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

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

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

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 concentration, confusion, frustration, boredom, surprise, and delight were compared. We found that boredom (D’Mello’s $L = 0.3, p < 0.01$), concentration ($L = 0.21, p < 0.01$), and frustration ($L = 0.10, p < 0.01$) tended to persist while interacting with NP. The most frequently observed affective states were concentration (72%), confusion (8%), boredom (7%), and frustration (7%).

Index Terms— Affective computing, affective persistence, Newton’s Playground, serious games.

I. INTRODUCTION

Serious games are utilized to educate within virtual environments [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 the capacity to improve learning, especially among the lower-performers [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 the learning experience and learning outcomes. 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 behavioral and cognitive task analyses between game features and desired learning outcomes, and understanding emotion-learning connections.

Since then, attention has been turned to the non-cognitive effects of games, specifically games’ effects on “academic emotions” such as confusion, boredom, and concentration. 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 particular, this paper investigates the most frequently experienced affective states, the persistent affective states, and how these findings compare against those found within other serious games. In the discussion, we speculate as to the reasons behind the patterns that we found and suggest future lines of investigation.

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, which is a conceptual understanding of how the physical world works along the lines of Newtonian physics [10]. It is characterized by an inherent understanding of the concepts surrounding Newton’s three laws of motion: balance, mass, the conservation and transfer of momentum, gravity, and potential and kinetic energy [11].

NP is a two-dimensional 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 Fig. 1. Using the mouse, the player can nudge the ball to the left and right, but the foremost manner of getting the ball to the balloon is by using the mouse to draw and create simple machines. The objects come to life once the object is drawn. All objects drawn and all interactions the student has with the game follow the basic rules of physics relating to gravity and motion [11]. The game’s mechanics are explained in [20].

C. The Observation Protocol

The Baker-Rodrigo-Ocumpaugh Monitoring Protocol (BROMP) is a protocol for quantitative field observations of student affect and behavior. This holistic coding procedure has been used in thousands of hours of field observations, and across all student levels. 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. Fig. 2 shows two BROMP coders observing as students as they interact with Newton’s Playground. BROMP is explained in greater detail in [12].

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], Rodrigo [8], and Rodrigo et al. [6]:

1) Concentration — immersion and focus on the task at hand, leaning toward the computer and attempting to solve the level, a subset of the flow experience described in [14].
2) Confusion — scratching his head, repeatedly attempting to solve the same level, statements such as “I don’t understand?” and “Why didn’t it work?”
3) Boredom — slouching, sitting back and looking around the classroom for prolonged periods of time, statements such as “Can we do something else?” and “This is boring!”
4) Delight — smirking, smiling at the computer monitor, statements such as “Yes!” or “I got it!”
5) Happy — clapping hands or laughing with pleasure, less contained expressions of joy
6) Frustration — banging on the keyboard or lying his head down on the table, wanting to leave and do something else, 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!”
functions are discussed in more detail in [12].

The device used to run HART. The application and all its observations are logged on a text file that is locally stored on behavior until the session is manually terminated. All Bromp observers to more conveniently code affect and then presents the student IDs in the order entered, allowing the students to be observed during the session. The application information, coding schemes to be used, and the student IDs of synchronizing with log file data from Newton’s Playground.

Observations to Internet time, allowing for more precise Bromp protocol. The application synchronizes the coded conducting quantitative field observations according to the D.

Computing literature. 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 consolidated these findings and compared them against those from two previous studies that employed 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 – helping a mouse get to a piece cheese, getting balls into bins, etc. – using a combination of tools each level provides.

The second study made use of Math Blaster, a pre-algebra game [8]. The player takes the role of a galactic commander stranded on a monkey planet. To escape 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 confusion and the percentage of level attempts 1) that resulted in the student earning gold and silver badges, and 2) wherein the student was flagged for stacking, a form of gaming the system – or the systematic misuse of the system features to advance in the game without learning the content [18] – within NP. The study found confusion to be positively correlated with earning a silver badge, negatively correlated with earning a gold badge, and positively correlated with stacking.

Earning a gold badge is indicative of mastery of the four simple machines in the game. As such, the insight behind the paper’s findings was that confused students lacked mastery, experimenting with different objects in trying to solve levels. Its positive correlation with stacking is another indication of lack of mastery.

In a paper by Andres and Rodrigo [20], analyses were conducted to investigate the relationships that exist between affect and learning within NP. In particular, the study identified four learning trajectories (one for each of the simple machines used to solve levels in the game) that traced student performance over time, and two affect trajectories that traced the percentage of participants observed to be bored and confused over the 2 hours of gameplay. The study then attempted to find a relationship between these learning and affect trajectories. The study found that 1) students performed more poorly over time, 2) a steady number of students were observed to be confused throughout gameplay, 3) the percentage of students observed to be bored increased over time, and 4) confusion was positively correlated with earning a silver badge, which supports findings in [19].

IV. 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 pretest and the posttest, 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 consolidated these findings and compared them against those from two previous studies that employed 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 – helping a mouse get to a piece cheese, getting balls into bins, etc. – using a combination of tools each level provides.

The second study made use of Math Blaster, a pre-algebra game [8]. The player takes the role of a galactic commander stranded on a monkey planet. To escape 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 confusion and the percentage of level attempts 1) that resulted in the student earning gold and silver badges, and 2) wherein the student was flagged for stacking, a form of gaming the system – or the systematic misuse of the system features to advance in the game without learning the content [18] – within NP. The study found confusion to be positively correlated with earning a silver badge, negatively correlated with earning a gold badge, and positively correlated with stacking.

Earning a gold badge is indicative of mastery of the four simple machines in the game. As such, the insight behind the paper’s findings was that confused students lacked mastery, experimenting with different objects in trying to solve levels. Its positive correlation with stacking is another indication of lack of mastery.

In a paper by Andres and Rodrigo [20], analyses were conducted to investigate the relationships that exist between affect and learning within NP. In particular, the study identified four learning trajectories (one for each of the simple machines used to solve levels in the game) that traced student performance over time, and two affect trajectories that traced the percentage of participants observed to be bored and confused over the 2 hours of gameplay. The study then attempted to find a relationship between these learning and affect trajectories. The study found that 1) students performed more poorly over time, 2) a steady number of students were observed to be confused throughout gameplay, 3) the percentage of students observed to be bored increased over time, and 4) confusion was positively correlated with earning a silver badge, which supports findings in [19].

IV. 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 pretest and the posttest, 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 consolidated these findings and compared them against those from two previous studies that employed 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 – helping a mouse get to a piece cheese, getting balls into bins, etc. – using a combination of tools each level provides.

The second study made use of Math Blaster, a pre-algebra game [8]. The player takes the role of a galactic commander stranded on a monkey planet. To escape 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.
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 occurring affective state. This finding is consistent with those from the studies conducted on The Simple Machine (62%) [6] and Math Blaster (63%) [8].

The second most frequently occurring affective 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].

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 the participants transitioned from one affective state to another. In conducting these analyses, the study took into consideration the base rates of each affective state. In order to properly account for the base rate of each state in assessing how likely a transition is, this study adopted 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. It is explained in greater detail in [13][8], and is computed as in

\[
L = \frac{\Pr(\text{NEXT} | \text{PREV}) - \Pr(\text{NEXT})}{(1 - \Pr(\text{NEXT}))}. 
\]

In Table II, previous affective states are listed horizontally, and the affective states observed 200 seconds later are listed vertically. The first number in each cell is the mean value of D’Mello’s \( L \), the number in the parenthesis is standard error. Cells in light gray signify transitions that were statistically significant (\( p \leq 0.05 \)). We found seven significant transitions.

### Table I. Incidences of Affective States within NP

<table>
<thead>
<tr>
<th>Affective State</th>
<th>Incidence of Affective State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boredom</td>
<td>7%</td>
</tr>
<tr>
<td>Concentration</td>
<td>72%</td>
</tr>
<tr>
<td>Confusion</td>
<td>8%</td>
</tr>
<tr>
<td>Delight</td>
<td>1%</td>
</tr>
<tr>
<td>Frustration</td>
<td>7%</td>
</tr>
<tr>
<td>Surprise</td>
<td>0%</td>
</tr>
</tbody>
</table>

### Table II. Persistence of Affective States within NP

<table>
<thead>
<tr>
<th></th>
<th>BOR</th>
<th>CONC</th>
<th>CONF</th>
<th>FRU</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOR</td>
<td>0.30</td>
<td>-0.02</td>
<td>-0.05</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.20)</td>
<td>(0.01)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>p</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>CONC</td>
<td>-0.03</td>
<td>0.21</td>
<td>-1.56E^3</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.07)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>p</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>CONF</td>
<td>-5.68E^3</td>
<td>-0.28</td>
<td>0.04</td>
<td>4.17E^2</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.17)</td>
<td>(0.03)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>p</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>FRU</td>
<td>0.02</td>
<td>-0.46</td>
<td>-9.38E^4</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td>(0.18)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>p</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>

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.

As in the previous studies, this analysis found a significant probability that a student who is bored will stay bored (\( L = 0.30, p < 0.01 \)) and is unlikely to transition into concentration 200 seconds later (\( L = -0.92, p < 0.01 \)). Likewise, frustrated students were likely to stay frustrated (\( L = 0.10, p < 0.01 \),) and were unlikely to transition into concentration (\( L = -0.46, p < 0.01 \)).

Boredom’s vicious cycle is consistent among 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 found it difficult to transition the participants out of boredom and into a more productive state like concentration. The persistence of frustration, however, was not present in both previous studies.

The study also found a significant probability that a student who is concentrating is likely to stay concentrating (\( L = 0.21, p < 0.01 \)), 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 significant probability 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.

V. 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.

ACKNOWLEDGMENT

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, Lue Paquette and the Gates Foundation Grant #OP106038 for collaborating with us.

REFERENCES


