Much of the research currently undertaken in the area of intelligent tutoring systems hails from
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 into
account, experts recognize the need for more intercultural evaluations and collaborations. For these
collaborations to be successful, though, methods and materials require modification. Field work
methodologies used in developed countries have to be nuanced when transferred to developing
world contexts. In specific, the paper describes five challenges that researchers must address in the
transfer process: technology adoption, school support, infrastructure, student culture, and force
majeure.

Keywords: Intelligent tutoring systems; field study; research methods; developing countries.

1. Introduction

Researchers interested in designing educational software for the developing world or
studying educational software’s use and effects in emerging economies have found that
both content and methods need to be adjusted for the products and processes to be
effective. In designing mobile educational games for rural children in India, the group of
Kam (2009) found that these children did not understand many of the game patterns commonly found in Western games. For example, they generally did not pay attention to scores. They were unfamiliar with renewable, hidden, or discoverable game resources. Finally, they were used to playing fields that were visible and constant throughout the game—they were unfamiliar with scrolling screens. Hence, any educational game intended for a rural Indian population has to be designed subject to these constraints.

Note that educational software development is an expensive and time-consuming undertaking. Those educators and researchers who do not have the resources to invest in software development work but who wish to conduct studies on the use and effects of technology in classrooms have to make use of existing software. The caveat is that much of existing educational software, particularly intelligent tutoring systems (ITSs), are developed in the West (Blanchard, 2012; Ogan et al., 2012).

Using existing educational software designed for developed countries and studying their adoption in the classroom presents its own challenges. The paper of Ogan et al. (2012) described the experience of transferring an intelligent tutoring system (ITS) developed in the United States to Brazil, Mexico, and Costa Rica. They observed, while the ITS was designed for individual use, students from Latin America tended to work collaboratively. This may be attributable to the strong collectivist nature of these societies. They also observed that students lacked basic math skills and computer experience than their US counterparts. This implies that the deployment of the ITS should most probably emphasize subject matter fluency, enrichment, or remediation rather than the development of advanced skills. Ogan et al. (2012) point out that these fundamental differences between the context in which educational software is designed and studied (e.g. the US) and the context in which it is deployed (e.g. Latin America) should lead to variations in instructional methodologies and software usage, so that the benefits of these systems can be maximized.

Indeed, educational technology’s success is determined in large part by its alignment with local contexts. Hoadley, Honwad, and Tamminga (2010) suggest alignment along four dimensions: infrastructure, expertise, pedagogical strategy, and values. Infrastructure refers to the selection of technologies that can operate reasonably in the environmental conditions (temperature, dust, and availability of electricity) of the target school. Expertise refers to the schools’ professional capacity to operate and maintain these technologies. Schools adopting technology also have to harmonize technology choices and pedagogical strategies. For instance, a school that favors lecture-based methods might benefit more from a single computer and projector rather than stand-alone computers meant for individual use. Finally, values refer to the ultimate goal or purpose that technology is intended to serve. Hoadley et al. (2010) emphasize that technology has a greater chance of making an impact if technology choices and usage serve community values (e.g. to support communication between and within groups) rather than values imposed by external agencies (e.g. the development of IT-related work skills). Consistency with local values dictates the extent to which the community will embrace the technology and take ownership of its usage.
Since 2006, the Ateneo Laboratory for the Learning Sciences (ALLS) has taken an active role in fostering these types of intercultural collaborations. In its studies of learners’ behaviors and affective states, the research team made use of western field research methods described in Baker, Rodrigo, and Xolocotzin (2007) and learning systems described in Baker, Corbett, Koedinger, and Wagner (2004), Davidson and Associates (1997), Nicaud, Bouhineau, and Chaachoua (2007), Rebolledo-Mendez, du Boulay, and Luckin (2006), Sierra Online (2001), and University of Kent (no date). The purpose of these studies varied from determining which affective states occurred with which behaviors (Rodrigo et al., 2007), common transitions between affective states (Baker et al., 2007), and models of student affect and behavior (Rodrigo & Baker, 2009). All of the data gathering for these studies took place in the Philippines, an emerging economy in Southeast Asia. ALLS chose to study ITSs and related systems because they have proven effective in developed countries. For these to be beneficial to developing countries, however, could they be adopted and studied without modification or will the transfer of these technologies and their related studies have to be adjusted to take local contexts into account?

In transferring existing educational software and study methods to the Philippine context, the ALLS research team found it necessary to make methodological changes. The team found that these systems, how they are used, and the methods by which their usefulness is assessed have to be adapted to the local human and social context in which they will be operated (Ogan et al., 2012). 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. Among many other factors, effective transfer of these technologies and evaluation methodologies must take into account the socio-economics, digital literacy, and cultural pre-dispositions of the target audience.

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 to transfer the methods effectively. The paper first describes the past studies then moves on to discuss the challenges that the research team encountered and the ways in which the team coped with them.

2. 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. To this end, we typically asked students to use a computer-based learning system for 40 to 80 minutes. During the interactions, a team of observers noted student behavior and affective states. In many cases, software also logged student-computer interactions. Table 1 summarizes the characteristics of our studies.

2.1. 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
Table 1. Summary of different studies.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Factor</th>
<th>Aplusix</th>
<th>Ecolab</th>
<th>M-Ecolab</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User Characteristics</strong></td>
<td>Number of Participants</td>
<td>140</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>12-15</td>
<td>9-13</td>
<td>9-13</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>83F + 57M</td>
<td>29F+61M</td>
<td>47F+43M</td>
</tr>
<tr>
<td><strong>Environment Characteristics</strong></td>
<td>School Location</td>
<td>4 urban; 1 rural</td>
<td>1 urban; 1 rural</td>
<td>1 urban; 1 rural</td>
</tr>
<tr>
<td></td>
<td>School Ownership</td>
<td>Private</td>
<td>Private</td>
<td>Private</td>
</tr>
<tr>
<td><strong>Learning Interaction Characteristics</strong></td>
<td>Domain</td>
<td>Algebra</td>
<td>Ecology</td>
<td>Ecology</td>
</tr>
<tr>
<td></td>
<td>System Type</td>
<td>Computer tutor</td>
<td>Computer Tutor</td>
<td>Computer Tutor</td>
</tr>
<tr>
<td><strong>Methodological Characteristics</strong></td>
<td>Interaction time (mins)</td>
<td>45</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Sampling Rate (secs)</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Kappa for affect</td>
<td>.63</td>
<td>.73</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>Kappa for behavior</td>
<td>.78</td>
<td>.75</td>
<td>.77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Factor</th>
<th>Scatterplot Tutor (Control Group)</th>
<th>Scatterplot Tutor (with Agent)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User Characteristics</strong></td>
<td>Number of Participants</td>
<td>59</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>12-14</td>
<td>12-14</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>35F+24M</td>
<td>42F+26M</td>
</tr>
<tr>
<td><strong>Environment Characteristics</strong></td>
<td>School Location</td>
<td>1 urban</td>
<td>1 urban</td>
</tr>
<tr>
<td></td>
<td>School Ownership</td>
<td>Public</td>
<td>Public</td>
</tr>
<tr>
<td><strong>System Characteristics</strong></td>
<td>Domain</td>
<td>Graphing Cognitive Tutor</td>
<td>Graphing Cognitive Tutor</td>
</tr>
<tr>
<td></td>
<td>System Type</td>
<td>Cognitive Tutor</td>
<td>Cognitive Tutor</td>
</tr>
<tr>
<td><strong>Methodological Characteristics</strong></td>
<td>Interaction time (mins)</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Sampling Rate (secs)</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Kappa for affect</td>
<td>.68</td>
<td>.74</td>
</tr>
<tr>
<td></td>
<td>Kappa for behavior</td>
<td>.74</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Factor</th>
<th>The Incredible Machine</th>
<th>Math Blaster</th>
<th>BlueJ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User Characteristics</strong></td>
<td>Number of Participants</td>
<td>36</td>
<td>30</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>14-19</td>
<td>12-13</td>
<td>17-20</td>
</tr>
<tr>
<td></td>
<td>Gender</td>
<td>17F + 19M</td>
<td>30M</td>
<td>25F+121M</td>
</tr>
<tr>
<td><strong>Environment Characteristics</strong></td>
<td>School Location</td>
<td>1 urban</td>
<td>1 urban</td>
<td></td>
</tr>
<tr>
<td></td>
<td>School Ownership</td>
<td>Private</td>
<td>Private</td>
<td>Private</td>
</tr>
<tr>
<td><strong>Learning Interaction Characteristics</strong></td>
<td>Domain</td>
<td>Logic puzzles</td>
<td>Pre-algebra</td>
<td>Java</td>
</tr>
<tr>
<td></td>
<td>System Type</td>
<td>Serious game</td>
<td>Serious game</td>
<td>Integrated Development Environment</td>
</tr>
<tr>
<td><strong>Methodological Characteristics</strong></td>
<td>Interaction time (mins)</td>
<td>10</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Sampling Rate (secs)</td>
<td>60</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>Kappa for affect</td>
<td>.63</td>
<td>.77</td>
<td>.65</td>
</tr>
<tr>
<td></td>
<td>Kappa for behavior</td>
<td>.71</td>
<td>.59</td>
<td>.75</td>
</tr>
</tbody>
</table>
(IDE): Aplusix, Ecolab, M-Ecolab, the Scatterplot Tutor with and without a pedagogical agent, the Incredible Machine, Math Blaster 9-11, and BlueJ, an IDE for Java.

Aplusix: Algebra Assistant (Nicaud et al., 2004; Nicaud, Bouhineau, Mezerette, & Andre, 2007) 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 Rebolledo-Mendez et al. (2006).

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 (Baker, Corbett, Koedinger, & Wagner, 2004).

The Incredible Machine: Even More Contraptions (Sierra Online, 2001) 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 (Davidson and Associates, 1997) 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 (University of Kent, no date) is an integrated development environment for Java that enabled students to edit, compile, debug, and run their programs.

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.

2.2. Schools

To find participants to include in these studies, we sent out letters of request and invitation to different schools. Those who responded willingly in our favor would then become part of the study. Schedules for data gathering were set at the convenience of each school.

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.

2.3. Populations

The populations under each study varied from grade school students to college students. The student participants were selected at random from the pool of students in each school. Selection of a target grade level depended on the nature and content of the learning environment to be used in the study.

Aplusix, The Incredible Machine, and Scatterplot data (Lagud & Rodrigo, 2010; Rodrigo et al., 2007; San Pedro, Baker, & Rodrigo, 2011) were gathered from first year high school students. Ecolab/M-Ecolab participants (Rodrigo et al., 2008) were in the 4th grade. MathBlaster participants (Rodrigo & Baker, 2011) were from the 7th grade. Finally, the BlueJ data (Dy & Rodrigo, 2010) was collected from novice programmers in the first to second year of college.

2.4. 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 D’Mello, Craig, Witherspoon, McDaniel, and Graesser (2005): boredom, confusion, delight, engaged concentration (a subset of flow; Csikszentmihalyi, 1990), frustration, surprise and neutral. They also coded seven behaviors from Baker et al. (2004): on-task, on-task conversation, off-task, 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. 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 20 seconds. If two distinct behaviors or affective states were seen during an observation, only the first behavior or affective state observed was coded. The findings from these observations are out of the scope of this paper but are detailed in Dy and Rodrigo (2010), Lagud and Rodrigo (2010), Rodrigo and Baker (2009, 2011), Rodrigo et al. (2007), Rodrigo et al. (2008) and San Pedro et al. (2011).
3. Challenges

In the process of collecting the data for these studies, we encountered many challenges. It is the purpose of this paper to describe these challenges according to five dimensions: technology adoption, school support, infrastructure, student culture, and force majeure. The sections below describe these challenges and the adaptations the research team made in order to transfer technologies and methods to the Philippine context.

3.1. 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 (Rodrigo, 2005). 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 a 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.

3.2. 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. We typically had to send formal letters of request to the school 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 (Rodrigo, 2005).

Obtaining permission from private schools was 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.

3.3. Infrastructure

For each of 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 were also the least adequately equipped. Of the 45,971 grade schools and high schools in the Philippines, 26,026 (57%) are equipped with computers in varying working conditions. Of the 7,470 high schools, 2,988 (40%) are connected to the Internet.

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 added to the tasks of 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, install it on another machine, extract the data, and return the hard disk.

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.

3.4. 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 (at least in the beginning) 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. After a few minutes, 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.

3.5. Force majeure
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. While a school year lasts for 40 weeks, ending in March, 25-27 (63%-68%) weeks are covered by the wet season, lasting from June to November. In the first 12 weeks of a school year, an average 10 days (2 weeks) are lost to the suspension of classes due to floods and inclement weather.

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.

4. Conclusions and Implications on International Transfer and Adaptation of Materials and Methodologies

When ITSs are deployed and studied in WEIRD countries, the research teams do encounter their own share of difficulties. Colleagues speak of test participants who opt out in the middle of the process or who use abusive language (Ryan Baker, personal communications). The data gathering methodologies anticipate and provide for these difficulties. The kinds of difficulties described in this paper, though, are not among those contingencies. Indeed, data gathering methods assume that computers are available, that the school is cooperative, and that there are no untoward interruptions in the process. We see from this paper that these assumptions are not always true when methods are transferred to the developing world.

This paper flags the unique difficulties that come with transferring these systems and field study methodologies to 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 a 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.

Acknowledgments

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


