Gathering Behaviors and Affective States of Learners Using Educational Software: Challenges to Research in an Emerging Economy

Ma. Mercedes T. RODRIGO[a], Jessica O. SUGAY[a], Jenilyn AGAPITO[b], and Salvador REYES Jr.[a]
[a]*Department of Information Systems and Computer Science, Ateneo de Manila University, Philippines*
[b]*Department of Computer Science, Ateneo de Naga University, Philippines*
mrodrigo@ateneo.edu

Abstract: Much of the research currently undertaken in the area of intelligent tutoring systems hails from the Western countries. To counteract any bias that this situation produces, to gain greater representation from the rest of the world, and to produce systems and publications that take cultural factors in to account, experts recognize the need for more intercultural evaluations and collaborations. To inform researchers planning field studies in developing countries, this paper discusses five challenges that researchers must address: technology adoption, school support, infrastructure, student culture, and force majeur.

Keywords: Intelligent tutoring systems, field study, research methods, developing countries

Introduction

Since 2006, the Ateneo Laboratory for the Learning Sciences (ALLS) has studied the behavior and affective states of learners using educational software through human observations and computer interaction logs. The purpose of these studies varied from determining which affective states occurred with which behaviors [15], common transitions between affective states [1], and models of student affect and behavior [12].

All of the data gathering for these studies took place in the Philippines, an emerging economy in Southeast Asia. The circumstances under which the research was conducted posed many challenges that, in some cases, threatened portions of the work, if not overall feasibility of the projects. The purpose of this paper is to document and discuss some of these challenges, their implications on the research (if any), and the ways in which the research team worked around them. These challenges vary from students' familiarity with technology, the availability of technological infrastructure, and extreme weather conditions.

1. Study Descriptions

The goal of ALLS is to derive new insights about how learners learn best through a quantitative analysis of student-computer interaction data. We typically ask students to use a computer-based learning system for 40 to 80 minutes. During the interactions, a team of observers notes student behavior and affective states. In many cases, software also logs student-computer interactions. Table 1 summarizes the characteristics of our studies.

Table 1. Summary of Different Studies

Dimension	Factor	Aplusix	Ecolab	M-Ecolab	Scatterplot Tutor (Control Group)	Scatterplot Tutor (with Agent)	The Incredible Machine	Math Blaster 9-12	BlueJ
User Characteristics	Number of Participants	140	90	90	59	68	36	30	146
	Age	12-15	9-13	9-13	12-14	12-14	14-19	12-13	17-20
	Gender	83F + 57M	29F + 61M	47F + 43M	35F + 24M	42F + 26M	17F + 19M	30M	25F + 121M
Environment Characteristics	School Location	4 urban; 1 rural	1 urban; 1 rural	1 urban; 1 rural	1 urban	1 urban	1 urban	1 urban	1 urban
	School Ownership	Private	Private	Private	Public	Public	Private	Private	Private
Learning Interaction Characteristics	Domain	Algebra	Ecology	Ecology	Graphing	Graphing	Logic puzzles	Pre-algebra	Java
	System Type	Computer Tutor	Computer Tutor	Computer Tutor	Cognitive Tutor	Cognitive Tutor	Serious game	Serious game	Integrated Development Environment
	Interaction time (mins)	45	40	40	80	80	10	40	50
Methodological Characteristics	Sampling Rate (secs)	200	200	200	200	200	60	200	200
	Kappa for	0.63	0.73	0.71	0.68		0.63	0.77	0.65
	Kappa for behavior	0.78	0.75	0.77	0.74		0.71	0.59	0.75

1.1 Descriptions of the Learning Environments

Data was gathered from different sets of students using nine learning environments that ranged from intelligent tutors to serious games to integrated development environments (IDE): Aplusix, Ecolab, M-Ecolab, the Scatterplot Tutor with and without a pedagogical agent, SimStudent, the Incredible Machine, Math Blaster 9-11, and BlueJ, an IDE Java.

Aplusix: Algebra Assistant [9][10] is an intelligent tutor for pre-algebra and algebra. It provides students with drill and practice with feedback in six content areas: numerical calculation, expansion and simplification, factorization, solving equations, solving inequations, and solving systems. Ecolab and M-Ecolab are ecology tutors that assist primary school children in learning about food chains and food webs. In terms of content, the two software packages are exactly the same. The difference between the systems is that M-Ecolab provides learners with motivational support in the form of a more-able virtual partner named Paul. A more detailed description of M-Ecolab's motivational support is provided in [12]. The Scatterplot tutor teaches students how to create and interpret scatterplots of data. Baker et al. developed a second version of the Scatterplot Tutor with a pedagogical agent, "Scooter the Tutor", designed to mitigate gaming the system (defined as misusing system features to progress through the curriculum without learning [2]). The Incredible Machine: Even More Contraptions [19] is a simulation environment where students complete a series of logical puzzles. The student must combine given objects in a creative fashion to accomplish each puzzle's goal.

Math Blaster 9-12 [5] 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 escape the planet, 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. Finally, BlueJ is an integrated development environment for Java. The environment allows students to edit a Java program, compile it, and test it. With the exception of The Incredible Machine and Math Blaster 9-12, all the systems logged student-system data for later analysis. The version of BlueJ used for our studies was instrumented to be able to store every student submission to the compiler on a central server. All other systems logged the data locally.

1.2 Schools

The schools under study were principally located in urban schools. Only one was located in a rural area. Most were privately owned. Only the school in which we deployed the Scatterplot tutor study was public. With the exception of the Scatterplot study, all data was gathered in the school computer laboratory. The Scatterplot study was conducted in the computer room of the school's library.

1.3 Populations

The populations under each study varied from grade school students to college students. Aplusix, The Incredible Machine, and Scatterplot data [8][15][18] were gathered from first year high school students. Ecolab/M-Ecolab participants [16] were in the 4th grade. MathBlaster participants [17] were from the 7th grade. Finally, the BlueJ data [7] was collected from novice programmers in the first and second years of college.

1.4 Data Collection Methods

We used a uniform observation protocol to record the behaviors and affective states of students using these different learning environments. Pairs of observers--Masters students in Education or Computer Science--conducted the observations. Most of our observers had teaching experience. All had been trained for the task through a series of pre-observation discussions on the meaning of the categories they were coding and through a pilot observation exercise conducted at a different school.

The observers coded a set of affective categories drawn from [6]: boredom, confusion, delight, engaged concentration (a subset of flow [4]), frustration, surprise and neutral. They also coded seven behaviors from [2]: on-task, on-task conversation, off-task, off-task conversation, off-task solitary, inactive, and gaming the system. The observers attempted to conduct observations in a fashion that did not make students aware that they were being observed at a given moment. To this end, students were observed through quick glances, through using peripheral vision, or by pretending they were looking at another student, so as to minimize the effects of the observations. Each pair of observers was assigned to a group of 10 students. Observers rotated among students in a pre-determined order, and conducted all observations in synchrony. Each observation lasted twenty seconds. If two distinct behaviors or affective states were seen during an observation, only the first behavior or affective state observed was coded.

Video recording and playback has been considered as a less intrusive method of gathering data on behavior and affect, so that observers would not be physically in the same room as the students. However, the need to acquire and set up recording equipment presents additional challenges.

2. Challenges

We identify five dimensions by which the challenges of each study can be characterized: technology adoption, school support, infrastructure, student culture, and force majeur.

2.1 Technology Adoption

Over the last three decades, the Philippines has seen an increasing investment in information and communication technologies for education. Both public and private schools have been acquiring or receiving computer hardware and software and Internet connectivity to support teaching and learning. However, usage of these technologies is often limited to the teaching of information technology literacy [13]. Hence, few schools use learning software as a regular part of their curriculum. The rare to non-existent use of sophisticated learning technologies such as intelligent tutors, serious games, simulations, or microworlds confounds cross-cultural comparisons in that analogous populations from a developed and developing country might be impossible to find.

At the student level, the level of technology adoption was an issue as well. Some students, particularly those in the public schools, had little experience in operating computers. The researchers had to provide them with assistance until they were comfortable in using the computers. Some of the student participants were unaccustomed to using laptop computers in particular. As a result, they would seek for assistance when they got perplexed with how to work with the machines. This may have had effects on the manner with which they worked with the software; however, this issue was not further explored.

2.2 School Support

For any school-based study to prosper, particularly one in which tens of students are involved, it must have approval from the school's administration, teachers, and the students' parents[1]. We typically had to send formal letters of request to the school

[1] The age of consent in the Philippines is 18. As most participants were not of legal age, parental consent was required in order for them to participate in the study.

principal, asking for permission to come on to the premises, use the school's computers, and interact with the students. The public school principal whose school participated in the Scatterplot study was quick to give us her permission. This experience was consistent with the first author's prior experience of data gathering for a previous study on computer availability and usage in Metro Manila schools [13].

Obtaining permission from private schools was much more challenging. Some school officials ignored our requests outright. In one case, the request was referred to a subject area coordinator. The coordinator asked us to explain the purpose of our study. When we attempted to explain that the long-term vision of the study was to develop models of human emotion, she literally laughed. We were not able to gather any data from that school.

2.3 Infrastructure

For these studies, we needed 10 computers installed with some version of Windows. With the exception of the BlueJ study, none of the learning environments required network access. Although public schools are the quickest to give us their permission, they are also the least adequately equipped. The public school for the Scatterplot tutor study had a computer laboratory; however, we were not allowed to use it because it was occupied throughout the day. The computer coordinator helped us assemble a suite of 12 computers. Most had been in storage, others were computers from the library. All were in disuse. When we visited the school to install the software on to the computers, we found that some did not have functioning CD drives or USB drives. Many of the mice were not working. We had to replace them with new ones. Windows versions ranged from 3.11 to XP.

During the dry run, the computers stalled or automatically turned off one after the other. At the end of 80 minutes, we were down to 4 computers. The computer coordinator asked us for a few days to "see what he could do". Of the 12 computers that the school assigned to the study, 8 were restored to a usable state. The remaining computers were beyond repair as the graphic cards of the computers used legacy ports (AGP), which meant finding replacements would be difficult or very expensive. The research team brought in two extra laptop computers to complete the set. The team also had to bring in a projector and speakers as these were needed for the introductory lecture. These initiated extra tasks for the research team as the laptops, the projector and a make-shift projector screen made of paper had to be set up before every session commenced. Extracting data proved to be difficult as well. Data transfer was bottlenecked by the speed of the old processors, the capacity of the memory modules, and the read speed of the hard disks.

The private schools generally had computer laboratories that we could use. Computers were uniformly configured. Hardware and software were reasonably up-to-date. There was one exception, though. One private school could only spare their mathematics laboratory, which was full of their oldest computers. After the study, we discovered that one computer's USB ports and disk drive were not working—there was no way to copy our data. We had to remove the computer's hard disk and install it on another machine.

Finally, the room / laboratory layouts varied from school to school. In some cases, the computers were arranged in classrooms-style, with all computers in rows and all students facing a common blackboard. There were other cases in which computers were arranged facing the wall, along the perimeter of the room. This posed a challenge to the observers because it was difficult to find a spot that could give them a good view of the facial expressions of the students. On the other hand, students' body language and computer activity were easily observable.

While the use of video recording might minimize the challenges to observers in space and positioning, the lack of available recording equipment and the need to install them for each individual student in each study posed even greater infrastructure challenges.

2.4 Student Culture

To minimize the effect of the observations on student behavior, quantitative field observations involve the use of peripheral vision and observation at a distance. One of the more difficult behaviors to code or note, from the observers' experiences, is on-task conversation. Because observers have to distance themselves from the student under observation during a given time, it is difficult to listen to what the student might be saying when in conversation. To hear what is said in conversation, observers would casually walk by the student, in the hope that something audible would be heard.

In most of the studies, students behaved as naturally as could be expected. Students in the grade school were particularly more expressive. They smiled, laughed, frowned, furrowed their eyebrows, exclaimed aloud, yawned, stretched, pulled at their hair, and even put their feet up when they felt like doing so. However, there were instances when students tried to behave at their best when they felt they were being observed. They would stop conversing with their seatmates, sit up straight, and focus on the work they were doing. But after some time, they forgot they were being watched and would act more naturally. One of the exceptions to the generally natural student behavior was observed in an all-girls private high school. The students were extremely well-behaved. They all sat upright and faced their computer screens during the entire session. They hardly spoke with one another and addressed any questions to the teacher. They resisted showing any facial expressions or making any gestures. They seemed uptight and rigid, which the observers attributed to the culture of discipline that this particular school implements.

2.5 Force Majeur

The Philippine school year begins in June. This coincides with the start of the wet season, the time of year marked with monsoon rains and typhoons. In the middle of the Scatterplot study, the Philippines was hit by Tropical Storm Ondoy (international name Typhoon Ketsana), the most devastating typhoon to hit Metro Manila since 1970. Typhoon Ondoy brought down 341.3 millimeters or one month’s worth of rainfall in six hours, resulting in landslides and floods ranging from two to six feet high. In the aftermath of Ondoy, official reports indicated 464 deaths and damages to public and private property amounting to USD 237 million. Because of the rain and the flooding, classes (and data collection) were suspended during the typhoon and up to two weeks thereafter.

The typhoon introduced a possible confound to the study: post-traumatic stress. As we were studying student affect, we wondered if the devastation might have had an impact on students’ overall mood, disposition, and motivations—and hence the findings from the study. Fortunately, when data collection resumed, the post-Ondoy user groups did not behave differently from the pre-Ondoy groups, though admittedly this data was not statistically studied.

3. Discussion

Much of the research currently undertaken in the area of intelligent tutoring systems hails from the Western countries [3][11]. This results in an unintentional bias—both empirical studies and theory are founded on a Western mindset. To counteract this bias, to gain greater representation from the rest of the world, and to produce systems and publications that take cultural factors in to account [11], experts recognized the need for more intercultural evaluations and collaborations [3]. This paper's authors are eager to participate in these collaborations to further explore intercultural comparisons. However, we flag the unique difficulties that come with deploying these systems and conducting field studies in developing countries.

Limited technology adoption has an impact on the fluency with which students can interact with the software as well as on the direct comparability of a developing world sample against one from a developed country. School support is essential but sometimes difficult to acquire. The availability of hardware and software dictates whether a school is a feasible test site or not. Indeed, in our experience, there seemed to be a tradeoff between schools support and infrastructure: the schools with the better infrastructure often refused or ignored our requests, whereas the schools with limited infrastructure were much more hospitable. Whether or not the phenomenon of interest is observable at all is dictated in part by student culture. Finally, natural and man-made calamities can have serious impact on school schedules and, consequently, data collection. This paper is not meant to discourage these collaborations. Rather it is meant to inform researchers, to help them plan their field studies with a broader range of considerations.

Acknowledgements

The research projects discussed in this paper were made possible by the Ateneo de Manila University, the Ateneo's Department of Information Systems and Computer Science, the Ateneo Center for Educational Development, the Philippines Department of Science and Technology, and the participating schools, students, and research staff. We thank them all for their cooperation and support.

References

[1] Baker, R. S. d. J., Rodrigo, M. M. T. & Xolocotzin, U. E. (2007). The dynamics of affective transitions in simulation problem-solving environments. In A. Paiva, R. Prada & R. W. Picard (Eds.): *ACII 2007, LNCS 4738*, 666-677. Germany: Springer-Verlag Berlin Heidelberg.

[2] 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.

[3] Blanchard, E. (2012). On the WEIRD nature of ITS/AIED conferences: A 10 year longitudinal study analyzing potential cultural biases. In S. Cerri et al. (Eds.: ITS 2012, Lecture Notes in Computer Science 7315, pp. 280-285. Springer-Verlag: Berlin Heidelberg

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

[5] Davidson and Associates. (1997). *Math Blaster 9-12* [Computer Software].

[6] D'Mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., & Graesser, A. (2005). Integrating affect sensors in an intelligent tutoring system. *Proceedings of Affective Interactions: The Computer in the Affective Loop Workshop*, 7-13.

[7] Dy, T. & Rodrigo, M. M. (2010). A detector for non-literal Java errors. In *Proceedings of the 10th Koli Calling International Conference on Computing Education Research (Koli Calling '10)*. ACM, New York, NY, USA, 118-122. DOI=10.1145/1930464.1930485.

[8] Lagud, M. C. V. & Rodrigo, M. M. T. (2010). The affective and learning profiles of students while using an intelligent tutoring system for algebra. In V. Aleven, J. Kay, & J. Mostow (Eds.) *ITS 2010 Part I, Lecture Notes in Computer Science 6094*, 255-263. Berlin Heidelberg: Springer Verlag.

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

[10] Nicaud, J. F., Bouhineau, D., Mezerette, S., Andre, N. (2007). Aplusix II [Computer software].
[11] Ogan, A., Walker, E., Baker, R.S.J.d., de Carvalho, A., Laurentino, T., Rebolledo-Mendez, G., Castro, M.J. (2012) Collaboration in Cognitive Tutor Use in Latin America: Field Study and Design Recommendations. *Proceedings of ACM SIGCHI: Computer-Human Interaction*, 1381-1390.
[12] Rebolledo-Mendez, G., du Boulay, B., & Luckin, R. (2006) Motivating the learner: an empirical evaluation. In M. Ikeda, K. Ashley, & T.-W. Chan (Eds.) *ITS 2006, Lecture Notes in Computer Science 4053*, 545-554, Berlin Heidelberg: Springer-Verlag.
[13] Rodrigo, M. M. T. & Baker, R. S. J. d. (2009). Coarse-grained detection of student frustration in an introductory programming course. *International Computer Education Research Workshop* ICER 2009.
[14] Rodrigo, M. M. T. (2005). Quantifying the divide: A comparison of ICT usage of schools in Metro Manila and IEA-surveyed countries. *International Journal for Educational Development*, *25*, 53-68
[15] 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*, 145-152. Amsterdam, The Netherlands: IOS Press.
[16] Rodrigo, M. M. T., Rebolledo-Mendez, G., Baker, R. S. J. d., du Boulay, B., Sugay, J. O., Lim, S. A. L., Lahoz, M. B. E., Luckin, R. (2008). The effects of motivational modeling on affect in an intelligent tutoring system. In Chan, T.-W., Biswas, G., Chen, F.-C., Chen, S., Chou, C., Jacobson, M., Kinshuk, Klett, F., Looi, C.-K., Mitrovic, T., Mizoguchi, R., Nakabayashi, K., Reimann, P., Suthers, D., Yang, S., & Yang, J.-C. (Eds.). *International Conference on Computers in Education* 2008, 49-56.
[17] Rodrigo, M.M.T. & Baker, R.S.J.d. (2011) Comparing Learners' Affect While Using an Intelligent Tutor and an Educational Game. *Research and Practice in Technology Enhanced Learning, 6* (1), 43-66.
[18] San Pedro, M. O. C. Z., Baker, R. S. J D., & Rodrigo, M. M. T. (2011). Detecting carelessness through contextual estimation of slip probabilities among students using an intelligent tutor for mathematics. 2011 Artificial Intelligence in Education Conference. 28 June to 1 July 2011. University of Auckland, Auckland, New Zealand. In G. Biswas, S. Bull, J. Kay, & A Mitrovic. Artificial Intelligence in Education. Lecture notes in Computer Science Volume 6738. Springer Berlin/Heidelberg. 304-311.
[19] Sierra Online. 2001. The Incredible Machine: Even More Contraptions [Computer Software].