Comparing Learners' Affect While Using an Intelligent Tutoring System and a Simulation Problem Solving Game

Ma. Mercedes T. Rodrigo[1], Ryan S.J.d. Baker[2], Sidney D'Mello[3], Ma. Celeste T. Gonzalez[4], Maria C.V. Lagud[4], Sheryl A.L. Lim[1], Alexis F. Macapanpan[4], Sheila A.M.S. Pascua[4], Jerry Q. Santillano[1], Jessica O. Sugay[1,4], Sinath Tep[1], Norma J.B. Viehland[4]

[1] *Department of Information Systems and Computer Science, Ateneo de Manila University*
[2] *Human-Computer Interaction Institute, Carnegie Mellon University*
[3] *Institute for Intelligent Systems, University of Memphis*
[4] *Education Department, Ateneo de Manila University*
mrodrigo@ateneo.edu, rsbaker@cs.cmu.edu

Abstract. We compare the affect associated with an intelligent tutoring environment, Aplusix, and a simulations problem solving game, The Incredible Machine, to determine whether students experience significantly better affect in an educational game than in an ITS. We find that affect was, on the whole, better in Aplusix than it was in The Incredible Machine. Students experienced significantly less boredom and frustration and more flow while using Aplusix. This implies that, while aspects unique to games (e.g. fantasy and competition) may make games more fun, the interactivity and challenge common to both games and ITSs may play a larger role in making both types of systems affectively positive learning environments.

Keywords: Affect, intelligent tutoring system, game, Aplusix, The Incredible Machine

1 Introduction

Games are fun. The same adolescents who are often reluctant to put significant time into their studies are often enthusiastically willing to put dozens of hours into playing modern computer games [6]. In recent years, researchers have 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. Some educational games have successfully built upon competition, curiosity, challenge, and fantasy to make learning more enjoyable, increase students' desire to learn, and complete more difficult work than with traditional educational materials [1,8,16]. However, there is also evidence that educational games may not have entirely positive effects on learners' affect and motivation. Bragg [5] found that students exhibited negative attitudes towards the use of games as the main instructional method for learning mathematics. Similarly, Vogel [24] argues that games and simulations that fail to make seamless connections between the subject matter and the game play will also inhibit learners' engagement and motivation.

While it is commonly believed that educational games will lead to better affect than non-gamelike learning environments, the evidence supporting this belief is not yet conclusive. In many cases, educational games have been studied in relation to relatively weak comparison conditions, such as paper worksheets with no feedback [16] and games with game features ablated [8]. Furthermore, it has been found that intelligent tutoring systems lead to significantly improved affect and motivation as compared to traditional, non-computerized learning contexts [22], though not necessarily to expert human tutors. Intelligent tutors generally lack game-like features like competition and fantasy, but share in common with games features such as instant feedback, and measures of continual progress. 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 that both games and intelligent tutors share.

The differences among effects of educational games and intelligent tutoring systems on students' usage choices are also not yet fully studied. Consider hint abuse and systematic guessing, behaviors categorized as gaming the system, i.e. "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" [3]. As Rodrigo et al [20] discussed, students generally know that gaming behavior is undesirable in intelligent tutoring systems, as the primary goal is to learn the domain content – and students demonstrate this belief by hiding this behavior from their teachers. By contrast, there may be a perception that since games are primarily for fun, it is acceptable to use them in any fashion; hence, students may game the system more often in educational games than in intelligent tutoring systems.

In this paper, we compare the affect associated with an intelligent tutoring environment, Aplusix II: Algebra Learning Assistant [17,18] (http://aplusix.imag.fr/), and a simulation problem solving game, The Incredible Machine: Even More Contraptions [20]. Earlier studies [4,20] collected affect and usage data on students' affective states when using The Incredible Machine. We collect similar data for Aplusix, within similar populations and following virtually identical data collection and analysis procedures. By comparing these two data sets, we can determine whether students experience significantly better affect in an educational game than in an intelligent tutor, and in turn study which aspects of educational games explain their positive effects on student affect.

1.1 Descriptions of the Learning Environments

As mentioned in the introduction, affect and usage data were gathered from participants using two different interactive learning environments: the Incredible Machine and Aplusix.

The Incredible Machine [21], called TIM for short, is a simulation game where students complete a series of logical "Rube Goldberg" puzzles. In each puzzle, the student is given a limited set of objects, including mechanical tools like gears, pulleys, and scissors; more active objects like electrical generators and vacuums; and even animals. The student must combine these objects in a creative fashion to accomplish each puzzle's goal. Objectives 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 where items should be located in a correct solution to the current problem (without displaying which items should be placed in each location). TIM is thought to be highly entertaining, having won multiple awards for its innovative gameplay, most recently including an award for best casual mobile game at the 5th Annual Spike TV Video Game Awards in 2007. Hence, if an intelligent tutor can produce comparable levels of affect as those produced by this game, the ITS can be considered highly motivating. A screenshot from TIM is shown in Figure 1.

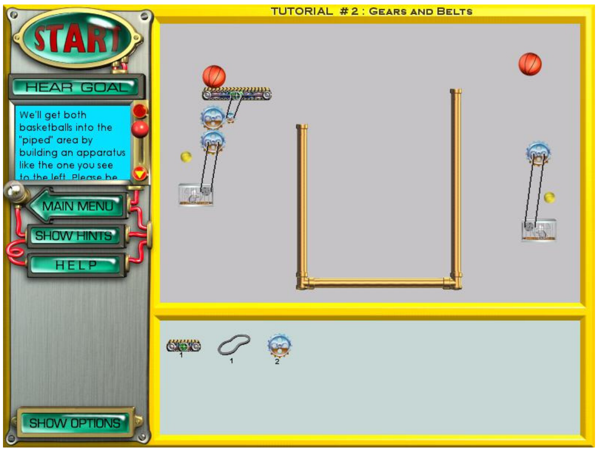

Figure 1. A screen shot from The Incredible Machine: Even More Contraptions (TIM).

Aplusix [18,19] (http://aplusix.imag.fr/) is an intelligent tutoring system for mathematics. 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 an arithmetic or algebraic problem from a problem set chosen by the student and allows the student to solve the problem one step at a time, 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 2). 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 with the final answer. Hence, Aplusix both reifies student thinking and gives instant feedback, two key characteristics of modern intelligent tutoring systems [cf. 2]. 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 [22], although, in a lab study, Jackson & Graesser [14] did not find evidence that progress-only skill bars improve motivation.

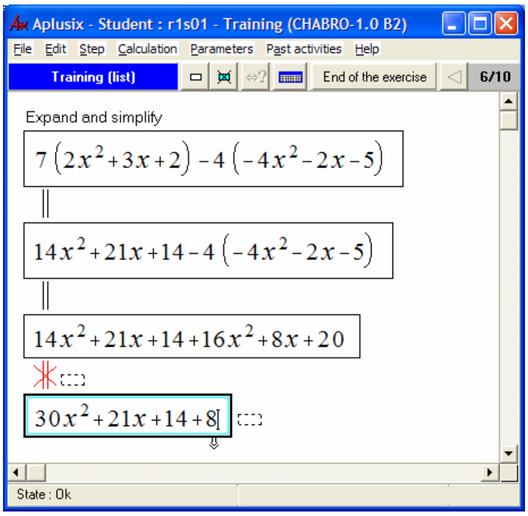

Figure 2. A screen shot from Aplusix: Algebra Learning Assistant

These two systems provide a strong comparison between intelligent tutoring systems and games. TIM has won awards for its enjoyable gameplay; Aplusix can be considered a fairly traditional intelligent tutoring system, as it includes the continual feedback and reification of student thinking that is characteristic of most intelligent tutoring systems, but lacks skill bars, which some researchers think lend intelligent tutors a game-like feel. Hence, the two systems are good representatives of their respective classes, and similarities or differences in learner affect between the two systems will be representative of similarities or differences in affect between games and ITSs in general. It is worth noting that the two systems do not cover the same educational material, as TIM covers general problem-solving skill while Aplusix covers algebra; this possible confound will be considered in the discussion section.

2 Methods

The data gathering procedures for the two environments was very similar. The subsequent section discusses the profile of the participants, the observers, the coding procedures, and the inter-rater reliability of the observations.

The participants for the TIM study were students in a private high school in Quezon City (Metro Manila), the Philippines. Student ages ranged from 14 to 19, with an average and modal age of 16. Thirty-six students participated in this study (17 female, 19 male). 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 and a modal age of 14 (high school begins earlier in the Philippines than in many other industrialized nations). One hundred and forty students participated in the Aplusix study (83 female, 57 male). The participants in both studies were computer-literate. However, none of them had previously used either TIM or Aplusix. The sample of participants did not overlap between studies.

Each student used TIM for ten minutes, and each student used Aplusix for 45 minutes. The different time spent in each system is a potential confound. In specific, this difference might lead to greater boredom or frustration within the Aplusix system (because students may experience more boredom or frustration later in a learning session) – if either of these effects is found, it may be due to differences between the studies rather than differences between the systems. Students used the software in small groups (9 for The Incredible Machine, 10 for Aplusix), one student per computer, during their class time. Each student's affect was observed several times as he or she used the learning software.

The observations were carried out by a team of six observers, working in pairs. The observers were Masters students in Education or Computer Science, and all but one had prior teaching experience. The set of observers was overlapping but not identical between systems. TIM was studied in 2006 [20] Aplusix was studied in 2007. Each observation lasted twenty seconds, and was conducted using peripheral vision, i.e. observers stood diagonally behind or in front of the student being observed and avoided looking at the student directly [cf. 3], in order to make it less clear when an observation was occurring. 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 a small number of students and alternated between them – more observers participated in the TIM study than the Aplusix study, thus a greater amount of time passed between observations in Aplusix (180 seconds) than The Incredible Machine (40 seconds).

In the studies, both affect and behavior were coded. The observers trained for the task through a series of pre-observation discussions on the meaning of the affective and behavior categories. Observations were conducted according to a guide that gave examples of actions, utterances, facial expressions, or body language that would imply an affective state, and observers practiced the coding categories during a pilot observation period prior to the studies. The guide was based on earlier work by [3,11], and is discussed in detail in [20]. The affective categories coded were boredom, confusion, delight, surprise, frustration, flow, and neutral, in line with earlier research by D'Mello et al [11] suggesting that these states are most relevant to students' affective experiences within an Intelligent Tutoring System. "Flow" refers to full immersion in an activity; the participant is focused on a task to the point that he or she is unaware of the passage of time [cf. 10]. The behavior categories coded were on-task, on-task conversation, off-task conversation, off-task solitary behavior, inactivity, and gaming the system; in both systems, gaming behavior consisted of systematic guessing – such as trying an object in every possible place in TIM – and use of help features to arrive at a solution without engaging in problem-solving.

706 observations were collected in TIM, for an average of 19.6 observations per student. Inter-rater reliability was acceptably high across all observations — Cohen's [7] κ=0.71 for usage observations, κ=0.63 for observations of affective state. Thirteen pairs of observations were collected per student in Aplusix, totaling 3,640 observations in all. Inter-rater reliability was again acceptably high: Cohen's κ=0.78 for usage observations, κ=0.63 for observations of affective state.

3 Results

3.1 Prevalence of Affective States

The most common affective state in both Aplusix and TIM was flow, occurring 62% of the time in TIM and 68% of the time in Aplusix. The difference in the prevalence of flow between environments was marginally statistically significant, t(174)= -1.66, two-tailed p=0.10, for a two-tailed, two-sample t-test with pooled variance.

The second most common affective state in both environments was confusion, occurring 11% of the time in TIM and 13% of the time in Aplusix. The difference in the prevalence of confusion between environments was also not statistically significant, t(174)= 0.73, two-tailed p=0.46. Delight was also not significantly different between environments, t(174) = 0.55, two-tailed p=0.58

However, the frequency of two negative affective states was significantly different between systems. Frustration was more common in TIM (6%) than Aplusix (2%), t(174)=3.25, two-tailed p=0.001. Boredom was also more common in TIM (7%) than Aplusix (3%), t(174)=2.27, two-tailed p=0.02.

The overall pattern of results (shown in Figure 3) is that the affective experiences were, on the whole, more positive within Aplusix than TIM, with the effect more pronounced among negative affective states than positive affective states.

3.2 Prevalence of Negative Usage Behaviors

Gaming the system occurred in both Aplusix and The Incredible Machine. The average student gamed the system 1.4% of the time in Aplusix, about half of the prevalence in previous observations of gaming behavior in Cognitive Tutors [cf. 3]; the average student gamed the system 7.5% of the time in The Incredible Machine, about double the prevalence in previous observations of gaming behavior in Cognitive Tutors. The difference between the prevalence of gaming in the two environments was statistically significant, *t*(174)=4.72, p<0.0001, for a two-tailed two-sample t-test with pooled variance.

There was the appearance of a difference in the prevalence of off-task behavior between the two environments, with students being off-task 2.2% of the time

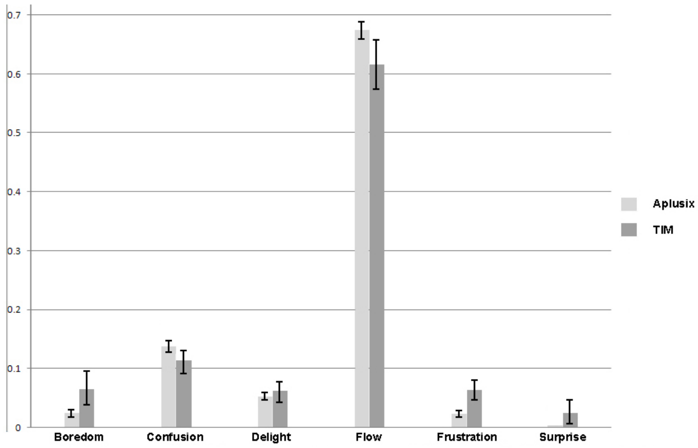

Figure 3. Affective categories' prevalence of occurrence (standard error bars shown)

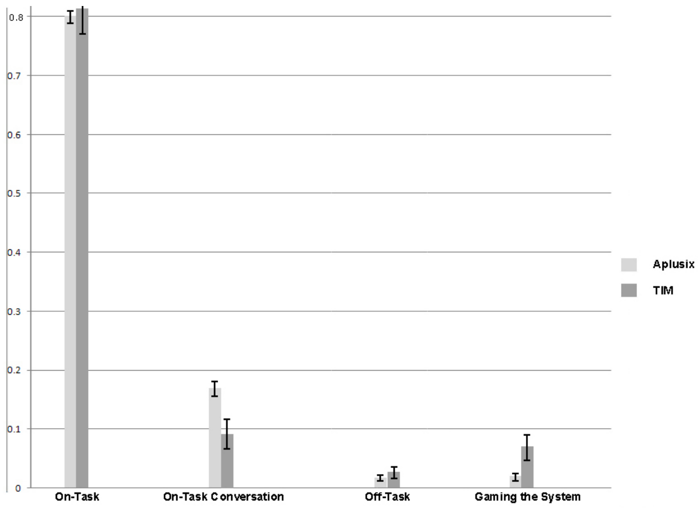

Figure 4. Usage categories' prevalence of occurrence (standard error bars shown)

in TIM and 1.3% of the time in Aplusix, but this difference was not statistically significant, t(174)= 1.25, two-tailed p=0.21.

The time spent on-task, working with the system, within the two environments, was almost identical: 80.9% on-task in TIM, 79.9% on-task in Aplusix, t(174)=0.33, two-tailed p=0.74. However, the time spent on-task, talking to another student or the teacher, was significantly higher in Aplusix (17.3%) than TIM (9.4%), t(174)= -3.14, two-tailed p=0.001. Hence, the overall pattern of results (shown in Figure 4) is that students spent significantly more time gaming the system in TIM, and significantly more time in on-task conversation in Aplusix.

4 Discussion and Conclusions

In this paper we have asked: are educational games associated with better affect because they are games, or simply because they are highly interactive learning environments? We investigated that question by comparing the incidence of positive and negative affective states and usage behaviors in an intelligent tutoring system, Aplusix, and a simulation problem solving game, The Incredible Machine.

Considering the high popularity of The Incredible Machine as a game, it would be reasonable to expect students using that environment to experience more positive affect, and less negative affect than students using an intelligent tutoring system such as Aplusix. At the same time, it might be reasonable to expect more students to game the system when playing The Incredible Machine than Aplusix, since by its very nature a game may encourage gaming the system relative to an intelligent tutor.

The evidence from our research partially aligns with these expectations. There is indeed more gaming the system in TIM than Aplusix; however, surprisingly, affect was on the whole better within Aplusix than TIM – there was significantly more boredom and frustration in TIM, and a less flow.

This suggests that a well-designed intelligent tutoring system can lead to equally positive – or even more positive – affect than an educational game. In turn, this suggests that while factors such as fantasy may make games more fun [cf. 8], the interactivity and challenge common to both games and intelligent tutors may play a larger role in making games affectively positive learning environments.

The results in this paper are not fully definitive, however, for four reasons. First, there are a number of differences between the two studies. Although the two studies were conducted by the same research group with a single methodology, TIM and Aplusix cover different subject matter and the studies were conducted with samples recruited in different years (and differing subtly, demographically) rather than with random assignment within a single population. This is not a fatal flaw for the study presented here, but does suggest that its result should be replicated before being treated as proven truth (as, in fact, all research results should be). Second, TIM and Aplusix differ pedagogically from each other in a number of ways. In comparing an intelligent tutor to an educational game, multiple substantial differences between environments are unavoidable; games have several characteristics that distinguish them from other types of interactive environments [19], as do intelligent tutoring systems [23]. A comparison that varied on only one factor would not fairly represent one type of environment or the other; however, determining *which* factors lead to the largest positive improvements on student affect and behavioral choice will be key. Third, TIM and Aplusix differed substantially in terms of curricular relevance. While TIM fostered problem solving skills in general, Aplusix focused specifically on Algebra, a subject that the participants were studying at the time. Participants may have perceived Aplusix as relevant to the larger goal of getting good grades in mathematics, motivating them to invest more effort and attention when using the software [cf. 15]. Finally, affect's impact on learning can be counterintuitive. Positive affect in some cases appears to reduce perseverance and increase distraction [12]. On the other hand, the affective state of confusion, sometimes considered negative, has been shown to promote deep thinking [9].

In recent years, there has been rapidly increasing interest in educational games. Some of this interest has been based on the hypothesis that games will lead to better affect than existing learning environments [cf. 8,13]. However, in the research reported here, we have found that a traditional intelligent tutoring system can produce equally good – or better – affect as an award-winning educational game. 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 engaging and educationally effective learning environments.

Acknowledgements

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. The work presented in this paper was made possible in part by a grant from the Ateneo de Manila University and by the Pittsburgh Science of Learning Center which is funded by the National Science Foundation, award number SBE-0354420.

References

1. Alessi, S. M. & Trollip, S. R. (2001). *Multimedia for Learning: Methods and Development* (3rd Ed). Needham Heights, MA: Allyn & Bacon.
2. 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.
3. Baker, R.S., Corbett, A.T., Koedinger, K.R., and 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*, 383-390.
4. Baker, R.S.J.d, Rodrigo, M.M.T., Xolocotzin, U. (2007). The dynamics of affective transitions in simulation problem-solving environments. In A. Paiva, R. Prada & R. W. Picard (Eds.): *ACII 2007, LNCS 4738*, pp. 666-677, Germany: Springer-Verlag Berlin Heidelberg.
5. 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.
6. Brown, J.S. (2005). New learning environments for the 21st century. *The Magazine of Higher Learning, 38(5), 18-24.*
7. Cohen, J. (1960) A Coefficient of Agreement for Nominal Scales. Educational and *Psychological Measurement,* 20, 37-46.
8. 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.

9. Craig, S.D., Graesser, A.C., Sullins, J., Gholson, B. (2004). Affect & learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, **29** (3), 241-250.
10. Csikszentmihalyi, M. (1990) *Flow: The Psychology of Optimal Experience*. New York: Harper and Row.
11. 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" In conjunction with International conference on Intelligent User Interfaces, 7-13.
12. Dreisbach, G. & Goschke, T. (2004). How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **30***(2), 343-353.*
13. Gee, J.P. (2003) *What video games have to teach us about learning and literacy.* New York: Palgrave Macmillan. Keller, C. (1983) Motivational Design of Instruction. In Reigeluth, C. *Instructional Design Theories and Models: An Overview of their current status.* Mahwah, NJ: Lawrence Erlbaum Associates.
14. 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.
15. Keller, J. M. (1987). Strategies for Stimulating the Motivation to Learn. *Performance and Instruction Journal*, 28(8), 1-7.
16. 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. *Proceedings of ACM SIGCHI 2004*, pp. 1375-1378.
17. 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.
18. Nicaud, J. F., Bouhineau, D., Mezerette, S., Andre, N. (2007). Aplusix II [Computer software].
19. Prensky, M. (2007). *Digital Game-Based Learning*. St. Paul, MN: Paragon House.
20. 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.
21. Sierra Online. (2001). The Incredible Machine: Even More Contraptions [Computer Software].
22. Schofield, J.W. (1995) *Computers and Classroom Culture.* Cambridge University Press, Cambridge, UK.
23. VanLehn, K. (2006). The behavior of intelligent tutoring systems. *International Journal of Artificial Intelligence in Education, 16*(3), 227-265.
24. 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.