A Comparison of Learners' Affect and Behaviors While Using an Intelligent Tutor and an Educational Game

Ma. Mercedes T. Rodrigo[1]
mrodrigo@ateneo.edu

Ryan S. J. d. Baker[2]
rsbaker@cs.cmu.edu

Nathaniel Abalos[1]
nathaniel.abalos@gmail.com

Karina Bacuyag[1]
karinapazb@gmail.com

Bronson Basuel[3, 4]
bronson.basuel@gmail.com

Melissa Bautista[1]
melissa.a.bautista@gmail.com

Michelle Cortez[1]
michellemanalangcortez@gmail.com

Godfrey Dulla[1]
gdulla8@gmail.com

Saichi Elomina[1]
ihcias@gmail.com

Mary Audrey Gineta[3]
narnian1019@gmail.com

Ahrren Rara[3]
aina.ahrren@gmail.com

Ramon Rodriguez[1]
mon.benny@gmail.com

Jeff Sanggalang[3, 4]
shiropot@gmail.com

Jessica Sugay[1]
jsugay@ateneo.edu

Ann Katherine Tan[1]
kathytan_15@yahoo.com

Mike Tan[3]
mike.araz@gmail.com

Emily Trajano[3]
emtrajano@yahoo.com

Fara Uy[1]
farauy@gmail.com

Noel Victorino[1]
nvictorino@gmail.com

Kathrina Blanca Villaflor[1]
blu.angel.17@gmail.com

ABSTRACT

We compared the extent to which of students exhibit certain affect states and behaviors while using an intelligent tutoring system, Aplusix, and an educational game, Math Blaster 9-11. We found that students using Math Blaster showed more boredom and less flow than the students who used Aplusix. Despite exhibiting less flow, students using the game spent more time on-task and less time in conversation or in gaming the system (defined as hint abuse or systematic guessing) that those using Aplusix. We were able to associate affective states with certain behaviors. We found that, in Aplusix, gaming the system tended to co-occur with boredom and confusion. In both Aplusix and Math Blaster, confusion and delight tended to dampen on-task behavior but foster on-task conversation. Also in both cases, flow was associated with on-task behavior and is not associated with off-task behaviors. Findings from this study suggest that, regardless of software format, boredom is an undesirable affective state. Flow is clearly desirable. Confusion and delight reduce solitary on-task behavior yet foster greater communication with teachers or classmates. The findings also provide some evidence that, regardless of the affective states they evoke, game formats tend to keep students on-task.

Keywords

Affect, Intelligent Tutoring Systems, Educational Games, Comparative Study

1. CONTEXT

In 2008, Rodrigo et al [9] published a study comparing the affect and behavior exhibited by students while using an intelligent tutor, Aplusix [8], and a simulation problem solving game, The Incredible Machine [11]. The study was intended to provide empirical evidence supporting or questioning the assumption that the embedding learning content in a game format will lead to better student motivation and therefore increased learning gains. The study found that student affect was in fact better in Aplusix than it was in The Incredible Machine, with students exhibiting less boredom and frustration, and more flow. While time on-task in both conditions was not significantly different, there was less "gaming the system" (defined as systematic guessing or help abuse [cf. 2]) in Aplusix. Students using Aplusix also exhibited more on-task conversation.

In the study's conclusions, though, the researchers cited the comparison's limitations: The subject matter of the two software packages differed radically. The Incredible Machine was a problem-solving game while Aplusix was an algebra tutor. Aplusix was clearly more relevant in terms of the curriculum. The populations using the two software packages differed. One of the conclusions, therefore, was that the study needed to be replicated under stricter conditions in order to validate the findings. This paper is the first report of the follow-up study conducted in conjunction with the 2008 publication.

2. GAMES AS VEHICLES FOR LEARNING

If learning is fun, will students learn more? Does presenting educational content in the form of a computer game lead to greater motivation and, by that token, greater achievement? The motivation, rewards, interactivity, challenge and fantasy that games provide seem to be natural ways to engage the digital natives [12], yet the answers to questions of motivational effect and educational effectiveness vary from case to case. Lepper and Cordova [7] found that some games excite but do not educate. Many commercial games are designed so that students can obtain winning scores through a series of rapid, mindless actions not predicated on the learning the material. Indeed, there are cases when the goals of learning and winning conflict rather than concur. In other studies, students have actually responded negatively to the use of games as learning agents [3]. Contrary to the researchers' observations and the teachers' impressions, students seemed to undervalue the games as learning tools, regarding them instead as rewards rather than actual carriers of content.

There have been cases, though, that illustrate the benefits of game formats. In more controlled experiments with carefully designed educational games, [7] found that motivational embellishments increased student achievement and led to greater interest in continuing work on similar subject matter. Lee et al [6] found that students using educational games spent more time on task. They also tended to increase the difficulty levels of the problems they were solving without prompting from the teacher or the researchers.

In other studies, the student responses were a qualified "it depends." Students tended to prefer games that required logic, memory, visualization, and problem solving over those that did not [1]. De Aguilera and Mendiz's [5] literature review singled out simulations as having enormous educational potential. Aside from being highly motivating, simulations help students acquire practical skills such as problem solving, strategy assessment, and organization.

3. RESEARCH QUESTIONS

We hope to contribute some answers to the debate described above by providing empirical answers to the following questions:

a. Does student achievement vary depending on educational software format?

b. Do students using a game format exhibit positive affective states more frequently?

c. What behaviors tend to accompany what affective states?

This study is part of a bigger research project with the medium term objective of arriving at computationally tractable models of student affect and behavior. By answering these questions above, we hope to find commonalities in the affective states and behaviors that different software formats evoke. We also hope to provide evidence connecting affective states with behaviors.

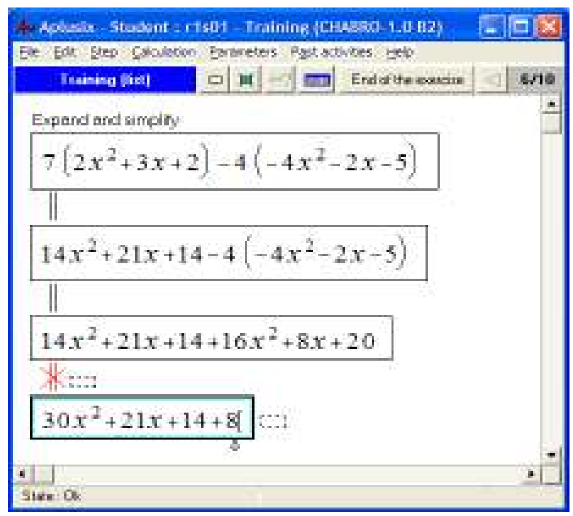
4. DESCRIPTIONS OF THE LEARNING ENVIRONMENTS

For this study, we retained Aplusix as our intelligent tutoring system. As a comparison condition, we with a math game, Math Blaster 9-12, with a similar scope. We describe the two software packages in this section.

4.1 Aplusix

Aplusix [8] 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. Students then solve the problem one step at a time. At each step, Aplusix displays equivalence feedback: two black parallel bars mean that the current step is equivalent to the previous step, two red parallel bars with an X mean that the current step is not equivalent to the previous step (Figure 1). Aplusix does not indicate which part of the current step requires further editing. When the student believes he or she is done, he can end the exercise. 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.

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



4.2 Math Blaster 9-12

Math Blaster 9-12 by Knowledge Adventure (Figure 2) 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.

5. METHODS

The study took place in the Ateneo de Manila University Grade School, an all-boys school located in Quezon City, Philippines. 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. Two sections used Aplusix while two sections used Math Blaster. In each case, contact time with the software lasted for 40 minutes.

Figure 2. Screen shot of Math Blaster 9-12

None of the students had prior experience with Math Blaster 9-12 or Aplusix before. All of them were computer-savvy and regularly played video games.

All students were given a pre-test and an identical post-test. The test had ten math problems taken from the software. The students were given 10 minutes at the start of the period to answer the pre-test and another 10 minutes at the end of the period to complete the post-test.

Of the 164 participants, 59 students were observed for affect and behavior—29 for Aplusix and 30 for Math Blaster. The students were not told who among them were being observed.

The observations were carried out by a team of six observers, working in pairs (Figure 3). The observers were Masters students in Education or Computer Science. Most had teaching experience. In this study, both affect and behavior were coded (discussed later). The observers trained for the task through a series of pre-observation discussions on the meaning of the affective and behavior categories and through an earlier observation exercise conducted at a different school. Observations were oriented with 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 this study.

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 and alternated between them. Since each observation lasted twenty seconds, each student was observed once 200 seconds. It was essential for observers look at the same student at the same time, therefore observers were synchronized using a timed PowerPoint presentation.

Within an observation, each observer coded the student's affective state. The affective categories coded were drawn from [10]. The categories coded were:

Figure 3. Two observers at work

1. **Boredom** – behaviors such as slouching, and resting the chin on his/her palm; statements such as "Can we do something else?" and "This is boring!".
2. **Confusion** – behaviors such as scratching his/her head, repeatedly looking at the same interface elemtents; statements such as "I don't understand?" and "Why didn't it work?"
3. **Delight** – behaviors such as clapping hands or laughing with pleasure; statements such as "Yes!" or "I got it!"
4. **Surprise** – behaviors such as jerking back suddenly or gasping; statements such as "Huh?" or "Oh, no!"
5. **Frustration** – behaviors such as banging on the keyboard or pulling at his/her hair; statements such as "This is annoying!" or "What's going on?!?"
6. **Flow** – complete immersion and focus upon the system [4]; behaviors such as leaning towards the computer or mouthing solutions to him/herself while solving a problem
7. The **Neutral** state, which was coded when the student did not appear to be displaying any of the affective states above, or the student's affect could not be determined for certain.

Behavior categories were also coded, using a coding scheme drawn from [10], and are as follows:

1. **On-task** – working within Aplusix or Math Blaster
2. **Giving or receiving answers** – discussing the mathematical content of the software the teacher or with a fellow student
3. **Other on-task conversation** – discussing some other aspect of the software with the teacher or with a fellow student, e.g. controls or objectives of the software
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)

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

To test the level of agreement between raters, we computed for Cohen's Kappa. Kappa was acceptably high in both cases, with κ=0.70 for affect. κ=0.63 for behavior, indicating that raters did not agree by chance.

6. RESULTS

We group the discussion of the results into three sub-parts: learning gains, incidence of affective states between the two software packages, and the co-occurrence of certain behaviors with certain affective states.

6.1 Learning gains

We scored the pre- and post-tests for all students. We used t-tests to compare the pre- and post-tests of the Aplusix group with the pre- and post-tests of the Math Blaster group, respectively. The test showed no significant differences, implying that the groups were homogenous. T-tests found no significant difference between pre- and post-test scores for the Aplusix group. Neither did we find significant differences in the pre-and post-test scores of the Math Blaster group. This implies that the students did not experience any learning gains from either software package.

Although differences were not significant, these findings should be approached with caution. Pre-test and post-test scores in both conditions were generally high, with means of 8.7 or 8.8 out of 10. This implies that the students already had mastery of the subject matter, so additional drills might not have been useful.

6.2 Frequency of affective states and behaviors

We examined the extent to which students exhibited certain affective states (Table 1) and behaviors (Table 2). We found that flow was the most prevalent affective state for both software packages, followed by boredom and delight. No frustration or surprise was observed during this study.

In terms of behavior, students were most commonly on-task or engaged in giving and receiving answers. The observers did not note any other on-task conversation in Aplusix nor any gaming in Math Blaster. The absence of gaming in Math Blaster is a novelty. In previous studies, incidence of gaming ranged from 7.5% in The Incredible Machine [10], 3% in Cognitive Tutor [2] and 1.4% in Aplusix[10]. The affective profile of the students using Aplusix varied somewhat from the profile published in [9]. In [9], students exhibited Boredom only 3% of the time, confusion 13% of the time, and flow 68% of the time and frustration 2% of the time.

To determine whether any of these differences were significant, we performed a between-subjects analysis using series of two-tailed t-tests assuming equal variances. We found that students using Math Blaster exhibited marginally significantly more boredom than those using Aplusix (t(57)=-1.91; p=0.06). Students exhibited marginally significantly more delight in Math Blaster than Aplusix (t(57)=-1.73; p=0.08). Students using Aplusix showed significantly more flow (t(57)=2.83; p=0.006). That students using Aplusix exhibited more flow and less boredom is consistent with findings published in [9].

Table 1. Incidence of affective states in Aplusix and Math Blaster; Significant differences in dark grey; Marginally significant differences in light grey

Affective State	Aplusix	Math Blaster
Boredom	13%	22%
Confusion	4%	2%
Delight	6%	12%
Flow	76%	63%
Frustration	0%	0%
Surprise	0%	0%
Neutral	1%	1%

Table 2. Incidence of behaviors in Aplusix and Math Blaster; Significant differences in dark grey; Marginally significant differences in light grey

Behavior	Aplusix	Math Blaster
On-task	77%	88%
Giving and receiving answers	16%	7%
Other on-task conversation	0%	1%
Off-task conversation	3%	3%
Solitary off-task	1%	1%
Inactive	1%	1%
Gaming	2%	0%

Students using Math Blaster tended to exhibit more on-task behavior (t(57)=-3.34; p=0.001). Students using Aplusix exhibited more giving and receiving answers (t(57)=3.35; p=0.001) and gaming (t(57)=2.41; p=0.019).

Once again, findings in [9] differed from those presented here. [9] reported the same amount of on-task behavior in both Aplusix and The Incredible Machine, and less gaming behaviors in Aplusix. [9] and this study agree, though, that more conversation took place in Aplusix than in the game being used as the point of comparison.

6.3 Co-occurrence of behaviors and affective states

To determine the extent to which behaviors co-occurred with affective states, we first computed for the percentage of time a student engaged in a behavior while exhibiting a particular emotion, e.g. the percentage of time a student was on task while bored, giving and receiving answers while bored, and so on. We then compared that percentage with the percentage of time these behaviors were observed overall. We also compared the percentages of affective-and-behavior pairs in Aplusix with those in Math Blaster. We focused the analysis on the four most common affective states (boredom, confusion, delight, and flow). In Tables 3, 4, and 5, we report instances in which co-occurring affective states and behaviors were several times higher than normal.

Students who were bored while using Aplusix (Table 3) were half as likely to give or receive answers. They were more than twice as likely to engage in off-task conversation, eight times as likely to be inactive, and four times more likely to game the system.

When students were confused, they were not likely to be on-task. Rather, they were almost three times more likely to be giving and receiving answers, six times more likely to be inactive and three times more likely to be gaming.

Table 3. Aplusix student behavior choices in during specified affective states

Affective State	Behavior	Percent of time
Boredom	Giving and receiving ans.	8%
	Off-task conversation	8%
	Inactive	8%
	Gaming	6%
Confusion	On-task	23%
	Giving and receiving ans.	52%
	Inactive	6%
	Gaming	6%
Delight	On-task	34%
	Giving and receiving ans.	52%
	Other on-task conversation	2%
	Off-task conversation	24%
	Off-task solitary	7%
Flow	On-task	85%
	Off-task conversation	1%
	Inactive	0%
	Gaming	0%

Delight tended to be coupled with giving and receiving answers and off-task conversation.

Finally, flow tended to be coupled with working. Students in flow were also not likely to engage in off-task conversations, inactivity or gaming the system.

In the Math Blaster environment (Table 4), confusion generally reduced on-task behavior. Students who were confused were almost twice as likely to give and receive answers, 12 times more likely to engage in solitary off-task behavior and 24 times more likely to be inactive. On-task behavior also dropped as a result of delight. Students who were delighted were likely to give and receive answers or else engage in off-task conversation. Finally, students in flow tended to be on-task. They were unlikely to engage in any kind of conversation or off-task behavior.

When we compared the affect-behavior frequencies within Aplusix against those of Math Blaster (Table 5), we found that boredom is more likely to be associated with on-task behavior in Math Blaster than in Aplusix. Boredom in Aplusix is more likely to lead to off-task conversation, inactivity or gaming the system than in Math Blaster. Within Math Blaster, confusion led to more off-task solitary behavior while delight and flow lead to more on-task behavior than in Aplusix. Flow in Aplusix tended to be associated with more giving and receiving answers that in Math Blaster.

Table 4. Math Blaster student behavior choices during specified affective states; all results are significant at the 0.05 level or less

Affective State	Behavior	Percent of time
Confusion	On-task	29%
	Giving and receiving ans.	12%
	Off-task solitary	12%
	Inactive	24%
Delight	On-task	56%
	Giving and receiving ans.	26%
	Off-task conversation	13%
Flow	On-task	96%
	Giving and receiving ans.	0%
	Off-task conversation	0%
	Off-task solitary	0%
	Inactive	0%

Table 5. Comparison of usage choices in both software packages, during specific affective states; all results are significant at the 0.05 level or less

Affective State	Behavior	Greater in which software
Boredom	On-task	Math Blaster
	Off-task conversation	Aplusix
	Inactive	Aplusix
	Gaming	Aplusix
Confusion	Off-task solitary	Math Blaster
Delight	On-task	Math Blaster
Flow	On-task	Math Blaster
	Giving and receiving answers	Aplusix

7. DISCUSSION AND CONCLUSIONS

The results presented in this study were inconclusive about the effect of game formats on achievement. We suspect that the students whom we studied had already mastered the content that we presented, so neither software package was able to contribute to their learning.

Furthermore, some students complained that they were "too old" for the game's format. We had based our selection of the game and the content on information we received about the math content that the students were currently studying. In retrospect, that may we could have been rigorous in our evaluation of both the target audience's prior knowledge and maturity level.

Despite this, we did still find significant differences in affective states and behaviors between two groups. The students using the game tended to exhibit more delight. Consistent with prior findings published in [9], they were also exhibited more boredom and less flow.

Students using the game spent more working on the software whereas those using the intelligent tutor tended to converse and game the system more.

We found that certain affective states tend to be associated with certain behaviors. In Aplusix, as in [10], that gaming the system tended to co-occur with boredom and confusion. In both Aplusix and Math Blaster, confusion and delight tended to dampen on-task behavior but foster on-task conversation. Also in both cases, flow is associated with on-task behavior and is not associated with off-task behaviors.

Findings from this study suggest that, regardless of software format, boredom is an undesirable affective state. Flow is clearly desirable. Confusion and delight are more ambiguous. On the one hand, they do reduce solitary on-task behavior, however they seem to foster greater communication with teachers or classmates. The findings also provide some evidence that, regardless of the affective states they evoke, game formats tend to keep students on-task.

8. ACKNOWLEDGMENTS

We thank Jean-Francois Nicaud of the Laboratoire d'Informatique de Grenoble for the use of Aplusix. We thank the Ateneo de Manila University's Department of Information Systems and computer Science, 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. We the 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**. Finally, Dr. Rodrigo thanks the Philippine American Educational Foundation and the Council for International Exchange of Scholars for her 2008-2009 Advanced Research and University Lecturing Fulbright Scholarship.

9. REFERENCES

[1] Amory, A., Naicker, K., Vincent, J. and Adams, C. "The Use of Computer Games as an Educational Tool: Identification of Appropriate Game Types and Game Elements," *British Journal of Educational Technology*, 30(4), 311-321 (October 1999).

[2] Baker, R.S., Corbett, A.T., Koedinger, K.R., and Wagner, A.Z., "Off-task behavior in the Cognitive Tutor classroom: When students 'Game The System'," *ACM CHI 2004: Computer-Human Interaction*, 383-390 (2004).

[3] Bragg, L., "Students' Conflicting Attitudes Towards Games as a Vehicle for Learning Mathematics: A Methodological Dilemma," Mathematics Education Research Journal, 19(1), 29-44 (2007).

[4] Csikszentmihalyi, M, *Flow: The Psychology of Optimal Experience*, New York: Harper and Row (1990).

[5] De Aguilera, M. and Mendiz, A. 2003. "Video Games and Education: (Education in the Face of a 'Parallel School')". *ACM Computers and Entertainment,* 1(1), 1-14 (Oct. 2003).

[6] Lee, J., Luchini, K., Michael, B., Norris, C., and Soloway, E. "More Than Just Fun and Games: Assessing the Value of Educational Video Games in the Classroom," In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (Vienna, Austria, April 24 - 29, 2004). CHI '04. ACM, New York, NY, 1375-1378 (2004).

[7] Lepper, M. R. and Cordova, D. I., "A Desire to be Taught: Instructional Consequences of Intrinsic Motivation," *Motivation and Emotion*, 16(3), 187-208 (1992).

[8] Nicaud, J. F., Bouhineau, D., and Chaachoua, H., "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 (2004).

[9] 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., "Comparing Learners' Affect While Using an Intelligent Tutoring System and a Simulation Problem Solving Game," In Woolf, B. P., Aimeur, E., Nkambou, R. & Lajoie, S. P. (Eds.) Intelligent Tutoring Systems, 9th International Conference, ITS 2008, Montreal, Canada, June 23-27, 2008, Proceedings. *Lecture Notes in Computer Science 5091*, Springer, 40-49 (2008).

[10] 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., and Viehland, N. J. B. "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*, 145-152 (2007).

[11] Sierra Online, Inc, *The Incredible Machine: Even More Contraptions* (2001).

[12] Vogel, J. J. Greenwood-Ericksen, A., Cannon-Bowers, J., and Bowers, C. A., "Using Virtual Reality with and without Gaming Attributes for Academic Achievement," Journal of Research on Technology in Education, 39(1), 105-118 (Fall 2006).

10. AUTHORS' INSTITUTIONAL AFFILIATIONS

[1]Department of Information Systems and Computer Science, Ateneo de Manila University, Loyola Heights, Quezon City, Philippines, +63 (2) 426-6071

[2]Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA, +1 (412) 268-9690

[3]Education Department, Ateneo de Manila University, Loyola Heights, Quezon City, Philippines, +63 (2) 426-6001 loc 5230

[5]Ateneo de Manila University Grade School, Loyola Heights, Quezon City, Metro Manila, Philippines, +63 (2) 426-6001 loc 7100