The Incidence and Persistence of Affective States While Playing Newton's Playground

Juan Miguel L. Andres
Department of Information Systems & Computer Science
Ateneo de Manila University
Quezon City, Philippines
miglimjapandres@gmail.com

Ma. Mercedes T. Rodrigo
Department of Information Systems & Computer Science
Ateneo de Manila University
Quezon City, Philippines
mrodrigo@ateneo.com

***Abstract* - We study the incidence (rate of occurrence), and persistence (rate of reoccurrence immediately after occurrence) of students' affective states while playing Newton's Playground (NP), an educational game for physics. We compare findings to those of previous studies', which were conducted using different populations and different educational games. Students' affective states are studied using quantitative field observations on public high school students. The incidence and persistence of boredom, frustration, confusion, concentration, delight, and surprise were compared. We found that boredom, concentration, and frustration tended to persist while interacting with NP. Concentration, boredom, frustration, and confusion were the most common states observed.**

Index Terms – Affective computing, affective persistence, Newton's Playground, serious games.

I. Introduction

A. *Serious Games and Affect*

Serious games are "virtual environments explicitly intended to educate or train" [1]. They are characterized by the coupling of content with game formats, and are most commonly used in educational settings [2]. Well-designed games integrate learning and enjoyment – where learning opportunities occur seamlessly within the game – while keeping the learner engrossed in the playing/learning experience.

Educators and researchers recommend the use of serious games to increase student motivation. Motivation is the desire of a student to commit to learning new things [3], and is said to play a crucial role in driving learning [4]. Good games are highly motivating [4]. Integrating games with school materials has potential to increase learning, especially among lower-performing students [1]. As such, studies have expressed the need for research on the effectiveness of serious games on learning, stating that focus on learning outcomes is just as crucial as focus on the learning and training process [2].

In [5], Graesser et al. note that there are few studies that empirically measure the effects of serious games on learning and the learning experience. The article speculates that interacting with a serious game could increase enjoyment and general interest in the topic. In order to properly assess how effective a game is in contributing to learning, the article recommends conducting behavioural and cognitive task analyses between game features and desired learning outcomes, and understanding emotion-learning connections.

Since then, researchers have turned their attention to the non-cognitive effects of games, specifically games' effects on "academic emotions" such as confusion, boredom, and flow. An assessment of games' effects on emotions is relevant precisely because games are used to foster intrinsic motivation.

As with previous work [6][7][8][9], we focus our attention on the dynamics of affective states among students using a serious game. Of particular interest to us are the incidence of affective states and the persistence of emotions leading to virtuous or vicious cycles.

This paper discusses the incidence and persistence of affective states within a serious game for physics, Newton's Playground. In the discussion, we speculate as to the reasons behind the patterns that we found and suggest future lines of investigation.

B. *Research Questions*

This paper focuses on three research questions related to students' cognitive-affective states during interactions with Newton's Playground:

1) What affective states do students experience more often?

2) Which affective states persist over time?

3) How do the incidence and persistence compare against those found within other serious games?

II. Methods and Learning Environment

Participants were asked to take a 16-item pre-test for 20 minutes, play with the software for 2 hours, and then take an equivalent 16-item post-test for another 20 minutes. As the students used the software, two trained observers noted their behaviors and affective states. This section provides the details regarding the participant profile, instruments, and data collection procedures.

A. *Participant Profile*

Data was gathered from 60 eighth grade public school students in Quezon City, Philippines. As of 2011, the school had 66 teachers and 1,976 students, divided into 37 advisory classes, occupying only 34 classrooms. Students are predominantly Filipino. In 2011, the school received a donation of 23 computers. A laboratory was built to house the units. The school is part of an urban neighborhood. Residents in this neighborhood are engaged in such occupations as shoemakers, barbers, meat or vegetable vendors, drivers, and laborers.

Students ranged in age from 13 to 16. Of the participants, 31% were male and 69% were female. Participants were asked to rate how frequently they played video games and watched television on a scale of 1 (not at all) to 7 (everyday, for more than 3 hours), and the resulting average frequency of gameplay is 3.2 (in between a few times a month, and a few times a week), and the resulting average frequency of watching television is 5.9 (in between everyday, but for less than 1 hour, and everyday, for 1-3 hours). Participants were asked for their most frequent grade on assignments, and on a scale of 0 (F) to 4 (A), the average most frequent grade of the participants is 3.1 (B).

B. *The Software*

Newton's Playground (NP) is a computer game for physics patterned after Crayon Physics Deluxe. It was designed to help secondary school students understand qualitative physics [10]. Qualitative physics is a nonverbal conceptual understanding of how the physical world operates, along the lines of Newtonian physics. Qualitative physics is characterized by an implicit understanding of Newton's three laws: balance, mass, and conservation and transfer of momentum, gravity, and potential and kinetic energy [11].

NP is a two-dimensional computer-based game that requires the player to guide a green ball to a red balloon. An example level that requires a pendulum to solve the level is shown in Figure I. The player uses the mouse to nudge the ball to the left and right (if the surface is flat), but the primary way to move the ball is by drawing or creating simple machines on the screen with the mouse and colored markers. The objects come to life once the object is drawn. Everything obeys the basic rules of physics relating to gravity and Newton's three laws of motion [11].

Fig. I Example level of Newton's Playground.

The 74 levels in NP require the player to solve the problems via drawing different simple machines, representing agents of force and motion: inclined plane/ramps, levers, pendulums, and springboards. Again, all solutions are drawn with colored markers using the mouse. A ramp is any line drawn that helps to guide a ball in motion. A ramp is useful when a ball must travel over a hole. A lever rotates around a fixed point, usually called a fulcrum or pivot point. Levers are useful when a player wants to move the ball vertically. A swinging pendulum directs an impulse tangent to its direction of motion. The pendulum is useful when the player wants to exert a horizontal force. A springboard (or diving board) stores elastic potential energy provided by a falling weight. Springboards are useful when the player wants to move the ball vertically.

C. *The Observation Protocol*

The Baker-Rodrigo-Ocumpaugh Monitoring Protocol (BROMP) is a protocol for quantitative field observations of student affect and behavior. BROMP is a holistic coding procedure that has been used in thousands of hours of field observations of students, from kindergarten to undergraduate populations. It has been used for several purposes, including to study the engagement of students participating in a range of classroom activities (both activities involving technology and more traditional classroom activities) and to obtain data for use in developing automated models of student engagement with Educational Data Mining (EDM) [12]. Within BROMP, each student observation lasts 20 seconds, and the observers move from one student to the next in a round robin manner during the observation period. Figure II shows two BROMP coders observing as students interact with the Newton's Playground.

Fig. II BROMP coders at work.

During each 20-second period, each observer independently coded the student's affective state. The affective categories were drawn from D'Mello, Craig, Witherspoon, McDaniel, and Graesser [13] and Rodrigo et al. [6]:

1) Boredom — slouching, and resting the chin on his palm, statements such as "Can we do something else?" and "This is boring!"

2) Confusion — scratching his head, repeatedly looking at the same interface elements, statements such as "I don't understand?" and "Why didn't it work?"

3) Delight — smirking, smiling at the computer monitor, statements such as "Yes!" or "I got it!"

4) Happy — clapping hands or laughing with pleasure, less contained expressions of joy

5) Concentration — immersion and focus on the system, a subset of the flow experience described by Csikszentmihalyi [14], leaning toward the computer or mouthing solutions to himself while solving a problem.

6) Frustration — banging on the keyboard or pulling at his hair, statements such as "This is annoying!" or "What's going on?!"

7) Surprise — jerking back suddenly or gasping, statements such as "Huh?" or "Oh, no!"

8) Curious — asking for help from seatmate, statements such as "How did you do that?" or "Please help me out."

9) Pride — boasting of progress, challenging of other students to get more badges

10) Anxious — uneasy movement, checking of watch to see how much time was left

11) Sad — laying the head on the table, expressions of hopelessness, such as "I don't want to do this anymore."

12) Angry — some violent behavior such as banging hand or head against the table.

If the student exhibits two or more distinct states during a 20-second period, the observers only coded the first state. To illustrate: Suppose that, at the start of the 20-second period, a student was asking a teacher or classmate for help with the soft- ware. At that point in time, the student is considered to be confused. If the student's problem is solved and he returns to work, he is considered to have transitioned into engagement. In cases such as these, only the first affective state, confusion, was recorded.

The behaviors observed were on-task, off-task, stacking, and a behavior called without thinking fastidiously (WTF), a behavior in which, despite a student's interaction with the software, "their actions appear to have no relationship to the intended learning task [15]." The analysis of the behaviors, however, is outside this paper's scope.

The inter-coder reliability for affect was acceptably high with a Cohen's [16] Kappa of 0.67. The typical threshold for certifying a coder in the use of BROMP is 0.6, established across dozens of studies as well as the previous affective computing literature.

D. *The Affect Coding Tool*

The Human Affect Recording Tool, or HART, is an Android application developed to guide researchers in conducting quantitative field observations according to the BROMP protocol. The application synchronizes the coded observations to Internet time, allowing for more precise synchronization with log file data from the educational software under study.

HART asks for input regarding school and classroom information, coding schemes to be used, and the student IDs of the students to be observed during the session. The application then presents the student IDs in the order entered, allowing BROMP observers to more conveniently code affect and behavior until the session is manually terminated. All observations are logged on a text file that is locally stored on the device used to run HART. The application and all its functions are discussed in more detail in [12].

III. RESULTS

We collected pre-test and post-test data from each student (N=60). Scores were generally poor. Students averaged 6.02 correct answers on both the pre-test and the post-test, out of a highest possible score of 16. This indicates that no learning improvements were detected.

We also collected a total of 36 observations per student (N=60) per observer over the 2 hours of gameplay, for a total of 3,456 observations. We summarized these observations and compared them against results from two earlier studies that used similar methodologies.

The first study made use of The Incredible Machine, a simulation problem-solving game [6]. The game challenges the player to solve different problem scenarios – getting a mouse to a block of cheese, getting a ball into a bin, etc. – using a combination of tools each level provides.

The second study made use of Math Blaster, a pre-algebra game [8]. The player plays the role of a galactic commander stranded on a planet of monkeys. To escape the planet and return home, the player must collect medallions that he can then offer the monkey king. He earns the medallions by solving whole number, decimal, and fraction arithmetic problems.

The affective states observed in these two studies were boredom, concentration, confusion, delight, frustration, and surprise. For comparability, the analyses done in this paper were limited to just these 6 affective states.

A. *Incidence of Affective States*

Table I shows that concentration (72%) was the most commonly observed affective state. This finding is consistent with quantitative field observations conducted on The Simple Machine (62%) [6] and Math Blaster (63%) [8].

The second most frequently observed state was confusion (8%). This finding is consistent with the incidence observed in The Incredible Machine (11%) [6], but higher than the incidence observed in Math Blaster (2%) [8].

TABLE I
INCIDENCE OF AFFECTIVE STATES WITHIN NEWTON'S PLAYGROUND

Affective State	Incidence of Affective State
Boredom	7%
Concentration	72%
Confusion	8%
Delight	1%
Frustration	7%
Surprise	0%

The third most frequently observed states were boredom and frustration (7% for both). The incidences are consistent with those observed in The Incredible Machine (7% for both) [6]. However, incidence of boredom was high in Math Blaster (22%) [8], which the study speculates was caused by how easy the students found the problems. The Math Blaster study did not observe any frustration during gameplay.

B. Persistence of Affective States

The study analyzed how frequently a student transitioned from one affective state to another. In conducting these analyses, the study takes into account the base rates of each category. In order to appropriately account for the base rate of each cognitive-affective category in assessing how likely a transition is, we adopt D'Mello et al.'s [13] transition likelihood metric, *L*. D'Mello et al.'s *L* gives the probability that a transition between two states will occur, given the base frequency of the destination state. Thus, if engagement occurred 70% of the time, then a 70% probability exists for any given affective state to transition into engagement. If confusion transitions to engagement 70% of the time, the transition is no better than chance. If, however, confusion transitions to engagement 85% of the time, this transition may be significant. *L* is computed as in (1).

$$L = \frac{\Pr(NEXT \mid PREV) - \Pr(NEXT)}{(1 - \Pr(NEXT))}, \qquad (1)$$

L is scaled between 1 and $-\infty$. A value of 1 means that the transition will always occur; a value of 0 means that the transition's likelihood is exactly what it would be given only the base frequency of the destination state. Values above 0 signify that the transition is more likely than it could be expected to be, given only the base frequency of the destination state. Values under 0 signify that the transition is less likely than it could be expected to be, given only the base frequency of the destination state.

For a given transition, we calculate a value for L for each student and then calculate the mean and standard error across students. We can then determine if a given transition is significantly more likely than chance (0), given the base frequency of the next state, using the two-tailed t test for one sample.

TABLE II
PERSISTENCE OF AFFECTIVE STATES WITHIN NEWTON'S PLAYGROUND

	BOR	CONC	CONF	FRU
BOR	0.30 (0.05), $p < 0.01$	-0.92 (0.20), $p < 0.01$	-0.05 (0.01), $p < 0.01$	0.02 (0.03), $p = 0.62$
CONC	-0.03 (0.01), $p < 0.01$	0.21 (0.07), $p < 0.01$	$-1.56E^{-3}$ (0.01), $p = 0.88$	-0.01 (0.01), $p = 0.16$
CONF	$-5.08E^{-3}$ (0.03), $p = 0.88$	-0.28 (0.17), $p = 0.11$	0.04 (0.03), $p = 0.12$	$4.17E^{-3}$ (0.02), $p = 0.84$
FRU	0.02 (0.03), $p = 0.62$	-0.46 (0.18), $p = 0.01$	$-9.38E^{-4}$ (0.02), $p = 0.96$	0.10 (0.04), $p < 0.01$

Note: BOR = Boredom; CONC = Concentration; CONF = Confusion; FRU = Frustration. The first number in each cell is the mean value of D'Mello's *L*, the number in the parenthesis is standard error.

In Table II, horizontal rows represent previous affective states, and vertical columns represent affective states 200 seconds later. The first number in each cell is the mean value of D'Mello's *L* across students, the number in parenthesis is the standard error. Cells in light gray represent transitions that were statistically significant ($p \leq .05$). We found seven significant transitions.

We found significance to the likelihood that a student who is bored will stay bored and is unlikely to transition into concentration 200 seconds later. Likewise, frustrated students were likely to stay frustrated, and were unlikely to transition into concentration.

Boredom's vicious cycle is consistent quantitative field observations conducted in both The Incredible Machine and Math Blaster. This finding supports previous implications that boredom is an undesirable state. Both Math Blaster and The Incredible Machine had a hard time transitioning the students out of boredom and into a more productive state such as concentration. The persistence of frustration, however, was not present in both previous studies.

We also found significance to the likelihood that a student who is concentrating is likely to stay concentrating, which is consistent with the marginally significant finding that concentration persists within The Incredible Machine. The Math Blaster study, however, reported that concentration was not persistent.

Interesting to note was that confusion did not persist within Newton's Playground. The Incredible Machine study reported a marginal significance to the likelihood that a student who was confused was likely to stay confused. Conversely, students who were confused in the Math Blaster study were likely to transition out into concentration.

IV. Discussion, Conclusions, and Future Work

Motivated by the call to study the effects of games on students' emotions, this study sought to examine the affective states experienced by students playing a serious game for physics, Newton's Playground (NP), and investigate which of these affective states persisted during gameplay. The study then sought to conduct a comparative analysis between this study's findings and those that resulted of two other previously studied serious games, The Incredible Machine and Math Blaster.

This study found that concentration, boredom, frustration, and confusion were the most commonly observed affective states among students using NP. The incidence of concentration was consistent with those found in both The Incredible Machine and Math Blaster studies. Incidence of confusion was consistent with findings in The Incredible Machine study, but was higher than the incidence found in Math Blaster. Conversely, while findings regarding incidences of boredom and confusion were consistent between Newton's Playground and The Incredible Machine, findings on boredom were much higher in Math Blaster. However, Math Blaster reported no observations of frustration.

In running analyses to find which affective states persisted over time, this study found that both boredom and frustration followed vicious cycles wherein students who were observed to be bored were likely to stay bored, and students who were observed to be frustrated were likely to stay frustrated. Students observed to be in either affective state were unlikely to transition out into more productive affective states, like concentration. Boredom's vicious cycle is consistent with findings in both previous studies, but neither study found any persistence in frustration.

Students who were observed to be in more productive affective states, however, such as concentration, were likely to stay concentrating, and were unlikely to transition into boredom. This finding is consistent with findings in The Incredible Machine.

All three studies reported different results regarding the persistence of confusion. No persistence of confusion is present in NP. The Incredible Machine reported that confused students were likely to stay confused, while Math Blaster reported that confused students were likely to transition into concentration.

That boredom and frustration occurred and persisted within Newton's Playground and other educational games, why they occur and persist, and how they affect learning are of interest to educational game developers and researchers because they demand that we take a nuanced approach to the designing and using educational games. What makes the incidence and persistence of boredom and frustration even more interesting to us is that they occurred among students with limited educational computer usage experience—not even the novelty effect disrupted these vicious cycles.

We speculate that there are a number of relationships that are worth further exploration. Poor prior knowledge (as evidenced by students' poor pre-test results) might have made the game daunting. The game interaction time of two hours may have been too long, leading to boredom. Indeed, the researchers noticed that the students rushed through the post-test, implying that they wanted to leave the testing area as quickly as possible. Boredom might have led to systematic guessing and other similar non-learning behaviors, leading in turn to poor post-test scores [17]. In future work, we intend to verify which among these hypotheses the data support. In doing so, we hope to contribute to principles that guide the development of good educational games.

ACKNOWLEDGMENTS

We would like to thank the Ateneo Center for Educational Development, Carmela C. Oracion, Christopher Ryan Adalem, the officials at Krus Na Ligas High School for making this study possible. We also thank Ryan Baker, Valerie Shute, Matthew Small, Matthew Ventura, Luc Paquette and the Gates Foundation Grant #OP106038 for collaborating with us.

REFERENCES

[1] Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. Serious games: Mechanisms and effects, 295-321.

[2] Breuer, J.S., & Bente, G. "Why so serious? On the relation of serious games and learning." Eludamos. Journal for Computer Game Culture 4.1 (2010): 7-24.

[3] DiSessa, A.A. (2000). "Changing Minds." MIT Press, Cambridge, MA.

[4] Gee, J. P. (2003). What video games have to teach us about learning and literacy. Computers in Entertainment (CIE), 1(1), 20-20.

[5] Graesser, A., Chipman, P., Leeming, F., & Biedenbach, S. (2009). Deep learning and emotion in serious games. Serious games: Mechanisms and effects, 83-102.

[6] 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., et al. (2007). Affect and usage choices in simulation problem-solving environments. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), 13th International Conference on Artificial Intelligence in Education (pp. 145-152).

[7] Baker, R. S., Rodrigo, M. M. T., & Xolocotzin, U. E. (2007). The dynamics of affective transitions in simulation problem-solving environments. In Affective Computing and Intelligent Interaction (pp. 666-677). Springer Berlin Heidelberg.

[8] Rodrigo, M. M. T. (2010). Dynamics of student cognitive-affective transitions during a mathematics game. Simulation & Gaming.

[9] Rodrigo, M. M. T., & d Baker, R. S. (2011). Comparing the incidence and persistence of learners' affect during interactions with different educational software packages. In New perspectives on affect and learning technologies (pp. 183-200). Springer New York.

[10] Shute, V., & Ventura, M. (2013). Stealth assessment: Measuring and supporting learning in video games. MIT Press.

[11] Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and Learning of Qualitative Physics in Newton's Playground. The Journal of Educational Research, 106(6), 423-430.

[12] Ocumpaugh, J., Baker, R.S.J.d., Rodrigo, M.M.T. (2012) Baker-Rodrigo Observation Method Protocol (BROMP) 1.0. Training Manual version 1.0. Technical Report. New York, NY: EdLab. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.

[13] D'Mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., & Graesser, A. (2005). Integrating affect sensors in an intelligent tutoring system. In Proceedings of the Workshop on Affective Interactions: The computer in the affective loop workshop, International conference on intel- ligent user interfaces (pp. 7-13). New York: Association for Computing Machinery.

[14] Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. New York: Harper Perennial.

[15] Wixon, M., Baker, R. S., Gobert, J. D., Ocumpaugh, J., & Bachmann, M. (2012). WTF? detecting students who are conducting inquiry without thinking fastidiously. In User Modeling, Adaptation, and Personalization (pp. 286-296). Springer Berlin Heidelberg.

[16] Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20 (1960), 37-46.

[17] Baker, R. S. d. J., D'Mello, S., Rodrigo, M. M. T. & Graesser. (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.