Research and Practice in Technology Enhanced Learning
Vol. 6, No. 1 (2011) 43－66


# COMPARING LEARNERS' AFFECT WHILE USING AN INTELLIGENT TUTOR AND AN EDUCATIONAL GAME

MA. MERCEDES T. RODRIGO

*Department of Information Systems and Computer Science, Ateneo de Manila University, Loyola Heights, Quezon City, Philippines*
*mrodrigo@ateneo.edu*

RYAN S. J. D. BAKER

*Worcester Polytechnic Institute, Worcester, Massachusetts, USA*
*rsbaker@wpi.edu*

We compare the affect associated with students learning from an intelligent tutoring system, Aplusix, and a game, Math Blaster 9-12, covering very similar mathematical content. Quantitative field observations of student affect were conducted in classrooms in private schools in the Philippines. Students experienced large amounts of positive affect in both environments. It has been hypothesized that educational games will lead to better affect than other forms of educational software, but it was found that students experienced more positive affect (specifically, engaged concentration) and less negative affect (specifically, boredom) in the intelligent tutor than in the game, though there was a trend towards more delight within the game. These results suggest that intelligent tutors may be more affectively successful than had previously been realized. An alternate possible implication is that games' motivational benefits are not solely in terms of moment-to-moment affective experience. At the same time, the differential affective benefits of the two genres of educational software suggest that some blend of the two types of learning environments may be more engaging than any existing educational software.

*Keywords*: Affect; intelligent tutoring systems; educational games; serious games; quantitative field observations.

## 1. Introduction

Games are fun. The same adolescents who are often reluctant to put significant time into their studies are often enthusiastically willing to spend dozens of hours on playing modern computer games (Brown, 2005). In recent years, it has been repeatedly suggested that embedding games into education can be a way to improve students' affect, interest, and motivation towards education, and in turn improve their learning (e.g. Gee, 2003; Squire, 2003; Shaffer et al., 2005; Wideman et al., 2007). In particular, it has been suggested that games lead to positive affect states (cf. Sherry, 2004; Voiskounsky et al., 2004), physical and mental changes that take place as a person experiences an emotion (Picard, 1997). Affect encompasses numerous and diverse human emotional and cognitive experiences, examples of which include amusement, anger, contempt,

contentment, disgust, embarrassment, excitement, fear, guilt, pride in achievement, relief, sadness/distress, satisfactions, sensory pleasure and shame (Ekman, 1999). Not all of these states, though, are relevant to learning and the identification of relevant affective states is a subject of both debate and research. Craig et al. (2004) suggest that the key learning-related affective states are boredom, confusion, eureka, flow, and frustration. D'Mello et al. (2005) later amended this list to boredom, confusion, delight, flow, frustration, neutral and surprise. Research since then has suggested that, of this set, boredom and confusion are particularly strongly related to learning (cf. Baker et al., 2010). In this study, we use D'Mello et al.'s (2005) list of affective states; we will discuss it in greater detail later in the text. Games have been hypothesized to be particularly strongly associated with the "Flow" experience, a form of heightened concentration and engrossment in an activity (Csikszentmihalyi, 1990), a benefit potentially extending to educational games as well.

Some of the most successful educational games have succeeded in improving student motivation and learning, successfully building upon factors such as competition, curiosity, challenge, fantasy, and well-designed game mechanics, in order to make learning more enjoyable, increase students' desire to learn, and lead students to complete more difficult work (Ainsworth & Habgood, 2009; Alessi & Trollip, 2001; Cordova & Lepper, 1996; Lee et al., 2004).

However, while it is commonly hypothesized that educational games will lead to better affect and motivation than non-game-like learning environments (cf. Gee, 2003; Prensky, 2007), the evidence supporting this hypothesis is not conclusive. In many cases, educational games have been studied in relation to relatively weak comparison conditions, such as paper worksheets with no feedback (Lee et al., 2004) and games with key features ablated, such as a decontextualized math game called "Math Game" (Cordova & Lepper, 1996). Despite the popularity of educational games among education researchers, with thousands of published papers describing different games (O'Neill et al., 2005), educational games seem to be viewed rather negatively by students (Bragg, 2007), and are only used by children around one hour a week outside of class (Kerawalla & Crook, 2005). One important factor may be how educational games are designed. Bruckman (1999) argues that many educational games are designed as "chocolate-dipped broccoli", failing to integrate fun elements with learning content. Vogel and colleagues (2006) correspondingly find that educational games that alternate between gameplay and didactic instruction fail to promote motivation and engagement. However, the best way to design educational games is not yet clear. Kafai (2001) argues that effective educational games must integrate subject matter into the game fantasy context in an intrinsic fashion. Ainsworth and Habgood, by contrast, argue that connection between the subject matter and game play is not as important as integration of the subject matter and core game mechanic (Ainsworth & Habgood, 2009; Habgood, 2007).

While issues of how to design games remain to be answered by the field, we feel that the design of many evaluations of games shows that there is also significant question as to how to evaluate games' concrete impacts on student affect and motivation. Many

studies beyond the ones mentioned above use weak or inappropriate comparison conditions (an extensive review of this topic is given in O'Neill et al., 2005). We propose that comparisons will be most meaningful if they are with another type of educational software (to avoid confounding attitudes towards a specific game with attitudes towards computers) which is known to be educationally successful and reasonably engaging (to show the genuine benefit of the educational game relative to a reasonable comparison condition).

One type of educational intervention that meets these conditions is the intelligent tutoring system. Within intelligent tutoring systems, a student engages in problem-solving tasks and receives automated help and feedback tailored to his or her knowledge (Wenger, 1987). Intelligent tutors have been shown to lead to significantly better performance on standardized mathematics exams than traditional curricular methods (cf. Koedinger & Corbett, 2006). Past qualitative research has suggested that intelligent tutoring systems also may lead to significantly improved affect and motivation as compared to traditional, non-computerized learning contexts (Schofield, 1995).

It is worth noting that some intelligent tutors have been built into games, and that game features have been incorporated into other intelligent tutoring systems, leading to systems that can be considered both games and intelligent tutoring systems. For example, Goldstein (1979) built intelligent tutoring support into a game, WUMPUS, that required logical and geometric reasoning to avoid getting eaten by a cave-dwelling monster. Burton and Brown (1982) built intelligent tutoring support into an educational game designed to teach students about arithmetic and operator precedence. The version of the game with intelligent tutor support was reported by students to be more enjoyable than the original version of the game (Burton & Brown, 1982). More recently, the Tactical Language and Culture Training System (TLCTS) (Johnson, 2007) integrated intelligent tutoring support into the dynamic behavior of adaptive agents within an educational game that teaches foreign language and cultural skills. The TLCTS system was shown in an independently-conducted evaluation to improve learning and engagement relative to standard training practice (Surface et al., 2007). Surprisingly, however, soldiers using TLCTS rated the system's game components as less engaging than a traditional intelligent tutoring system bundled with the TLCTS (Surface et al., 2007). Another recent system, the BiLAT (Bilateral Engagement) system, built intelligent tutoring support into a game-based learning environment which teaches soldiers strategies for effective negotiations with people from other cultures (Kim et al., 2009). BiLAT has been shown to lead to significant improvements in the situational judgment skills of individuals inexperienced at cross-cultural negotiations (Kim et al., 2009). An interesting example part-way between an intelligent tutor and educational game is Crystal Island, which explicitly incorporates narrative into a learning system which is otherwise a standard intelligent tutor (McQuiggan et al., 2008); however, while students experienced positive affect within the system (McQuiggan et al., 2010), and learned from the system (McQuiggan et al., 2008), students using Crystal Island did not have significantly better learning than students learning from a PowerPoint presentation of the same content

(McQuiggan et al., 2008). Other recent systems combining intelligent tutoring support with game-like features include Operation ARIES (Wallace et al., 2009) and Mily's World (Rai et al., 2010).

Even intelligent tutors that do not incorporate specific game elements such as competition and narrative often utilize features also found in games, such as instant feedback and measures of continual progress. These features are typically found in games, but their presence alone in a system is not sufficient for a system to be considered as a game. It is possible that the additional motivational features of educational games lead to more positive affect than intelligent tutors (i.e. more delight and engagement, and less frustration and boredom), but it is also possible that the largest motivational benefits come from the interactivity and feedback that both games and intelligent tutors share.

In this study, we compare the affective states exhibited by students using an educational game and students using an intelligent tutor which is not explicitly designed to also be a game (simply called an intelligent tutor in the remainder of the paper, for brevity). The goal of this study is to determine whether students indeed experience significantly better affect in educational games than in an intelligent tutor. We ask: how do students' affect differ in detail between these two types of environments? Relatedly, do students display less disengagement within educational games than in non-game intelligent tutors, manifested through disengaged behaviors such as "gaming the system" (Baker et al., 2004), a form of behavior known to be associated with negative affect (Baker et al., 2010)?

Within this paper, we focus our analyses on students' affect, rather than on their learning of domain content. Both dimensions of the learning experience are of the utmost importance for education in the 21st century. One of the hypothesized key benefits of games, as compared to other forms of educational content, is improved affect, engagement, and enjoyment (cf. Malone & Lepper, 1987; Cordova & Lepper, 1996; Kirriemuer & McFarlane, 2004). While creating the maximum learning within a fixed amount of time may be necessary in formal educational settings such as schools, improved affect may lead students to play an educational game outside of school, increasing total time on task and potentially leading to greater total learning. Positive affect may also lead to increased situational interest (Hidi & Anderson, 1982), which in turn has been theorized to lead to greater long-term personal interest in the content domain (Krapp, 2002). Correspondingly, if one form of educational interface is associated with greater proportions of negative affect, there may be costs beyond just direct and immediate impacts on learning. Boredom has been shown to be correlated with greater incidence of disengaged behavior, in both games and intelligent tutors (Baker et al., 2010). Boredom has also been shown to be correlated with less use of self-regulation and cognitive strategies for learning (Pekrun et al., 2010). Student frustration, similarly, can lead to decreased perception of skills in a domain (cf. Bergin & Reilly, 2005) and eventually lead to disengagement from a learning task (Perkins et al., 1985).

Comparing these two environments provides us with an opportunity to study the impact of games on student disengaged behaviors associated with differences in learning.

In specific, we will consider hint abuse (Aleven & Koedinger, 2000) and systematic guessing, behaviors categorized as gaming the system ("attempting to succeed in an educational environment by exploiting properties of the system rather than by learning the material and trying to use that knowledge to answer correctly" – Baker et al., 2004). Gaming the system has been found within intelligent tutors to lead to worse learning (Baker et al., 2004; Walonoski & Heffernan, 2006; Cocea et al., 2009). Gaming the system behaviors have been observed in both intelligent tutors (cf. Aleven & Koedinger, 2000; Baker et al., 2004) and in educational games (cf. Miller et al., 1999; Magnussen & Misfeldt, 2004; Rodrigo et al., 2007; Baker et al., 2010). Hence, it is likely to be feasible to compare the occurrence of gaming the system in these two types of learning environments.

## 2. Preliminary Study

Prior to the study reported in this paper, our research group published a post-hoc quasi-experimental analysis of the affect experienced by two groups of students in the Philippines (Rodrigo et al., 2008). One group of students used an award-winning simulation problem-solving game, The Incredible Machine: Even More Contraptions (Sierra Online, 2001). The other group of students used an intelligent tutoring system, Aplusix II: Algebra Learning Assistant (Nicaud et al., 2004; Nicaud et al., 2007) (http://aplusix.imag.fr/). Each student's affect was assessed using quantitative field observations as they used the software in classrooms in private schools in the Philippines (significantly more detail on the quantitative field observation method will be given later in this paper).

Within this comparison, the incidence of boredom and frustration was – surprisingly – higher within the game than the intelligent tutoring system. The incidence of positive engagement was higher in the intelligent tutor.

However, it was not clear whether this counter-intuitive result was due to a genuine difference, or one of the many confounds inherent in a post-hoc analysis of studies conducted for other reasons (the two studies were originally conducted separately, in order to study affective dynamics and the relationships between affect and student behavior in each environment – cf. Rodrigo et al., 2007). First, the two software environments involved different domain content, and the intelligent tutor's content was more immediately relevant to the school environment. Second, the populations had some important differences, specifically in terms of student age. Third, the studies did not run for identical periods of time.

Hence, in the current paper, we present an experiment which addresses each of the three limitations in this earlier work. The study presented in this paper has random assignment to conditions (at the classroom level), identical experiences for subjects in each condition except for the treatment variable, and highly similar domain content between the two systems studied.

Conducting this study will enable us to discover whether the counter-intuitive result previously obtained was simply a result of one of the three confounds, or is a replicable

pattern, suggesting that the motivational differences between intelligent tutors and educational games may not be in terms of moment-to-moment differences in affect.

## 3. Descriptions of the Learning Environments Studied

Within this paper, we study two learning environments that cover similar domain content: Aplusix II: Algebra Learning Assistant, and Math Blaster 9-12 (Davidson, 1997). Higher levels of Math Blaster 9-12 cover similar pre-algebra content as can be found in lower levels of Aplusix. The Math Blaster series has been popular, selling copies and remaining commercial viable for over a decade. Mathematics is integrated into the core game mechanic of Math Blaster 9-12, thought to be a key aspect of successful game design (Habgood, 2007). At the same time, Math Blaster has been criticized for having poor connection between its cover story and the domain content (Squire, 2006), although the degree to which these features are necessary for a game to be successful is still under debate (Ainsworth & Habgood, 2009; Habgood, 2007).

Aplusix: Algebra Assistant (Nicaud et al., 2004; Nicaud et al., 2007) (http://aplusix.imag.fr/) is an intelligent tutoring system for pre-algebra and algebra. Topics are grouped into six categories (numerical calculation, expansion and simplification, factorization, solving equations, solving inequations, and solving systems), with four to nine levels of difficulty each. Aplusix presents the student with a mathematics problem from a problem set chosen by the student and allows the student to solve the problem one step at a time, more or less as he or she would using a paper and pen. 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 Figure 1). This informs the student about the state of the problem in order to guide him or her towards the final solution. Students can end the exercise when they believe they are done. 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 has the option of looking at the solution, a "bottom out" hint (cf. Aleven & Koedinger, 2000) with the final answer. Overall, Aplusix both reifies student thinking and gives instant feedback, the two key characteristics of modern intelligent tutoring systems (cf. Anderson et al., 1995). However, Aplusix lacks one game-like feature found in many intelligent tutoring systems – indications of the probability that students have learned relevant skills, in the form of "skill bars". It has been suggested that students view skill bars as being like points in games, and that skill bars give students the perception of progress and encourage competition between students (Schofield, 1995), although, in a lab study, Jackson & Graesser (2007) did not find evidence that progress-indicating skill bars improve motivation.

We asked students to focus on level A1 of Aplusix. At this level, students were asked to perform numerical calculations in which they add, subtract, multiply and divide positive and negative whole numbers, real numbers, and fractions. Problems at this level

of difficulty can range from one type of operation with two operands, e.g. -17 + 10, to two or more types of operations on four to six operands (see Figure 1).

Math Blaster 9-12 is a popular mathematics game from the 1990s, published by Davidson (1997) (Figure 2). Current versions of the "Blaster" series are published by Knowledge Adventure. Math Blaster is a collection of pre-algebra drills embedded in an adventure game. The premise of the game is that a galactic commander is stranded on a planet of monkeys. To help the commander escape, the player has to collect medallions that the commander can then offer to the monkey king. In order to win the medallions, the player has to engage in pre-algebra games that require him or her to add, subtract, multiply or divide positive and negative whole numbers, decimals, or fractions.

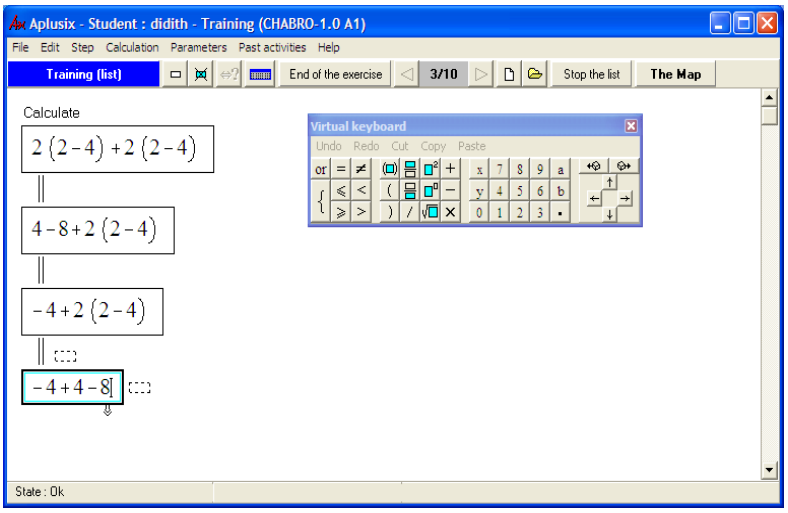


Figure 1. A screen shot from Aplusix: Algebra Learning Assistant.

The participants were asked to focus on three activities within the game: Crater Crossing (Figure 2), Banana Splat (Figure 3) and Bridge Builder (Figure 4). In Figure 2, Crater Crossing, the participants had to jump on pods whose solutions ranged from -14 to -1. In Figure 3, Banana Splat, the participants had to throw a banana at the flying monkey carrying the number that completes the equation, -9 - -5 - ___ = -13. Finally, in Figure 4, Bridge Builder, participants had to complete a bridge by selecting the combination of fractions and decimals that add up to 1. The scope of these exercises matched the scope of level A1 of Aplusix, arithmetic operations with integers, decimals, and fractions, though the exact mathematical problems chosen differed between systems. The students were not asked to play the two other games within Math Blaster 9-12, which involved logic puzzles rather than mathematics.

One key aspect of the design of the study is that the content of the game and the intelligent tutor were very similar. The students had to practice the same mathematical operations with the same types of operands in both environments. Some of the more

difficult problems in Aplusix were composed into multi-step problems (a feature not present in Math Blaster), but initial problems did not have this aspect. Finding a game that is recognizably a game, and a tutor that is recognizably a tutor, which cover the same mathematics content, is not a trivial task. The close match between systems in content and difficulty level help us determine whether differences in interaction style will make an impact on learner affect. Aplusix offers the learner a relatively bland and straightforward intelligent tutor interface while Math Blaster embeds math problems in an animated, interactive game environment with colorful animation, music, and humor. These instances of an intelligent tutor and a game, respectively, were chosen as representative of their classes of educational software (to be more specific, Math Blaster represents games that integrate domain content with the core game mechanic, but that do not integrate domain content into the game fantasy context – cf. Habgood, 2007). Our goal in comparing these two environments was to inform thought about how engaging in these different types of educational interactions will have an impact on learner affect.

However, one potential limitation of the study design is that the mathematics content covered in the two environments could be too easy or too hard for the population studied (discussed later in this paper). While some arithmetic items appeared likely to be too easy, the fraction comparison problems (involving multiple representations of fractions) have been found to be particularly challenging for this age cohort in large-scale research in other countries (e.g. Fennell et al., 2008). Pre-study discussions with the students' teachers suggested that they found the content age-appropriate for the student population, and in line with national education standards.

As one check on this, all students were given a 10-item pre- and post-test with questions taken from Aplusix, covering student ability to successfully solve mathematics problems drawn from this environment. The results of this test are presented later in this paper. However, this test may not be the most appropriate measure of learning in the two conditions in any event. One of the common uses of games is to increase fluency (e.g. ability to perform correctly quickly) (cf. Baltra, 1990; Repenning & Lewis, 2005; Nelson, 2009), a key aspect of early learning in mathematics. It has been argued that fluency, in particular with the skills covered in these two learning environments, can be just as important as correctness for determining the later ability to use early mathematics knowledge to learn later mathematics skills such as algebra (Fennell et al., 2008). Measuring fluency is challenging; fluency is typically measured through instruments such as timed math problems given by computer (cf. Habgood, 2007) or individually timed problem sets (Siegler & Jenkins, 1989); as previously mentioned, we consider a full investigation of the learning and cognitive differences between intelligent tutors and games to be outside of the scope of this paper.

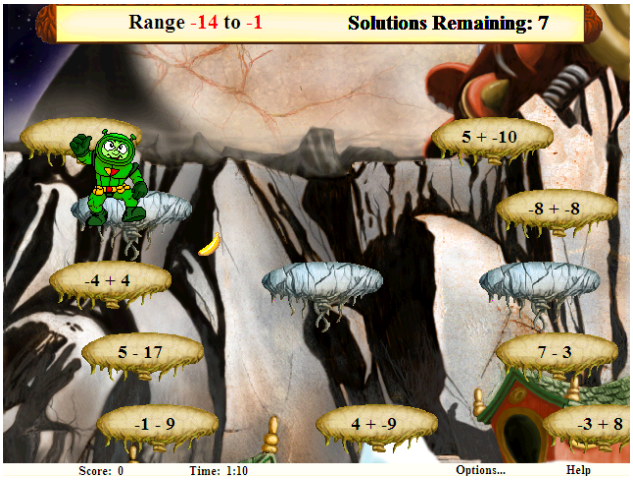


Figure 2. Crater Crossing.

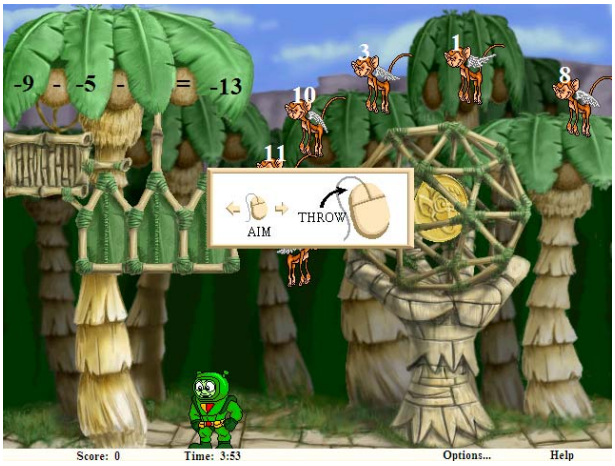


Figure 3. Banana Splat.


Figure 4. Bridge Builder.

## 4. Methods

The study took place in the Ateneo de Manila University Grade School, an all-boys school located in Quezon City, Philippines whose students typically perform well on Philippines National Achievement Test (NAT). In 2009-2010, the NAT scores of the Ateneo's Grade 6 cohort were among the highest in Quezon City. Quezon City, part of the Metro Manila National Capital Region, is the most populous city in the Philippines. Most of the students come from the upper socio-economic bracket of the national population. Because the Ateneo was an all-boys school, no girls were included in the study. While this population limits the generalizability of this study, investigation in this context is valid because single-gender schools are common in the Philippines educational system.

Four sections of grade 7 boys participated in this study. Each section had 40 to 42 students, for a total of 164 participants. The participants had an average age of 12.8 years old and a modal age of 13. In each condition, usage of the software lasted for 40 minutes.

Two sections were randomly selected to use Aplusix while two sections were selected to use Math Blaster. Entire sections were assigned due to the difficulty of separating students from their classmates during class in the schools studied, and the potential effects (such as resentful demoralization) of students in the same computer laboratory receiving highly different software. A between-subjects design was chosen rather than a within-subjects design, due to concerns that students might experience different motivation towards one environment after having used the other (for instance, a student might be displeased about using the intelligent tutor after having used the game, leading to a different pattern of affect). Though these effects can be controlled to some extent through counter-balancing, doing so would lead to greater variance for each condition if there was an order effect or interaction.

To establish that the two groups were homogenous, all students were given a 10-item pre- and post-test with questions taken from Aplusix's Level A1, covering student ability

to successfully solve mathematics problems drawn from this environment. There was not a significant difference between the two groups of students at the time of the pre-test, $t(162)=0.002$, two-tailed $p=0.998$, for a t-test assuming equal variances. There were also no differences between the pre-test and post-test scores for either group – for Aplusix, $t(80)=0.05$, two-tailed $p=0.96$; for MathBlaster, $t(82)=0.96$, two-tailed $p=0.34$. This appears to be because performance on the pre-test was very high (for each environment, an average of 8.8 items were correct on the pre-tests), and was likely at ceiling. This result was somewhat surprising, given our review of relevant literature and discussions with the teachers, but likely represents both the advanced population studied, and the distribution of difficulty in the items studied (including both arithmetic and fraction comparison). As discussed earlier, a test of problem-solving skill may not have been sensitive enough to establish differences between conditions for this population; a more difficult to administer test of fluency may have had greater power to establish differences in learning between conditions, but would have been outside of the scope of the research goals.

None of the students reported prior experience with Math Blaster 9-12 or Aplusix. Prior to the study, each student was asked if they had ever used a computer, and if they regularly played computer games. Every student reported prior computer experience, and 162 of 164 students reported regularly playing computer games. Hence, the population of students assigned to the control condition and experimental condition had similar experience with computers and games, as well as having similar domain knowledge prior to the study.

Affect was assessed using the field observation protocol first articulated in Rodrigo et al. (2007), refined across several studies (e.g. Rodrigo et al., 2008a, 2008b, 2009) and discussed in full detail in Baker et al. (2010). Similar methods have also been recently used by Dragon et al. (2008) and Ziemek (2006). This method draws upon the rich history of classroom observation of student engagement (e.g. Baker et al., 2004; Karweit & Slavin, 1982; Lahaderne, 1968; Lee et al., 1999; Lloyd & Loper, 1986). Field observations of affect carried out in this manner have achieved good inter-rater reliability in prior research (cf. Rodrigo et al., 2007, 2008a, 2008b, 2009; Baker et al., 2010). This approach has significant advantages over video coding, in terms of feasibility and tractability, for assessing student affect in classroom settings. There are several challenges to coding affect in classrooms using video, including the difficulty of getting a full picture of each student's affect with a tractable number of video cameras (unlike in laboratory settings, students often are not sitting right in front of their computers; they go off-task, walk around to help each other, lean to the side, and so on; this requires synchronizing data from multiple cameras per student, a highly time-consuming and challenging process).

In this study, observers coded both affect and behavior during live observation sessions. The observations were carried out by a team of eight observers, working in pairs. The observers were Masters students in Education or Computer Science. Most had teaching experience. During any given live observation session, only two pairs of

observers were available to code the data. The other two pairs were not present because of conflicts in schedule. Each pair of observers was randomly assigned a row of 10 students to observe. Given that we had two pairs of observers per session over four sessions, we could collect data for a maximum of 80 students. In analysis, though, we only considered affect and behavior for 59 students—29 for Aplusix and 30 for Math Blaster. This lower number was due to two factors. First, there was one row that only had 9 students instead of 10. Second, data from one pair of observers had to be discarded because of unacceptably low inter-rater reliability. (We will elaborate on the examination of inter-rater agreement later in the text.) The students were not told who among them was being observed.

The observers trained for the task through a series of pre-observation discussions on the meaning of the affective and behavior categories, oriented around an existing coding manual, and through a pilot observation exercise conducted at a different school. Observations were conducted according to a guide that gave examples of actions, utterances, facial expressions, or body language that would imply an affective state. Observers practiced the coding categories during a pilot observation period prior to the studies. The same set of observers coded student behavior in both systems, in order to make sure that the same constructs were coded in both systems (e.g. that the same student behaviors were coded as "bored" in both conditions).

The guide was based on earlier work by Baker et al. (2004) and D'Mello et al. (2005), and is discussed in detail in Rodrigo et al. (2007) and Baker et al. (2010). The affective categories coded were taken from D'Mello et al. (2005). It is worth noting that two of the affective states studied, engaged concentration and confusion, have significant cognitive components. For this reason, some researchers have termed these two states *cognitive-affective* states (cf. Baker et al., 2010). However, in this paper, for simplicity of discussion, we refer to the full set of states as affective states. Considering them in this fashion does not signify that these states do not have a cognitive component; it simply allows us to focus on the affective aspect of these states. The affective states coded (given along with examples from our coding manual), consisted of:

(1) **Boredom** – slouching, resting the chin on his/her palm; statements such as "Can we do something else?" or "This is boring!"
(2) **Confusion** – 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** – clapping hands; laughing with pleasure; statements such as, "Yes!" or "I got it!"
(4) **Engaged concentration** – immersion, focus, and concentration on the system; leaning towards the computer; mouthing solutions; pointing to parts of screen. Engaged concentration can be considered a subset of the construct of flow proposed by Csikszentmihalyi (1990).
(5) **Frustration** – banging on the keyboard or mouse; pulling his/her hair; deep sighing; statements such as, "What's going on?!"

(6) **Surprise** – sudden jerking or gasping; statement such as "Huh?" or "Oh, no!"
(7) **Neutral** – coded when the student did not appear to be displaying any of the other affective states or when the student's affect could not be determined for certain

Behavior categories were also coded, using the following coding scheme from Baker et al. (2004):
(1) **On-task** – working within the software
(2) **Giving and receiving answers** – asking a teacher or another student for the answer to a problem
(3) **Other on-task conversation** – asking a teacher or another student for help with interface problems or game mechanics
(4) **Off-task conversation** – talking about any other subject
(5) **Off-task solitary behavior** – any behavior that did not involve the software or another individual (such as surfing the web or checking a cell phone)
(6) **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.
(7) **Gaming the System** – sustained and/or systematic guessing, and repeated and rapid help requests used to iterate to a solution, without reflection

Observers were trained to conduct observations in a way that did not make students aware that they were being observed at a given moment. To this end, students were observed through quick glances. Observers used peripheral vision or pretended that they were looking at another student, so as to minimize the effects of the observations. Each observation lasted twenty seconds. If two distinct affective states were seen during an observation, only the first affective state observed was coded; similarly, if two distinct behaviors were seen during an observation, only the first behavior observed was coded. Any behavior by a student other than the student currently being observed was not coded.

Each pair of observers was assigned to ten students. Observers rotated among students in a pre-determined order, and conducted all observations in synchrony, that is, both observers looked at the same student at the same time. Since each observation lasted twenty seconds, each student was observed once every 200 seconds. Students used the software for 40 minutes. Each observer coded 12 observations per student. Each student therefore had a total of 24 observations.

As a training exercise, the observers practiced the observation protocol on a set of pilot students prior to the study. Each pair of observers compared their observations during this training session, to resolve differences in judgments. After the observations of the actual study participants, we discarded the data from one pair of observers because of low inter-rater reliability (for Aplusix, $\kappa$=0.31 for affect, and $\kappa$=0.58 for behavior; for Math Blaster, $\kappa$=0.54 for affect, and $\kappa$=0.26 for behavior). Inter-rater reliability among the remaining three observer pairs was acceptably high for both systems. For Aplusix, Cohen's (1960) $\kappa$=0.58 for affect, and $\kappa$=0.61 for behavior. For Math Blaster, $\kappa$=0.79 for affect, and $\kappa$=0.59 for behavior. Kappa values in this range, considered "substantial agreement" by Landis & Koch (1977), are common in classroom observation research

coding these types of behavior and affect (cf. Chow & Kasari, 1999; Gest & Gest, 2005; Rodrigo et al., 2007, 2008a, 2008b, 2009; Baker et al., 2010).

## 5. Results

We examined the extent to which students exhibited each affective state (Table 1) and each behavior (Table 2). The percentages reported represent the propotion of observations classified as a specific affective state or behavior. In all computations, we included all observations, including cases where there was disagreement.

Table 1. Proportion of observations of each affective state across students; Statistically significant differences ($p<0.05$) in dark gray; marginally significant differences ($p<0.1$) in light gray.

| Affective State | Aplusix | Math Blaster |
|---|---|---|
| Boredom | 12.78% | 21.52% |
| Confusion | 4.45% | 2.36% |
| Delight | 5.89% | 12.08% |
| Engaged concentration | 76.29% | 63.05% |
| Frustration | 0.00% | 0.27% |
| Surprise | 0.00% | 0.00% |
| Neutral | 0. 57% | 0.69% |

Engaged concentration was the most prevalent affective state within both Aplusix (76.29%) and Math Blaster (63.05%) – e.g. students using Aplusix were engaged 76.29% of the time, while students using Math Blaster were engaged 63.05% of the time. There was a significantly greater proportion of engaged concentration in students using Aplusix than Math Blaster, $t(57)=2.83$, two-tailed $p<0.01$. Students exhibited marginally significantly more delight in Math Blaster (12.08%) than Aplusix (5.89%), $t(57)=-1.73$, two-tailed $p=0.08$. Students using Math Blaster also exhibited marginally significantly more boredom (21.52%) than students using Aplusix (12.78%), $t(57)=-1.91$, two-tailed $p=0.06$. Frustration and surprise were not observed in either condition during this study.

Table 2. Proportion of observations of each behavior across students; Statistically significant differences ($p<0.05$) in dark gray.

| Behavior | Aplusix | Math Blaster |
|---|---|---|
| On-task solitary | 77.15% | 87.91% |
| Giving and receiving answers | 15.51% | 6.81% |
| Other on-task conversation | 0.28% | 0.69% |
| Off-task conversation | 3.45% | 2. 50% |
| Solitary off-task | 0. 57% | 1.11% |
| Inactive | 1.43% | 1.11% |
| Gaming | 1.58% | 0.00% |

In terms of behavior, students were more frequently on-task in a solitary fashion in Math Blaster (87.91%) than in Aplusix (77.15%), to a statistically significant degree, $t(57)=-3.34$, two-tailed $p=0.001$. However, the total amount of time spent engaged with the material was comparable in the two systems: instead of working solitarily, students spent significantly more time exchanging answers within Aplusix (15.51%) than Math Blaster (6.81%), $t(57)=3.35$, two-tailed $p=0.001$). Exchanging answers involves some level of engagement with the software, but may not be an ideal problem-solving strategy, as simply learning that the answer is "7" may not lead to understanding *why* the answer is "7". Gaming the system, a strategy related to exchanging answers, was fairly rare within Aplusix (1.58%), but was never observed at all in Math Blaster (0.00%), a statistically significant difference, $t(57)=2.41$, two-tailed $p=0.02$.

Some affective states tended to co-occur with certain behaviors more than others. Table 3 shows that when students using Aplusix were on-task, whether in conversation or solitary, they were typically in a state of engaged concentration (they were engaged 84.54% of the time when on-task solitary; 63.89% when giving and receiving answers; and 50.00% when engaged in other on-task conversation). When students were off-task, they tended to appear to be delighted (41.67% when engaged in off-task conversation; 75.00% when in solitary off-task behavior). Those who were inactive or gaming the system were largely either bored (70.00% while inactive; 45.45% when gaming) or confused (20.00% while inactive; 36.36% when gaming). The tendency for boredom to co-occur with gaming is consistent with findings in Rodrigo et al. (2007) and Baker et al. (2010).

Within Math Blaster, the pattern appeared to be moderately different, as shown in Table 4. Students using Math Blaster mostly exhibited engaged concentration (68.88%) while solitarily on-task. Engaged concentration was less common among students giving and receiving answers (30.61%) – students giving and receiving answers were most frequently observed as being in a state of delight (46.94%). While engaged in other on-task conversation, students tended to be either bored (40.00%) or delighted (40.00%). Delight also co-occurred with off-task conversation 61% of the time. Students who were engaged in solitary off-task behavior tended to be bored (42.86%). Those students who were inactive tended to be either confused or neutral.

Table 3. Proportion of the time each affective state was observed to co-occur with each behavior while students used Aplusix. There were no observations of Frustration and Surprise.

| | Boredom | Confusion | Delight | Engaged concentration | Neutral |
|---|---|---|---|---|---|
| **On-task solitary** | 11.55% | 1.30% | 2.61% | 84.54% | |
| **Giving and receiving answers** | 6.48% | 14.81% | 12.04% | 63.89% | 2.78% |
| **Other on-task conversation** | | | 50.00% | 50.00% | |
| **Off-task conversation** | 29.17% | 8.33% | 41.67% | 20.83% | |
| **Solitary off-task** | 25.00% | | 75.00% | | |
| **Inactive** | 70.00% | 20.00% | | | 10.00% |
| **Gaming** | 45.45% | 36.36% | | 18.18% | |

Table 4. Proportion of the time each affective state was observed to co-occur with each behavior while students used Math Blaster. There were no observations of Surprise.

| | Boredom | Confusion | Delight | Engaged concentration | Frustration | Neutral |
|---|---|---|---|---|---|---|
| **On-task solitary** | 22.27% | 0.79% | 7.74% | 68.88% | 0.32% | |
| **Giving and receiving answers** | 12.24% | 10.20% | 46.94% | 30.61% | | |
| **Other on-task conversation** | 40.00% | | 40.00% | 20.00% | | |
| **Off-task conversation** | 16.67% | 5.56% | 61.11% | 11.11% | | 5. 56% |
| **Solitary off-task** | 42.86% | 28.57% | 28.57% | | | |
| **Inactive** | | 50.00% | | | | 50.00% |
| **Gaming** | | | | | | |

## 6. Discussion and Conclusions

The results here suggest a new picture of the affective and behavioral differences between students playing a game, represented by Math Blaster, and using an intelligent tutor, represented by Aplusix. Based on the general perception of games, as well as past theoretical accounts about their benefits, it might have been reasonable to hypothesize that students would be on-task more often in the game, and would experience more engaged concentration and delight within the game.

The pattern uncovered in this study is different. Though there was a trend towards more delight in Math Blaster, significantly more engaged concentration was observed within Aplusix. Students were on-task about the same amount of time in the two environments, but they spend more time on giving and receiving answers in Aplusix, and more time on working solitarily in Math Blaster. Finally, students gamed the system more often in Aplusix (gaming the system was fairly rare in Aplusix, but even small amounts of gaming the system have been shown to be associated with significantly poorer learning – e.g. Baker et al., 2004).

It appears that both environments successfully engage students, but in different ways. Students experienced more continuing engagement in Aplusix. However, Math Blaster appears to lead to more experience of delight. The difference in delight is unsurprising. Throwing bananas at monkeys, leading monkeys across a bridge, and jumping from one floating pod to another are likely to be more delightful than solving expressions on a featureless screen. One key question for future research will be which form of engagement is more important for learning – there is considerable research linking engaged concentration (e.g. flow) to positive long-term outcomes for both learning and desire for learning (e.g. Csikszentmihalyi, 1990). The research on delight and learning is more limited, in part because delight is relatively rare in non-game learning environments (cf. Baker et al., 2010). However, further study of affect in educational games may enable the field to better understand the impacts on learning of these different forms of positive affect. In particular, discovering and ablating features that produce delight may help to clarify how delight impacts learning.

At the same time as it promoted delight, students using Math Blaster experienced more boredom than students using Aplusix, suggesting that the delight-generating features may not have created engagement that persisted, unlike the more concentrated engagement produced by the less delightful interaction in Aplusix. This finding is somewhat non-intuitive, and suggests that the factors driving delight and boredom may be more orthogonal than previously suggested by models that treat these two affective states as opposites – an example of this is seen in the mapping between educationally relevant affective states and Russell's (2003) core framework, in Baker et al. (2010). Alternatively, the boredom seen in Math Blaster may be due to students' high initial proficiency with the material, although this does not directly explain why there was greater boredom in Math Blaster than Aplusix, given the similar content in the two systems.

Correspondingly, what led to the high engaged concentration in Aplusix? One possibility is that the high interactivity and challenge within Aplusix led to this result. Like most modern intelligent tutoring systems, Aplusix gives continual feedback to students, and students are given the choice of continually more difficult mathematics problems. Continual feedback and challenge are hallmarks of successful games (cf. Malone & Lepper, 1987; Gee, 2003). This is not to say that Aplusix is a game, as it lacks key aspects that most games have, in particular fantasy context and storyline characters, sounds and animation. However, the elements that Aplusix shares in common with highly successful games may be the factors that lead to engaged concentration, whereas the factors that it lacks — animated characters, sounds, story lines — may be those that lead to delight in games. In general, understanding the affective patterns associated with each aspect of games and other highly interactive learning environments may help us design interactive learning environments that students respond with high engagement. Another possibility is that presentation of the mathematics was more engaging in Aplusix than Math Blaster; although both required the same basic mathematical operators, the composition of problems in Aplusix (where the answer of one operation is frequently an

operator in the next operation) may have been more engaging than the more disconnected mathematics problems in Math Blaster, possibly leading to more engaged concentration and less boredom.

Another interesting difference between the two environments was in terms of behavior. Students using Aplusix showed more collaborative behavior, but had a tendency to focus on answers to the expense of the mathematics. This tendency showed up both in the students' collaborative behaviors, and through the incidence of gaming the system. This difference suggests that some element of games may lead students to attempt to master the game on its own terms, succeeding within the game's rules, rather than focusing on getting the correct answers in any way possible. In other words, games may lead students to display mastery goals instead of performance goals (cf. Dweck, 2000). If this finding replicates, it will be an essential difference between games and intelligent tutors, and discovering what aspects of design lead to this difference will be a key area for future research.

An alternate potential explanation for the greater degree of collaborative behavior in Aplusix is that problems were somewhat longer in later Aplusix problems than in Math Blaster problems (even though the content was quite similar), with the greater length caused by the exact same mathematical operations being composed into more complex expressions. These longer problems may have afforded students more opportunity to discuss the problems. At one level, this may appear a confound between the two environments; it can also be seen as an affordance of intelligent tutors. It is relatively difficult to embed lengthy problems in games and maintain a "game-like" feel, a challenge not present in intelligent tutors.

One limitation of the research presented here is that only one intelligent tutor and only one educational game were studied. In order to be considered a general property of games and intelligent tutors, as opposed to a contingent property of the design of the two environments studied in this research, the findings obtained here will need to be replicated across other pairs of tutors and games. Similarly, replicating the results across additional populations will be needed to establish that the results seen here can generalize across age groups, national cultures, and different learning settings. However, even within one pair of systems and population, the results obtained within this paper indicate that common hypotheses of how games and tutors impact student affect (e.g. Gee, 2003; Prensky, 2007) do not hold, at least in some cases. A second limitation of the study presented here is that the game and intelligent tutor differed among multiple dimensions, though the domain content was highly similar. This is a general difficulty for conducting research on games versus other types of learning environments. Games and intelligent tutoring systems are different genres (except when hybridized), and therefore it is hard to find or develop a paired comparison that differs on only one dimension but is still unambiguously representative of each type of learning system (e.g. an intelligent tutoring system with fantasy is still an intelligent tutoring system; a game with on-demand help is still a game). Even the comparison between the game-tutor hybrid "Mission Game", and the pure ITS "Skill Builder" components of the Tactical Language and Culture Training

System in (Surface et al., 2007) still differed along multiple dimensions. A potential line of future work towards determining which features of games and intelligent tutors influence student affect in different ways is to develop hybrids of games and tutors, and study their differences from more central examples of each genre. In doing so, it will be important to be able to turn on and off specific elements (such as fantasy, competition with computer players, and trivial choice) to test each element's individual effects on student affect. This program of research is similar to the approach in Cordova & Lepper (1996) where a mathematics game with combinations of these elements was compared to a game lacking all of them. Starting with an intelligent tutor as the baseline has potential advantages over existing research: an intelligent tutor can generally be expected to start with positive affect (as shown here, and also as reported in Schofield, 1995) and positive learning gains (so long as the tutor has been previously validated), bringing into sharper relief what is uniquely positive about game features.

In recent years, there has been rapidly increasing interest in educational computer games, based on the hypothesis that embedding computer games into education can be a way to improve students' affect, interest, and motivation towards education, and in turn improve their learning (e.g. Squire, 2003; Shaffer et al., 2005; Wideman et al., 2007). However, in the research reported here, we have found that a traditional intelligent tutoring system can produce comparable affect to a commercially successful educational game covering very similar domain content. The game achieves a greater proportion of delight, but the tutor achieves a greater proportion of engaged concentration. The key question, therefore, appears not to be which type of learning environment is better, but how we can leverage the best practices developed by each of these design communities, in order to develop a new generation of optimally engaging and educationally effective learning environments.

## Acknowledgments

We thank Jean-Francois Nicaud of the Laboratoire d'Informatique de Grenoble for the use of Aplusix. We also thank the Ateneo de Manila High School, Kostka School of Quezon City, School of the Holy Spirit of Quezon City, St. Alphonsus Liguori Integrated School and St. Paul's College Pasig for their participation in the studies conducted. We thank our colleagues, research assistants, and students at the Education Department and the Department of Information Systems and Computer Science at the Ateneo de Manila University. We thank the Ateneo de Manila Grade School, Fr. Norberto Maria L. Bautista, S.J., Mrs. Helen Amante, Mr. Marvin Marbella, and Mr. Paolo Sanchez for allowing us to conduct this study with their students. We thank Juan Miguel Andres for proofreading this paper before final submission. This publication was made possible through a 2008-2009 Advanced Research and University Lecturing Fulbright Scholarship provided by the Philippine American Educational Foundation and the Council for International Exchange of Scholars and through grants from the Ateneo de Manila, and the Philippines Department of Science and Technology's Engineering Research and Technology for Development Project entitled Multidimensional Analysis of User-Machine Interactions

Towards the Development of Models of Affect, and by the Pittsburgh Science of Learning Center which is funded by the National Science Foundation, award number SBE-0836012.

**References**

Ainsworth, S., & Habgood, J. (2009). Exploring the effectiveness of intrinsic integration in serious games. *Proc. 13th biennial conference of the European Association for Research on Learning and Instruction (EARLI)*. Amsterdam, The Netherlands.

Alessi, S. M. & Trollip, S. R. (2001). *Multimedia for learning: Methods and Development (3rd Ed)*. Needham Heights, MA: Allyn & Bacon.

Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson & K. VanLehn (Eds.), *Proc. 5th international conference on intelligent tutoring systems* (pp. 292-303), Montreal, Canada.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2) 167-207.

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". *Proc. ACM CHI 2004: Computer-Human Interaction* (pp. 383-390).

Baker, R. S. J. d., D'Mello, S., Rodrigo, M. M. T., Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223-241.

Baltra, A. (1990). Language learning through video adventure games. *Simulation & Gaming, 4*, 445–452.

Bergin, S., & Reilly, R. (2005). Programming: factors that influence success. In *Proceedings of the 36th SIGCSE Technical Symposium on Computer Science Education* (St. Louis, Missouri, USA, February 23 - 27, 2005). SIGCSE '05. ACM, New York, NY, 411-415.

Bragg, L. (2007). Students' conflicting attitudes towards games as a vehicle for learning mathematics: A methodological dilemma. *Mathematics Education Research Journal, 19*(1), 29-44.

Brown, J. S. (2005). New learning environments for the 21st century. *The Magazine of Higher Learning, 38*(5), 18-24.

Burton, R., & Brown, J. S. (1982). An investigation of computer coaching for informal learning activities. In D. Sleeman & J. S. Brown (Eds.), *Intelligent Tutoring Systems*, Orlando, FL: Academic Press.

Bruckman, A. (1999). Can education be fun? Paper Presented at the Game Developer's Conference, San Jose, CA.

Chow, V.T., & Kasari, C. (1999). Task-related interactions among teachers and exceptional, at-risk, and typical learners in inclusive classrooms. *Remedial and Special Education, 20*(4), 226-232.

Cocea, M., Hershkovitz, A., & Baker, R. S. J. d. (2009). The impact of off-task and gaming behaviors on learning: Immediate or aggregate? *Proc. of the 14th international conference on artificial intelligence in education* (pp. 507-514), Brighton, UK.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*, 715-730.

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*. New York: Harper and Row.

Davidson and Associates. (1997). Math Blaster 9–12 [Computer Software]. Torrance, CA: Davidson.

D'Mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., & Graesser, A. (2005). Integrating affect sensors in an intelligent tutoring system. In *Affective Interactions: The Computer in the Affective Loop.* Workshop conducted at the International Conference on Intelligent User Interfaces, San Diego, CA.

Dragon, T., Arroyo, I., Woolf, B.P., Burleson, W., el Kaliouby, R., & Eydgahi, H. (2008). Viewing student affect and learning through classroom observation and physical sensors. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 29-39.

Dweck, C. S. (2000). *Self-Theories: Their role in motivation, personality and development*. Hove, UK: Psychology Press.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & T. Power (Eds.), *The handbook of cognition and emotion* (pp. 45-60). Sussex, UK: John Wiley & Sons, Ltd.

Fennell, F., Faulkner, L. R., Ma, L., Schmid, W., Stotsky, S., Wu, H-H., & Flawn, T. (2008). *Foundations for success: Report of the National Mathematics Advisory Panel: Chapter 3: Report of the Task Group on Conceptual Knowledge and Skills.* Washington, DC: US Department of Education.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Gest, S. D., & Gest, J. M. (2005). Reading tutoring for students at academic and behvioral risk: Effects on time-on-task in the classroom. *Education and Treatment of Children, 28*, 25-47.

Goldstein, I. J. (1979). The genetic graph: A representation for the evolution of procedural knowledge. *International Journal of Man-Machine Studies, 11*(1), 51-77.

Habgood, M. P. J. (2007). *The effective integration of digital games and learning content.* Unpublished doctoral thesis, University of Nottingham, UK.

Hidi, S., & Anderson, V. (1992). Situational interest and its impact on reading and expository writing. In K. A. Renninger, S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development* (pp. 215-238). Hillsdale, NJ: Erlbaum.

Jackson, G. T., & Graesser, A. C. (2007). Content matters: An investigation of feedback categories within an ITS. In R. Luckin, K. Koedinger & J. Greer (Eds.), *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work*. Amsterdam, The Netherlands: IOS Press.

Kafai, Y. B. (2001). The educational potential of electronic games: From games-to-teach to games-to-learn. Retrieved October 15, 2009, from http://www.savie.ca/SAGE/Articles/1182_1232-KAFAi-2001.pdf

Karweit, N., Slavin, R.E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal of Experimental Psychology, 74*(6), 844-851.

Kerawalla, L., & Crook, C. (2005). From promises to practices: The fate of educational software in the home. *Technology, Pedagogy and Education, 14*(1), 107-125.

Kim, J., Hill, R. W., Durlach, P., Lane, H. C., Forbell, E., Core, M., Marsella, S., Pynadath, D., & Hart, J. (2009). BiLAT: A game-based environment for [racticing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education, 19*(3), 289-308.

Kirriemuir, J., & McFarlane, A. (2004). *Literature review in games and learning: A report for NESTA Futurelab*. Retrieved March 12, 2010, from http://hal.archives-ouvertes.fr/docs/00/19/04/53/PDF/kirriemuir-j-2004-r8.pdf

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 61-78). Cambridge, UK: Cambridge University Press.

Krapp, A. (2002). Structural and dynamic aspects of interest development: theoretical considerations from an ontogenetic perspective. *Learning and Instruction, 12* (4), 383-409.

Lahaderne, H. M. (1968). Attitudinal and intellectual correlates of attention: A study of four sixth-grade classrooms. *Journal of Educational Psychology, 59*(5) , 320-324.

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Lee, J, Luchini, K., Michael, B., Norris, C. & Solloway, E. (2004). More than just fun and games: Assessing the value of educational video games in the classroom. *Proc. ACM SIGCHI 2004*, 1375-1378.

Lee, S. W., Kelly, K. E., & Nyre, J. E. (1999). Preliminary report on the relation of students' on-task behavior with completion of school work. *Psychological Reports, 84*, 267-272.

Lloyd, J. W., & Loper, A. B. (1986). Measurement and evaluation of task-related learning behavior: Attention to task and metacognition. *School Psychology Review, 15*(3), 336-345.

Magnussen, R. & Misfeldt, M. (2004). Player transformation of educational multiplayer games. Retrieved October 15, 2009, from https://pure.dpu.dk/ws/fbspretrieve/82/player_transformation.pdf

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction: III, Conative and affective process analysis.* Hillsdale, NJ: Erlbaum.

McQuiggan, S. W., Robison, J. L., & Lester, J. C. (2010). Affective transitions in narrative-centered learning environments. *Educational Technology & Society, 13*(1), 40-53.

McQuiggan, S. W., Rowe, J. P., Lee, S., & Lester, J. C. (2008). Story-based learning: The impact of narrative on learning experiences and outcomes. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 530-539.

Miller, C., Lehman, J., & Koedinger, K. (1999). Goals and learning in microworlds. *Cognitive Science, 23*, 305-336.

Nelson, M. (2009). *The effects of computer math games to increase student accurracy and fluency in basic multiplication facts.* Unpublished doctoral dissertation, Caldwell College, Caldwell, NJ, USA.

Nicaud, J. F., Bouhineau, D., & Chaachoua, H. (2004). Mixing microworld and CAS features in building computer systems that help student learn algebra. *International Journal of Computers for Mathematical Learning, 9*(2), 169-211.

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

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal, 16*(4), 455-474.

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology, 102*(3), 531–549.

Perkins, D. N., Hancock, C., Hobbs, R., Martin, F., & Simmons, R. (1985). Conditions of learning in novice programmers. Concept Paper. Educational Technology Center, Harvard Graduate School of Education.

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

Prensky, M. (2007). *Digital game-based learning*. St. Paul, MN: Paragon House.

Rai, D., Beck, J. E., Heffernan, N. T. (2010). Mily's world: A coordinate geometry learning environment with game-like properties. *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*, 254-256.

Repenning, A., & Lewis, C. (2005). Playing a game: The ecology of designing, building, and testing games as educational activities. *Proceedings of ED-MEDIA: World Conference on Educational Multimedia, Hypermedia, and Telecommunications.*

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

Rodrigo, M. M. T., Baker, R. S. J. d., D'Mello, S., Gonzalez, M. C. T., Lagud, M. C. V., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., Santillano, J. Q., Sugay, J. O., Tep, S., & Viehland, N. J. B., (2008a). Comparing learners' affect while using an intelligent tutoring system and a simulation problem solving game. In B. P. Woolf, E. Aimeur, R. Nkambou & S. P. Lajoie (Eds.), *Proc. Intelligent tutoring systems* (pp. 40-49). Montreal, Canada.

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. In R. Luckin, K. R. Koedinger & J. Greer (Eds.), *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work*, Amsterdam, The Netherlands: IOS Press.

Rodrigo, M. M. T., Rebolledo-Mendez, G., Baker, R. S. J. d., du Boulay, B., Sugay, J. O., Lim, S. A. L., Espejo-Lahoz, M. B., & Luckin, R. (2008b). The effects of motivational modeling on affect in an intelligent tutoring system. *Proceedings of International Conference on Computers in Education*, 57-64.

Russell, J. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145-172.

Shaffer, D. W., Squire, K. D., Halverson, R. & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan, 87*(2), 104-111.

Sherry, J. L. (2004). Flow and media enjoyment. *Communication Theory, 14*, 328-347.

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

Schofield, J. W. (1995) *Computers and classroom culture*. Cambridge, UK: Cambridge University Press.

Siegler, R.S., & Jenkins, E.A. (1989). *How children discover new strategies.* London, UK: Psychology Press.

Squire, K. (2003). Video games in education. *International Journal of Intelligent Simulations and Gaming*, (2) 1.

Squire, K. (2006). From content to context: Videogames as designed experiences. *Educational Researcher, 35*(8), 19-29.

Surface, E. A., Dierdorff, E. C., & Watson, A. (2007). *Special Operations Language Training Software Measurement of Effectiveness Study: Tactical Iraqi Study Final Report.* Tampa, FL: Special Operations Forces Language Office.

Voiskounsky, A.E., Mitina, O. V., & Avetisova, A. A. (2004). Playing online games: Flow experience. *PsychNology Journal, 2*(3), 259-281

Vogel, J. J., Greenwood-Ericksen, A., Cannon-Bowers, J., & Bowers, C. A. (2006). Using virtual reality with and without gaming attributes for academic achievement. *Journal of Research on Technology in Education, 39*(1), 105-118.

Wallace, P., Graesser, A., Millis, K., Halpern, D., Cai, Z., Britt, M.A., Magliano, J., & Wiemer, K. (2009). Operation ARIES! A computerized game for teaching scientific inquiry. *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling,* 602-604.

Walonoski, J., & Heffernan, N.T. (2006). Detection and analysis of off-task gaming behavior in intelligent tutoring systems. Ikeda, Ashley & Chan (Eds.), *Proc. Eight International Conference on Intelligent Tutoring Systems*.

Wenger, E. (1987). *Artificial intelligence and tutoring systems: Computational and cognitive approaches to the communication of knowledge*. Los Altos, CA: Morgan Kaufmann.

Wideman, H. H., Owston, R. D., Brown, C., Kushniruk, F. H., & Pitts, K. C. (2007). Unpacking the potential of educational gaming: A new tool for gaming research. *Simulation and Gaming, 38*(1), 10–30.

Ziemek, T. R. (2006). Two-d or not two-d: Gender implications of visual cognition in electronic games. *Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games,* 183-190.