. International Journal of Artificial Intelligence in Education 12, 257–279.

Graesser, A.C., Witherspoon, A., McDaniel, B., D'Mello, S., Chipman, P., Gholson, B., 2006. Detection of emotions during learning with AutoTutor, Proceedings of the 28th Annual Meeting of the Cognitive Science Society. Erlbaum, Mahwah, NJ, pp. 285–290.

Grimm, M., Kroschel, K., Harris, H., Nass, C., Schuller, B., Rigoll, G., Moosmayr, T., 2007. On the necessity and feasibility of detecting driver's emotional state while driving. In: Paiva, A., Prada, R., Picard, R.W. (Eds.), Affective Computing and Intelligent Interaction. Springer-Verlag, Berlin, Heidelberg, pp. 126–138.

Grimm, M., Mower, E., Kroschel, K., Narayan, S., 2006. Combining categorical and primitives-based emotion recognition, 14th European Signal Processing Conference (EUSIPCO), Florence, Italy.

Guhe, M., Gray, W.D., Schoelles, M.J., Ji, Q., 2004. Towards an affective cognitive architecture. Poster presented at the Cognitive Science Conference, Chicago, IL.

Hess, U., Kappas, A., Scherer, K.R., 1988. Multichannel communication of emotion: synthetic signal production. In: Scherer, K.R. (Ed.), Facets of Emotion: Recent Research. Erlbaum, Hillsdale, NJ, pp. 161–182.

Hone, K., 2006. Empathic agents to reduce user frustration: the effects of varying agent characteristics. Interacting with Computers 18, 227–245.

Hudlicka, E., McNeese, D., 2002. Assessment of user affective and belief states for interface adaptation: application to an Air Force pilot task. User Modeling and User-Adapted Interaction 12 (1), 1–47.

Jones, E., Nisbett, R., 1971. The Actor and the Observer: Divergent Perceptions of the Causes of Behavior. General Learning Press, New York.

Kapoor, A., Burleson, W., Picard, R.W., 2007. Automatic prediction of frustration. International Journal of Human–Computer Studies 65 (8), 724–736.

Klein, J., Moon, Y., Picard, R., 2002. This computer responds to user frustration—theory, design, and results. Interacting with Computers 14 (2), 119–140.

Koedinger, K.R., Corbett, A.T., 2006. Cognitive tutors: technology bringing learning sciences to the classroom. In: Sawyer, R.K. (Ed.), The Cambridge Handbook of the Learning Sciences. Cambridge University Press, New York, NY.

Komatsu, T., Ohtsuka, S., Ueda, K., 2007. Comprehension of users' subjective interaction states during their interaction with an artificial agent by means of heard rate variability index. In: Paiva, A., Prada, R., Picard, R.W. (Eds.), Affective Computing and Intelligent Interaction. Springer-Verlag, Berlin, Heidelberg, pp. 266–277.

Kort, B., Reilly, R., Picard, R., 2001. An affective model of interplay between emotions and learning: reengineering educational pedagogy—building a learning companion, Proceedings IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges. IEEE Computer Society, Madison, Wisconsin pp. 43–48.

Lehman, B.A., Matthews, M., D'Mello, S.K., Person, N., 2008. Understanding students' affective states during learning. In: Woolf, B. (Ed.), Ninth International Conference on Intelligent Tutoring Systems. ITS 2008, Lecture Notes in Computer Sciences, vol. 5091. Springer-Verlag, pp. 50–59.

Lehman, B., D'Mello, S.K., Person, N., 2008. All Alone with your emotions: an analysis of student emotions during effortful problem solving activities, Supplementary Proceedings of the Workshop on Emotional and Cognitive issues in ITS (WECITS).

Lepper, M.R., Cordova, D.I., 1992. A desire to be taught: instructional consequences of intrinsic motivation. Motivation and Emotion 16 (3), 187–208.

Litman, D.J., Forbes-Riley, K., 2004. Predicting student emotions in computer–human tutoring dialogues, Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, East Stroudsburg, PA pp. 352–359.

Litman, D.J., Silliman, S., 2004. ITSPOKE: an intelligent tutoring spoken dialogue system, Proceedings of the human Language Technology Conference: Third Meeting of the North American Chapter of the Association of Computational Linguistics, pp. 52–54.

Mandler, G., 1984. Mind and Body: Psychology of Emotion and Stress. W.W. Norton & Company, New York.

Mandler, G., 1999. Emotion. In: Bly, B.M., Rumelhart, D.E. (Eds.), Cognitive science. Handbook of perception and cognition (2nd ed.). Academic Press, San Diego, CA.

Mandryk, R.L., Atkins, M.S., 2007. A fuzzy physiological approach for continuously modeling emotion during interaction with play environments. International Journal of Human–Computer Studies 6 (4), 329–347.

Marinier, R.P., Laird, J.E., 2007. Computational modeling of mood and feeling from emotion, Proceedings of the Annual Meeting of the Cognitive Science Society.

Marinier, R., Laird, J.A., 2006. Cognitive Architecture Theory of Comprehension and Appraisal. Agent Construction and Emotion.

McNeese, M.D., 2003. New visions of human–computer interaction: making affect compute. International Journal of Human–Computer Studies 59, 33–53.

McQuiggan, S.W., Lee, S., Lester, J.C., 2007. Early prediction of student frustration, Proceedings of the Second International Conference on Affective Computing and Intelligent Interaction, pp. 698–709.

Mentis, H.M., 2007. Memory of frustrating experiences. In: Nahl, D., Bilal, D. (Eds.), Information and Emotion. Information Today, Medford, NJ.

Meyer, D.K., Turner, J.C., 2006. Re-conceptualizing emotion and motivation to learn in classroom contexts. Educational Psychology Review 18 (4), 377–390.

Miserandino, M., 1996. Children who do well in school: individual differences in perceived competence and autonomy in above-average children. Journal of Educational Psychology 88, 203–214.

Nicaud, J.-F., Bouhineau, D., Chaachoua, H., 2004. Mixing microworld and CAS features in building computer systems that help students learn algebra. International Journal of Computers for Mathematical Learning 9, 169–211.

Nicaud, J.F., Bouhineau, D., Mezerette, S., Andre, N., 2007. Aplusix II [Computer software].

Norman, D.A., 2004. Emotional Design: Why We Love (or Hate) Everyday Things. Basic Books, New York.

Paiva, A., Prada, R., Picard, R.W. (Eds.), 2007. Affective Computing and Intelligent Interaction. Springer Verlag, Berlin, Heidelberg.

Pantic, M., Rothkrantz, L.J.M., 2003. Towards an affect-sensitive multimodal human–computer interaction. Proceedings of the IEEE, Special Issue on Multimodal Human–Computer Interaction (HCI) 91 (9), 1370–1390.

Patrick, B., Skinner, E., Connell, J., 1993. What motivates children's behavior and emotion? Joint effects of perceived control and autonomy in the academic domain. Journal of Personality and Social Psychology 65, 781–791.

Perkins, R.E., Hill, A.B., 1985. Cognitive and affective aspects of boredom. British Journal of Psychology 76 (2), 221–234.

Peterson, P.L., Fennema, E., 1985. Effective teaching, student engagement in classroom activities, and sex-related differences in learning mathematics. American Educational Research Journal 22 (3), 309–335.

Picard, R., 1997. Affective Computing. MIT Press, Cambridge, MA.

Picard, R.W., Papert, S., Bender, W., Blumberg, B., Berezeal, C., Cavallo, D., Machover, T., Resnick, M., Roy, D., Strohecker, C., 2004. Affective learning—a manifesto. BT Technology Journal 22 (4), 253–269.

Planalp, S., DeFrancisco, V.L., Rutherford, D., 1996. Varieties of cues to emotion in naturally occurring settings. Cognition and Emotion 10 (2), 137–153.

Prendinger, H., Ishizuka, M., 2005. The empathic companion: a character-based interface that addresses users' affective states. International Journal of Applied Artificial Intelligence 19 (3–4), 267–285.

Prensky, M., 2007. Digital Game-Based Learning. Paragon House, St. Paul, MN.

Rodrigo, M.M.T., Baker, R.S.J.d., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sevilla, L.R.S., Sugay, J.O., Tep, S., Viehland, N.J.B., 2007. Affect and Usage Choices in Simulation Problem Solving Environments, Proceedings of the 13th International Conference on Artificial Intelligence in Education, pp. 145–152.

Russell, J., 2003. Core affect and the psychological construction of emotion. Psychological Review 110, 145–172.

Russell, J.A., 1994. Is there universal recognition of emotion from facial expression?: a review of the cross-cultural studies. Psychological Bulletin 115, 102–141.

Schutzwohl, A., Borgstedt, K., 2005. the processing of affectively valenced stimuli: the role of surprise. Cognition & Emotion 19, 583–600.

Shafran, I., Riley, M., Mohri, M., 2003. Voice signatures, Proceedings IEEE Automatic Speech Recognition and Understanding Workshop. IEEE, Piscataway, NJ, pp. 31–36.

Sierra Online., 2001. The Incredible Machine: Even More Contraptions [Computer Software].

Silvia, P., Abele, A., 2002. Can positive affect induce self-focused attention? Methodological and measurement issues. Cognition and Emotion 16, 845–853.

Sylwester, R., 1994. How emotions affect learning. Educational Leadership 52 (2), 60–65.

Stein, N.L., Levine, L.J., 1991. Making sense out of emotion. In: Kessen, W., Ortony, A., Kraik, F. (Eds.), Memories, Thoughts, and Emotions: ESSAYS in Honor of George Mandler. Erlbaum, Hillsdale, NJ, pp. 295–322.

Stein, N.L., Hernandez, M.W., Trabasso, T., 2008. Advances in modeling emotions and thought: the importance of developmental, online, and multilevel analysis. In: Lewis, M., Haviland-Jones, J.M., Barrett, L.F. (Eds.), Handbook of Emotions 3rd ed Guilford Press, New York, pp. 574–586.

VanLehn, K., 1990. Mind Bugs: The Origins of Procedural Misconceptions. MIT Press, Cambridge, MA.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J.A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., Wintersgill, M., 2005. The Andes physics tutoring system: five years of evaluations. International Journal of Artificial Intelligence and Education 15 (3), 1–47.

VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., Rose, C.P., 2007. When are tutorial dialogues more effective than reading?. Cognitive Science.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., Baggett, W., 2003. Why do only some events cause learning during human tutoring? Cognition and Instruction 21 (3), 209–249.

Walonoski, J.A., Heffernan, N.T., 2006. Detection and analysis of off-task gaming behavior in intelligent tutoring systems, Proceedings of the Eighth International Conference on Intelligent Tutoring Systems, pp. 382–391.

Wagner, J., Vogt, T., Andre, E., 2007. A systematic comparison of different HMM designs for emotion recognition from acted and spontaneous speech. In: Paiva, A., Prada, R., Picard, R.W. (Eds.), Affective Computing and Intelligent Interaction. Springer-Verlag, Berlin, Heidelberg, pp. 114–125.

Whang, M.C., Lim, J.S., Boucsein, W., 2003. Preparing computers for affective communication: a psychophysiological concept and preliminary results. Human Factors 45 (4), 623–634.