Cross-Cultural Differences and Learning Technologies for the Developing World

Workshop Co-Chairs:

Ivon Arroyo
Worcester Polytechnic Institute, Worcester, MA, United States

Imran Zualkernan
American University of Sharjah, United Arab Emirates

Beverly P. Woolf
University of Massachusetts Amherst, Amherst, MA, United States

http://cadmium.cs.umass.edu/LT4D

Preface

The Learning Technologies for the developing world (LT4D) workshop aims to provide a forum for a discussion of corss-cultural differences and rational introduction of learning technologies in the developing world, exploring the economic constraints, socio-cultural differences, political and other constraints that shape the implementation and the affordances for learning technologies in the developing world.

Besides differences in socialization and cultural differences, well-intentioned introduction of learning technologies in developing countries can fail for mundane reasons such as teachers not willing to use the technology because of lack of comfort with technology, or simply lack of computers in sharp contrast to abundance of mobile devices.

Such constraints cannot be ignored. Rather than blindly implanting technologies, based on a rationalized discussion of such issue and constraints, and possibilities for the immersion of learning technologies, the workshop then aims to provide future visions and roadmaps of such technologies for the developing world and subsequent practical implementation for technology enhanced learning.

The workshop intends to touch on the following broad set of issues:

1. Cross-cultural differences in educational outcomes of AIED systems or non-adaptive learning technologies across countries, developing vs. developed, or across developing countries.
2. Ideas to solve issues of economic cost of adapting interactive learning environments (ILEs) to developing countries
3. Examples of Localization and Cultural translation of systems and interfaces
4. Issues of Social Inclusion: how to encourage and support both individuals and communities that are marginalized --economically, socially, or culturally; indigenous communities, and other special communities.
5. Science of sustainable design of learning technologies for the developing world; Sustainability of projects in the developing world; funding sources; ideas for maintaining technology resources and personnel
6. How education technology is used in the developing world; how is or should it be used? As a means? As an end?
7. Supporting Teacher Training via e-Learning in developing countries
8. How can Educational Data Mining help to support education and reveal information that would help developing countries
9. Differences in realities and possibilities of implementation of Interactive Learning Environments (ILEs) across the developing world? Is there a common ground, or are countries too different from each other?
10. Issues of timing: Are there key areas where learning technologies can have an immediate impact?
11. Models of adoption (or non-adoption) of learning technologies in the developing world

12. An analysis of great successes or drastic failures in applying ILEs to the developing world
13. Opportunities for leap frogging and avoiding mistakes in the developed world

July, 2013
Ivon Arroyo, Imran Zualkernan, Beverly Woolf

Program Committee

Co-Chair: Ivon Arroyo, *Worcester Polytechnic Institute, United States (iarroyo@wpi.edu)*
Co-Chair: Imran Zualkernan, *American University of Sharjah, United Arab Emirates* (izualkernan@aus.edu)
Co-Chair: Beverly P. Woolf, *University of Massachusetts Amherst, Amherst, MA* (bev@cs.umass.edu)

Fabio Akhras. *Renato Archer Center of Information Technology, Brazil*
Ryan Joazeiro de Baker. *Teachers College. Columbia University. USA.*
Paul Brna. *University of Leeds. United Kingdom.*
Melissa Sue John. *Worcester Polytechnic Institute, USA.*
Amy Ogan. *Carnegie Mellon University, USA*.
Ma. Mercedes T. Rodrigo. *Ateneo de Manila University, Phillipines.*
Rosa Maria Vicary. *Universidade Federal do Rio Grande do Sul, Brazil.*

Table of Contents

Towards Localization of Automated Tutors for Developing Countries

Imran Zualkernan[1], Ivon Arroyo[2], Beverly P. Woolf[3]

[1] American University of Sharjah, United Arab Emirates
[2] Social Sciences and Policy Studies, Worcester Polytechnic Institute
[2] School of Computer Science, University of Massachusetts Amherst

izualkernan@aus.edu, iarroyo@wpi.edu, bev@cs.umass.edu

Abstract. This paper describes localization issues in relation to AIED systems in the developing world, and analyzes the particular case of the successful immersion of learning technologies to schools in Pakistan. The paper analyzes the needs for personalized learning in the developing world in comparison to countries such as the United States. A model and a survey based on various types of localization dimensions like teacher, student, and cultural alignment was developed and deployed to conduct an evaluation of an AI tutor called the Wayang Outpost in Pakistan. The results are that teachers are likely to use such a system if available, and that their intention to use such a tutor in the future is strongly dependent on how well the tutor is aligned with their teaching practices, students' learning habits, and whether the language in the tutor is understood by students. On average these teachers were also willing to allocate about two hours per week for such automated tutors.

Keywords: Developing World, Adaptive Technologies, Localization

1. Introduction

A recent study [1] that used a self-contained traveling van to deliver Khan Academy (www.khanacademy.org) videos in conjunction with Android-based online assessments for children resulted in two interesting observations. First, a learning technology intervention for a seemingly culture-agnostic subject like grade 4 and 5 Mathematics required a significant effort in 'localization' that went far beyond language translation [2]. Second, the statistical effects observed between treatment and control groups were high when compared with those obtained using automated tutors in the West [1],[3]. Taken together, these two observations suggest that while there is a great potential for using automated tutors in developing countries, to be effective, these tutors may need to undergo extensive localization along a number of non-obvious dimensions.

Adaptive tutoring systems have been effective at improving students' achievement in a variety of tests, including standardized tests. For instance, the Wayang Outpost Mathematics Adaptive Tutoring System has shown improvements within 0.3-0.8 effect sizes on standardized tests compared classroom instruction, after controlling for time [4]. The Algebra Tutor has shown effect sizes of 0.3-1.2 standard deviation on a variety of tests [5]. Andes tutor for Physics has shown effect sizes of 0.92 [6]. This is particularly impressive considering that human tutors (subject-matter experts working synchronously with a single student) are one standard deviation better compared to a teacher in front of the classroom with a typical class size of 20 students. This is in contrast to what was originally believed in studies by Bloom [7], which claimed that a human tutor could be 2 standard deviations better than classroom instruction. Since the van study using Khan Academy cited earlier [1] showed effects of 0.87 to over 3, it is expected that the use of adaptive tutors in developing countries will yield much better results. However, as indicated earlier, when considering the implementation of tutoring systems in countries other than the United States, and particularly in the developing world, the important question that urges to be answered is whether and how much *localization* efforts are required to make these adaptive tutors be effective.

The remaining article shows that, while much language and cultural translation needs to be carried out, there is potential to achieve large effect sizes, and that there are specific needs of the developing world that make the use of adaptive learning environments particularly ideal for students with a large variety of unique circumstances. However, cultural differences need to be understood, as they can affect the ecological validity of the intervention, and the general effectiveness of the teaching tool.

2. Why Adaptive Tutors for Developing Countries?

This section motivates the need for automated tutors in developing countries.

2.1 Teachers and Students

Quality and availability of teachers is a key input into the educational quality of children. However, according to a recent study [8], 29 countries mostly in Arab or Sub-Saharan Africa regions have severe teacher gaps and need to grow annually by at least 3.0% during the 2010 to 2015 period. Even when the teachers are present, teacher absenteeism remains a problem, ranging from 3 percent in Malawi to 27 percent in Uganda [9]. [10] reports that teachers absenteeism in six developing countries was about 19%, with Peru at 11%, Indonesia at 19%, and in India at 25%. In a meta-study of developing country learning interventions from 1990 to 2010, out of 79 studies, 5 studies showed mostly negative impacts of teacher absenteeism [11].

There is also a wide variation (800-1000 hours) in the contact time between teacher and students in the developing world [12]. However, more contact time with teachers does increase performance in majority of cases [12]. There are also differences between various countries when it comes what a teacher actually does in the classroom. For instance, in Tunisia, Morocco, Brazil and Ghana, the students were engaged in learning 79%, 71%, 63% and 39% of the time respectively [13]. However, teachers in these countries mostly used “chalk-and-talk” which can result in limited student attention and subsequent recall [13]. Automated tutors have a potential to bring standardized learning processes to children and unlike teachers, tutors are less likely to suffer from absenteeism.

Pupil-teacher ratios are also much higher in the developing countries than the West [14]. For example, in US and UK, teacher-pupil ratios are 14 and 18 respectively. These ratios can be as high as 65 (Rwanda) or Zambia (68) in sub-Saharan Africa. South Asian countries like Pakistan and Bangladesh have pupil-teacher ratios of 40 and 43 respectively. However, a better pupil-teacher ratio does not necessarily guarantee better student performance. For example, [15] observed that reducing the pupil-teacher ratio alone from 88 to 40 did not have a significant impact on learning. However, contract teachers and a strong institutional PTA support did have a positive impact. A meta-study in developing countries also showed that out of 101 studies, 59 studies showed a negative impact of the larger class size, but only 30 studies were statistically significant [12]. Surprisingly, however, another 39 studies showed a positive effect, out of which 15 were statistically significant. Another meta-study of developing countries also showed that effect of teacher-pupil ratio alone on student performance was inconclusive [11].

In summary, in developing countries, there is a shortage of teachers, high teacher absenteeism, and use of ineffective pedagogical approaches, and high pupil-teacher ratios which makes automated tutors a good choice.

2.2 Alternatives to Tutors: Better Textbooks and More Homework assignments

One argument could be that rather than supplying schools with adaptive tutors that require computers and other additional infrastructure, perhaps better teaching materials is all that is required. However, providing textbooks to children in Kenya did not raise test scores of students overall, but did increase the scores for higher performing students suggesting that these books were primarily targeted to the smarter students [11]. A Meta-study of developing countries also confirmed that there is little evidence that just providing textbooks, workbooks and exercise books increased student learning [12].

While there is some debate about whether more homework impacts student performance, there is a general correlation between quality and quantity of homework and student achievement in the West (e.g., [16]). This trend also seems to hold for developing countries. For example, in a meta-study of developing country interventions from 1990 to 2010, 5 studies showed mostly positive impact of more homework [11]. However, one key problem is that in impoverished regions of many developing countries, children have to work after school leaving little time to do homework. For example, [17] observed that in a survey of 1030 children in three parts of South Africa, 26.5% of the children working on farms missed school, arrived late or were too tired to do their homework. Similarly, [13] points out that in a country like Ghana, 84% of the parents and 54% of children reported spending less than one hour per day on their homework. In contrast, technology is used in the United States for students to do homework, and provide immediate assessments to the teacher who can guide discussions about the questions that were wrong [18]. Merely showing that a question is wrong, without any further hints or explanations, provided a 0.5 standard deviation of improvement compared to not getting any immediate marking that a homework question was wrong.

2.3 What is known about Personalized Instruction and Adaptive Tutors?

This section summarizes what is known about adaptive learning and automated tutors (not necessarily adaptive) in developing countries and types of effects achieved by doing so. A large study of 140 schools in Keyna shows that the simple act of splitting students into two sections based on ability did have an impact, and effects of 0.14 to 0.18 were achieved [15]. Similarly, using specialized human tutors for remedial classes yielded performance effects of 0.18 to 0.28, while use of an automated tutor achieved 0.35 to 0.47 effects in Mathematics [19]. These two studies show that personalizing instruction at the group level works, and clearly automated tutors also tend to have an impact.

Another related issue is whether the tutor should complement, or replace existing instruction in the classroom. For example, [20] found that a complementary program where children were actually provided computer-aided learning had an effect of 0.28 as opposed to the replacement group whose performance actually got worse (-0.57); the replacement group replaced conventional teaching with an automated tutor. This tends to suggest that automated tutors can perhaps be more effective in a complementary mode.

2.4 Base Grades are lower and Variability is high

The benefits of personalization should be high especially if there is large variability in student achievement within the same classroom, so that the instruction by one teacher might not fit the needs of every student. We have data from a series of studies in the United States, which show performance in a standardized test for both high achieving and low achieving schools, urban vs. rural, in standardized tests that should have good psychometric properties. The test is the Northwest Evaluation Association MAP test, which evaluates students' Mathematics expertise across grades and at different points of the year. Table 1 shows scores of students in two schools during the 2012 academic year.

It can be seen from Table 1 that the standard deviations are small for the NWEA MAP test, which is a test administered by many schools across the United States. The standard deviations for the rural-area high achieving school are 15% of the full range of scores recorded by 7-8 grade students (note: *stdev / (maximum score – minimum score)* = 14.6/(267-167) = 0.146), and 20% of the full range of scores recorded by the low achieving urban school, for grades 9-10 (16.8/(262-180)).

Table 1. Variation in NWEA's MAP standardized test, in two schools in the United States

School	Grade	N	Median	Mean	Natl. Avg.	Std. Dev.	Min.	Max.	Range
Urban-Area School	9-10	97	221	220	234	16.8	180	262	82
Rural-Area School	7-8	223	234	233	228	14.6	167	267	100
Total						16.3	167	267	

No standardized testing data are available for Pakistan. However, results of a standardized Mathematics tests for grade 4 and 5 in semi-rural schools for twenty different school sections (18 different schools) from a recent study [1] show that this ratio (SD/Range) is higher than 20% (Mean = 26.71; SD = 2.97; minimum=21; maximum = 30; Anderson-Darlington, $p>0.05$; single sample t-test; DF = 19, T=10.09, $p<0.05$). In other words, in these semi-rural schools of Pakistan, the standard deviation is more than 20% of the range and sometimes as high as 30% showing more diversity in the learning achievements of students than their counterparts in the United States. Because of this higher variability, it is expected that automated and adaptive tutors will be more effective in developing countries like Pakistan.

Another key issue for students in developing countries is that the overall competencies tend to be much lower than students in the West. For example, in India only 19.5% of third grade children in Vadodara, and 33.7% in Mumbai, passed the grade one competencies (number recognition, counting and one digit addition and subtraction) in Mathematics [19]. Table 2 shows the Mathematics results from the *Programme for International Student Assessment* (PISA) standardized exam for a few third-world countries; Pakistan does not participate in PISA. As can be clearly seen, large proportions of children in a variety of developing countries from different continents tend to have lower proficiency in Mathematics skills than students from Western countries like the United States.

Table 2. Students in Developing Countries are at lowest proficiency of PISA Mathematics Scores

Country	% Students in the lowest proficiency
United States	8
Kazakhstan	30
Trinidad	30
Jordan	35
Argentina	37
Brazil	38
Columbia	39
Albania	40
Tunisia	43
Indonesia	44
Peru	48
Panama	51
Krygyz Republic	65

3 Case Study

As a first step towards localization, a case study was conducted to evaluate how teachers in Pakistan would respond to the use of an interactive adaptive tutor for their students.

3.1 Wayang Outpost – The Adaptive Tutor

An intelligent mathematics tutor for grades 5-12 developed at UMass-Amherst and named Wayang Outpost [4] was selected for this study (See Figure 1). Wayang Outpost has been used by thousands of students in the United States, and students using Wayang Outpost have consistently shown significant learning gains since 2003 on Mathematics tests involving standardized tests items (an increase of 20% achievement level after 3 time periods only), and significant gains on state-standard exams compared to control groups (0.5-0.7 standard deviations depending on the study). Students using Wayang Outpost have also improved more on specific areas of a national standardized test compared to control groups (MAP, a national test of NWEA) for specific mathematics knowledge units that were tutored by Wayang Outpost, and not for other areas of mathematics that were not tutored by Wayang Outpost during those sessions.

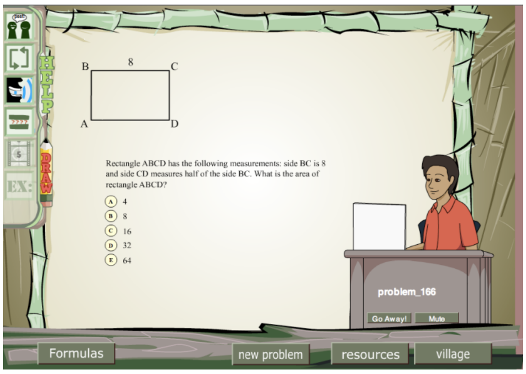


Fig. 1. The Wayang Outpost Math Tutor interface.

Wayang Output targets the United States Mathematics curriculum of grades 6 through 11 covering a large range of topics including number sense, pre-algebra, algebra, geometry, logical reasoning. The pedagogical approach of the Wayang Tutor is based on cognitive apprenticeship and mastery learning, and its internal mechanism of adaptive behavior of item selection is based on empirical estimates of problem difficulty and a variety of parameters that regulate its behavior, which are set by a combination of input from teachers and the researchers [21]. For feedback and scaffolding, Wayang Outpost relies on the Theory of Multimedia Learning, and implements many of its principles, and provides also videos and worked-out examples as part of its support. However, the main mechanism of support consists of hints that solve a small part of the solution for the student, and allow him/her to continue on his own, or ask for more support. Wayang Outpost carries out several instructional tasks: it models (introduces the topic via worked-out examples, making steps explicit, and working through a problem aloud); provides practice with coaching (offering multimedia feedback and hints to sculpt performance to that of an expert's); scaffolds (putting into place strategies and methods to support student learning); provides affective support (via affective characters that reflect about emotions, encourage students to persevere and demystifies misconceptions about mathematics problem solving), and encourages reflection (self-referenced progress charts that allow students to look back and analyze their performance) at key moments of loss, boredom, or un-excitement.

3.2 Survey Design and Data Collection

Based on experiences gained in localizing Khan Academy videos [2], a survey to isolate the various factors that could have an impact on teachers' utilization of Wayang Post was designed as shown in Table 3. The survey was delivered to nine Mathematics teachers on April 28, 2013 in a private urban school in Peshawar, Pakistan. The teachers were introduced to Wayang Outpost, and were led through a one hour session as a student through Wayang Outpost. Each teacher then filled out a survey shown earlier where each item was scored on a Licker-type scale with 1 = Strongly Agree and 5=Strongly Disagree.

3.3 Results

The teachers thought that they could spare on average of about 2 hours per week for an automated tutor session (Mean = 2.44; SD=1.13). This is a substantial amount of time considering that the teachers in this school spend a total of about 5 hours per week on teaching Mathematics. The statistics for the remaining factors are shown in Table 4.

Authenticity (A) and Cultural Alignment (CA) were dropped from further analysis because the value of Cronbach's alpha were lower than 0.7 indicating a lack of internal consistency in how teachers answered the various items; Cronbach's alpha was higher than 0.7 for all other factors. While BI, LC, PA and TA were normally distributed (Anderson-Darlington; $p>0.05$), since the total number of respondents was small ($n=9$), non-parametric analysis was used to analyze the data.

As Table 5 shows, all the internally consistent factors were highly correlated. One key variable in the experiment was whether teachers would use such a tutor in the future (BI). BI can be considered a response variable and based on Ordinal Logistic Regression, BI is strongly affected by PA ($G= 13.278$, $DF = 1$, P-Value = 0.000) with an odds ratio of 0.01. BI is also affected by TA in a similar fashion ($G = 10.899$, $DF =$

1, P-Value = 0.001) with an odds ratio of 0.02. Finally, BI is also affected by BI (G = 5.678, DF = 1, P-Value = 0.017) but the odds ratio is 0.14 indicating that its impact is lesser than those of the two other variables.

Table 3. Survey Design to Determine Factors of Teachers's Adoption

Factors	Items
Teacher Alignment – How well does the tutor fit with teaching style of the teacher?	Wayang Outpost System fits well with the way I teach Mathematics
	Wayang Outpost is consistent with how I like my students to learn Mathametics
	Wayang Outpost teaches Mathematics the way I teach it
Language Comprehension – How well does the child comprehend the language used in the tutor?	The children will understand the language used in Wayang Outpost
	The children will not have any difficulty reading the problems posed in Wayang Outpost
	The children will find it easy to follow the problem descriptions and feedback in Wayang Outpost
Authenticity – How authentic are the problem being posed in the tutor?	The children can relate to the problems being posed in Wayang Outpost
	The examples in Wayang Outpost are consistent with how these children live their lives
	The problems in Wayang Outpost are about things that these children care about
Cultural Alignment – Is the tutor aligned with cultural norms and taboos?	The problems in Wayang Outpost do not violate any traditions or taboos
	The problems in Wayang Outpost are consistent with the Pakistani culture
	The problems in Wayang Outpost are not alien to children from a cultural perspective
Pedagogical Alignment – Is the tutor consistent with how children learn?	Children will not have any difficulty following the way Wayang Outpost teaches
	Children will enjoy interacting with the various characters that help them out while solving problems using Wayang Outpost
	Children will like solving problems and getting feedback on their performance using Wayang Outpost
Behavioral Intention – What is the likely-hood of the teacher using the tutor in the near future?	If Wayang Outpost were available, I would use it in my classroom
	I would like to use Wayang Outpost for teaching Mathematics
	It would be great to use a system similar to Wayang Outpost to teach Mathematics

Table 4. Summary of Survey Responses (n=9)

Factor	Mean	StDev	Median	Min.	Max.	Median [95% Conf. Intrval]
Authenticity (A)	2.70	0.61	2.66	1.66	3.33	2.67 [2.17, 3.33]
Behavioral Intention (BI)	**2.29**	**0.82**	**2**	**1**	**3.66**	**2.17 [1.67, 3:00]**
Cultural Alignment (CA)	3.18	0.62	3.33	2.33	4	3.17 [2.67, 3.67]
Language Comprehension (LC)	2.96	0.94	3	1.33	4	3 [2.17, 3.67]

Pedagogical Alignment (PA)	2.25	0.83	2.33	1	3.33	2.33 [1.50, 3.67]
Teacher Alignment (TA)	2.29	0.77	2.333	1	3.33	2.33 [1.67, 3:00]

Table 5. Correlation between the Various Factors (Kendall-Tau; * = p<0.05)

	LC	BI	TA	PA	CA
BI	0.627*	1			
TA	0.618*	0.746*	1		
PA	0.618*	0.806*	0.941*	1	
CA	0.462	0.344	0.4	0.462	1
A	0.576*	0.646*	0.667*	0.637*	0.381

In summary, the data show that there was a reasonable probability that the teachers would use the system if available, and were willing to allocate about two hours per week for this activity. Further, their intention to use Wayang Outpost of or a similar system is contingent on teacher and pedagogical alignment, and whether Wayang Outpost's language would be understood by the children. However, is important to note that as Table 4 shows, while teachers were not negative about any of the factors, they were mostly not sure (closer to Neither Agree nor Disagree) about pedagogical and teacher alignment etc. This strongly implies the need to consider using these factors in localization of Wayang Post.

5 Conclusion and Future Work

While adaptive tutors like the Wayang Post have shown considerable impact in improving learning outcomes in countries like the United States, an exploitation of the full potential of such systems in the developing world is contingent on careful localization that goes beyond simple language translation. Clearly, there is a dire need for such systems in the developing world and even though the sample size was small, this paper shows that teachers in a developing country are likely to adopt such systems provided the issues of teacher, student alignment etc. are adequately addressed. The challenge now remains to find the resources to localize and deploy adaptive tutors such as Wayang Outpost.

Acknowledgements

This research was funded by an award from the National Science Foundation, NSF # 1109642, REESE, Personalized Learning: strategies to respond to distress and promote success, Ivon Arroyo (PI), with Beverly Woolf and Winslow Burleson. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies. The authors would also like to thank TeleTaleem (pvt) Limited, Islamabad, Pakistan for help with the logistics of conducting the teacher study.

References

1. Zualkernan, I. A. & Karim A. (2013a). Using a Traveling Van to deliver Blended Learning in a Developing Country, ICALT 2013, Beijing, July, 2013 (to appear).
2. Zualkernan, I. A. & Karim, A. (2013b). Zualkernan, I. A. & Karim A. Online Content Localization for Blended Learning in Developing Countries: A Case Study Using Khan's Academy, Proceedings of the 7th International Technology, Education and Development Conference, Valencia (Spain), March 4-6, 2013.
3. VanLehn, H. K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems, Educational Psychologist, vol. 46, no. 4, 2011, pp.197-221.
4. Arroyo, I,, Beal, C.R,, Murray, T,, Walles, R, Woolf, B.P. (2004) Web-Based Intelligent Multimedia Tutoring for High Stakes Achievement Tests. Proceedings of 7th International Conference on Intelligent Tutoring Systems Conference. Maceio, Brazil. LNCS, Springer.
5. Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. (1997). Intelligent tutoring goes to school in the big city. International Journal of Artificial Intelligence in Education, 8, 30-43
6. VanLehn, K., Lynch, C., Schulze, K. Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., & Wintersgill, M. (2005). The Andes physics tutoring system: Five years of evaluations. In G. McCalla, C. K. Looi, B. Bredeweg & J. Breuker (Eds.), Artificial Intelligence in Education. (pp. 678-685) Amsterdam, Netherlands: IOS Press.
7. Bloom, B. S. (1984). The 2 sigma problem: The search formethods of group instruction as effective as one-to-one tutoring. Educational Researcher
8. CTTG (2012). Closing the Trained Teacher Gap, Global Campaign for Education, Rosebank, Johannesburg, South Africa, 2012.
9. Guerrero, G., Leon, J., Zapata, M., Sugimaru, C. & Cueto, S. (2012). What works to improve teacher attendance in developing countries? A systematic review, London: EPPICentre, Social Science Research Unit, Institute of Education, University of London.
10. Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K. & Rogers, F. S. (2006). Journal of Economic Perspective, Volume 20, Number 1, Winter 2006, Pages 91–116.
11. Glewwe, P. W., Hanushek, E. A., Humpage, S. D., Ravina, R., (2010). School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010, Working Paper 17554, Available at: http://www.nber.org/papers/w17554.
12. USAID (2012). USAID Education Strategy Reference Materials, USAID, April, 2012.
13. Abadzi, H. (2007). Absenteeism and Beyond: Instructional Time Loss and Consequences, The World Bank Independent Evaluation Group, Thematic, and Global Evaluation Division, October 2007.
14. WBI. (2012). World Bank Development Indicators, The World Bank, Washington, D.C.
15. Duflo,E., Dupas,P. & Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya, *American Economic Review 101* (August 2011): 1739–1774.
16. CCLC. (2009). A systematic review of literature examining the impact of homework on academic achievement, Canadian Council on Learning, Canada.
17. Dewes, A. (2012). Agricultural work in South Africa: A contested Space. in *Childhood Poverty: Multidisciplinary Approaches* edited by Jo Boyden, Michael Bourdillon, 2012.
18. Kelly, Y., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, G. & Soffer, D. (submitted). Estimating the Effect of Web-Based Homework. Proceedings of the 16th International Conference on Artificial Intelligence in Education. Memphis, TN. 2013
19. Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying Education: Evidence from Two Randomized Experiments in India, The Quarterly Journal of Economics (2007) 122(3): 1235-1264.
20. Linden, L. (2008). Complement or Substitute? The Effect of Technology on Student Achievement in India, MIT Jameel Poverty Action Lab, Cambridge, Mass.
21. Arroyo, I., Mehranian, H, Woolf, B.P. (2010) Effort-based Tutoring: An Empirical Approach to Intelligent Tutoring. Proceedings of the Third International Conference on Educational Data Mining. Pittsburgh, PA.

A Case Study of the Localization of an Intelligent Tutoring System

Phaedra Mohammed and Permanand Mohan

Department of Computing and Information Technology, The University of the West Indies, St. Augustine, Trinidad and Tobago
{phaedra.mohammed@gmail.com, pmohan@tstt.net.tt}

Abstract. The matter of a one-size-fits-all approach towards the development of culturally-relevant educational software is debated with one side arguing for internationalization and the other side arguing for localization. This paper takes a pragmatic look at the issues involved in localization and aims to shed light on the strengths and limitations of undertaking culture as a design feature. With an emphasis on the application layer, the paper investigates the requirements and steps that need to be taken when using cultural contexts in educational software. It describes the design of a localized intelligent tutoring system developed for the context of Trinidad and Tobago and discusses how the prototype was evaluated in two separate studies which looked at learning gains, students' opinions, attitudes, and preferences for localization.

Keywords: Localization, cultural translation, intelligent tutoring systems

1. Introduction

This paper is set in the context of Trinidad and Tobago. With a GDP of $20,400US [2] Trinidad and Tobago is one of the wealthiest nations in the Caribbean. Liquefied natural gas, petroleum and its byproducts make up the bulk of the country's exports and account for approximately 40% of the country's GDP [7]. The average population size is 1.2 million, life expectancy is estimated at around 72 years, and literacy rates are over 98% for ages 15 and older [2]. The country is becoming more modernized as evidenced by the increasing number of Internet users (growth from 8% in 2000 to 48% of the population in 2012) and large number of cell phone users (over 1.8 million) [2]. Although access to personal computers is not as widespread with roughly less than 20% of the population having access, the government of the country provided free laptops to entry-level students in secondary schools in 2010. The challenge that now arises for the country's education sector is whether the software on these machines can support learning in the context of Trinidad and Tobago.

Accommodating for learner diversity based on cultural backgrounds is becoming a major personalisation focus with the increasing drive towards globalization as evidenced in Trinidad with the distribution of laptops. Despite this drive, the knowledge and processes for incorporating culture have not been clearly defined with automation in mind. The matter of a one-size-fits-all approach towards the development of educational software is debated with one side arguing for internationalization and the other side arguing for localization. Proponents of localized approaches argue that culture increases the credibility, realism, familiarity and acceptance of educational systems with the end result being a higher quality learning experience [4, 5]. On the other hand, techniques have been described for creating internationalized designs that do not target a particular culture and avoid cultural specificity but still cater for the needs of a learner. This viewpoint is based on arguments that cultural designs tend to be cosmetic and stereotypical, suffer from designer bias [1], and are overall difficult to automate [8].

The purpose of this paper is not to take a particular side or argue for or against a particular approach. Rather, this paper takes a pragmatic look at the issues involved, and aims to shed light on the strengths and limitations of undertaking culture as a design feature in educational systems. It focuses on the requirements and steps involved in carrying out localization at the application layer. The paper then investigates the requirements and steps that need to be taken in order to use the cultural context of student in educational software and then describes the design of a localized intelligent tutoring system (ITS) developed for the context of Trinidad and Tobago with these requirements in mind. Next, the paper discusses how the system was evaluated in two separate studies which looked at learning gains, students' opinions and attitudes towards the system, and student preferences for localization. The paper concludes with an outline of the lessons learned from these studies and potential research directions for localized systems.

2. The Strengths and Challenges of Localization

Educational frameworks and systems from different sources are typically repurposed in order to reduce costs associated with content development, to replicate proven results with learning gains, and to set standards in educational curricula. Repurposing of such environments entails some form of localization since the design of user interfaces, the selection of teaching strategies, the format and content of the educational material all vary depending on the contextual background of the developers. In order to relate to students and in some cases not offend others, localization of educational environments has been recommended and there have been benefits cited in the literature for students such as increased motivation levels [3, 6].

There are several challenges associated with localization such that actual systems are limited in practice. Many developers have shied away due to the complexity in reliably representing aspects of a particular culture and because of the ill-defined nature of culture [1, 8]. The costs can be higher in the long run because of the amount of effort required in cataloguing cultural knowledge. In addition, localization requires many pieces of metadata for adaptation to go beyond keyword insertion and colour changes. Learning gains seen in one country are not guaranteed in another since many variables exist across countries such as the system may not be used in the same way as in the original country, there might be limited internet connectivity, different hardware, untrained teachers, or even a different curriculum. Lastly, a localized solution runs the same risk of being as offensive or irrelevant as the original version since a student does not belong to a discrete cultural group, and more than just cultural knowledge is required in order to relate to students; a sound understanding of environmental context is critical.

3. Requirements for Effective Localization at the Application Layer

There are three entry points for localization in educational software systems: the presentation layer, the application layer, and the data layer. The application layer selects, modifies, aggregates, and formats raw content based on the user's input, history of events and intended instructional goals. Firstly, data is needed about the student's country of residence (target country) because this data defines the environmental context that the student is familiar with. Such data is typically available from the country's national statistical office or department and can be used to set the scope of the localization. For example, statistics on population density, economic activity, and religious group distribution can be used. Secondly, demographic data from the student is required in order to model his/her contextual background and to know what type of localization to carry out and to what extent. This data comes from users directly entering information about themselves in a form for instance and can include a user's age, gender, native language and residence location for instance. Thirdly, data is required on the contextual groups in the target country specifically what kinds of contextual elements are familiar, appropriate, and relevant to a specific group. This is needed in order to control the direction of the localization process. Next, language rules and localized natural language terms are needed in order to translate textual content and adapt the cultural meaning of the content to suit the student. Semantic markup on these cultural terms is required in order to know how to use the terms in sentences correctly and also to be able to interpret image context in a culturally appropriate manner. Lastly, real-time localization is required in order to adapt content to suit different student preferences.

4. Localized System Design

This section describes the design of a contextually-relevant Intelligent Tutoring System (ITS). A modular design was chosen because of the complexity involved in delivering culturally-relevant instructional experiences. Flexible alteration and improvement of the component features are easily accommodated as a result. Many of the components featured in the design of the localized web-based ITS are based on traditional ITS components but have been modified to include localized data and functionality. A cultural student model, a pedagogical module, cultural heuristics and, a content repository make up the major architectural units of the design.

The cultural student model records the pedagogical events, performance-related student data, suggested hints and instructional guidance given to the student. Cultural background data is assimilated into this model and is made up of demographic data and cultural influences. The pedagogical module consists of instructional rules that constantly access and update the student model in response to input data from the student and events captured from the screen. They control how and when instructional feedback is given, and they manage the selection and transition of learning activities.

A culturally-relevant instructional approach requires that cultural references made by software systems should be applicable to the learning content, familiar to the students, authentically rendered, and integrated into the context of the instructional material [4, 5]. Cultural rules modify the textual portions of instructional content such as question descriptions, scenarios, hints, and instructional feedback produced by systems. The localization of visual portions of these systems, namely the multimedia related to the learning activities, is also handled by these rules. Textual modifications include customizing the language of the instructional feedback, whereas the multimedia localization involves swapping in cultural assets for generalized assets such as images. The scope of the languages used for localization in this paper is restricted to English-based Creole languages. These languages are useful for educational environments because they foster comfortable learning experiences especially in domains that are seen by students as problematic or difficult to cope with [6]. Textual outputs of the localization process are therefore sentences expressed in a local dialect specifically mesolect forms[1], and feature equivalent cultural lexical terms. In order for these rules to properly localize the learning activities and determine appropriate feedback for the student, the learning materials need to be accompanied with metadata descriptions that identify content that can be localized such as hints, question descriptions, and images.

The content repository handles the organisation and distribution of all instructional and cultural assets to a content aggregator. Localized ITSs rely on reusable content more than non-cultural ITSs because of the additional dimension of cultural personalisation; this was the basis for having a separate asset repository - reusability. The content repository primarily hosts all of the educational and interface-related material used by the system and the student model. For example, multimedia files related to the interface's look and feel, such as icons, logos, and those related to the learning exercises (scenario pictures, feedback pictures) are stored here together with educational material such as question descriptions, solutions, feedback files, and topic hierarchies. Each of these assets is described by their asset metadata descriptions which define the context of use and the nature of the assets. These descriptions are indispensable in the design because they facilitate reuse and exchange of compatible assets. Both the pedagogical module and cultural heuristic component use these descriptions when making instructional and localization-related decisions.

5. System Implementation

Two web-based ITSs were implemented based on the software architecture described in the previous section. One system was localized for Trinidad and Tobago's

[1] A variety of language in a Creole continuum that is intermediate between the standard form (acrolect) and forms that diverge greatly from the standard form (basilect).

context (Culturally Relevant Instructional Programming System – CRIPSY) while the other remained generic (Instructional Non-CulturAl Programming System – INCAPS).

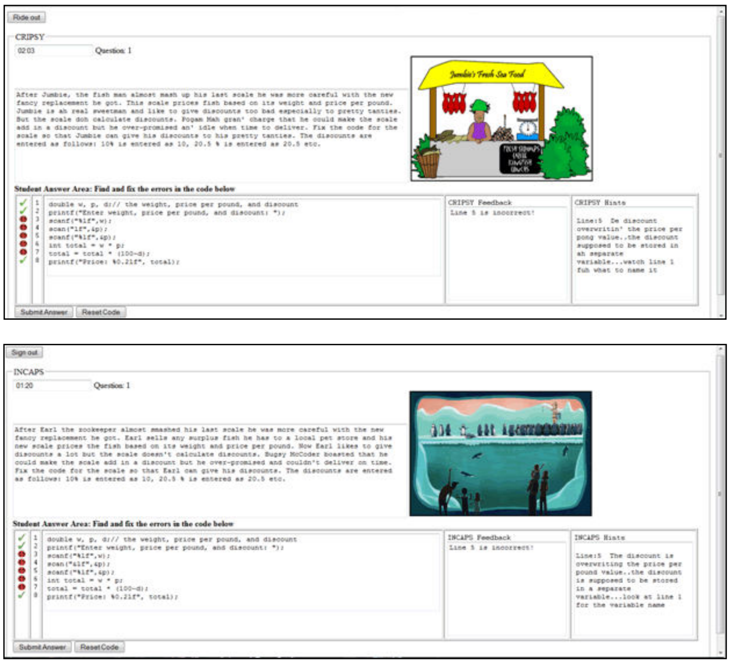


Fig.1. Screenshots of CRIPSY (top) and INCAPS (bottom) featuring localized and non-cultural versions of the same programming exercise.

Both systems were built for the same educational domain (Computer Science programming) were identical from a functional standpoint, and were implemented using Java-based tools and technology which facilitated seamless integration of the various components into complete systems. The pedagogical module, student model, and cultural heuristic component were implemented using JESS (Java Expert System Shell) rule engines and rules. At the data level, simple file formats were used to manage the content repository since rule engines handle data manipulation primarily using facts. Constructivism and situated cognition were selected as the major instructional strategies since analytical programming skills were being targeted and also because of the good fit between situated cognition and culturally-aware instruction. The development of localized assets for CRIPSY was done semi-automatically and manually. The programming exercise descriptions, parts of the exercise code and instructional hints were localized using subtle, careful use of cultural semiotics, specifically cultural names of objects and foods. As shown in the screenshots in Figure 1 above, both systems used the same instructional content but differed in the expression of the content, that is, cultural and non-cultural. A minimalist interface design was used and instructional feedback consisted of identification of correct/incorrect lines of code, hints for the incorrect lines, and general informative guidance.

6. Prototype Evaluation and Results

A previous study was done using both prototypes and participant details are reported in [6]. That study revealed that the use of both systems resulted in significant increases in student test scores and one design issue affected students in particular: the language used in the localized system. The density of the localisation distorted the system's content into basilect[2] Creole which for some students made the sentences difficult to read quickly and therefore difficult to understand quickly. This observation stimulated the consideration of incorporating a customizable language density scale in the localized system so that students may adjust the language to suit their own preferences. The desirability of such a feature was assessed in the second study.

The second study aimed to find out which system appealed more to the students, what kinds of conditions would impact if at all upon their preferences, and whether a language localization slider would be a desirable feature and why. Fifty-eight (58) students from the previous study participated in this study. The control and test groups were switched so that students would have used both systems at the end of this study. A similar procedure was followed where each student's username activated their newly assigned system and the systems timed out after 30 minutes. A short questionnaire was administered and the results are shown in Table 1 below. After using both systems, more students (68.9%) preferred the localized system (CRIPSY) over the non-localized one (INCAPS). The majority of students reported that they wanted the localization slider (79.3%) while the minority did not see the slider as necessary. Increasing question difficulty did not seem to influence student preference for either system. Lastly, more students (13.8%) changed their preferences from the localized system to the non-localized system when server glitches or software problems occurred compared to those whose preferences were reversed in favor of the localized system (3.4%).

Feedback Topic	**Student Preferences**	**Percentage of Students**
Localization Slider	Wanted localization slider	79.3
	Did not want localization slider	20.7
System of Choice	Preferred localized system in general	68.9
	Preferred non-localized system in general	31.1
Preference in Relation to Question Difficulty	Preferred localized system	48.3
	Preferred non-localized system	48.3
	No preference/no response	3.4
Preference change if glitches occurred in System of Choice	Changed to non-localized from localized	13.8
	Changed to localized from non-localized	3.4
	No change in preference	3.4
	No response	79.4

Table 1. Percentage of student preferences categorized by feedback topics

[2] The basilect variety of English-based Creoles is the most distorted from Standard English.

7. Analysis of Results

The results reported in the previous study using the localized ITS, CRIPSY, confirmed the assumption that cultural interventions do indeed have positive effects on learners and provide empirical evidence in support of localized learning systems. Overall, the students liked the localized system primarily because of the reasons outlined in the earlier studies in [6] namely enriching learning experiences and humour. The use of culture created a familiar setting and it was done in a way that was interesting to the students. A larger percentage of students rated the programming exercises as easier and the instructional feedback/hints as more helpful for the localized system although similar guidance was given in the control system.

In the second study, the majority of students (79.3%) wanted the localization slider. The most common reasons given by students for wanting a slider included: wanting control over the timing and the degree of localization, wanting to change the localization to suit their moods, and wanting to be able to explore the different degrees of localization. An interesting trend in the responses of the students who did not want the slider was their dislike of the localization. Many of them stated that the localization resulted in descriptions that were longer to read and had too much localized language which was confusing. Longer descriptions were indeed the case since the highest density of localization was used in generating content for the study, and therefore the maximum number of lexical insertions and replacements possible for the content was made. This indicates that there is a strong need for the slider since the students essentially wanted to choose their own levels of localisation. Another interesting result is the low student tolerance for software faults in the localized system evidenced by the larger numbers of students who changed their preference to the non-localized system when faults occurred in the localized one. A possible cause could be that students perceived the localized system as being inferior because of the cultural behaviour of the system in using language levels (basilect) typically associated with the less educated in Trinidad and Tobago. This implies that being able to dynamically adjust localization is crucial for system acceptance by students.

8. Conclusion and Future Research

Culture is rapidly becoming an important consideration in the design of educational software firstly because of the increase in the number of users accessing software over the Internet, and secondly because of the sheer diversity in the cultural backgrounds of these users. Conventional learning has often taken place in a localized setting with a teacher guiding one or more students in their search for knowledge. With the advent of the Internet, this traditional setting has changed drastically since students now have access to teachers and educational material from over wide distances. Consequently, these students are exposed to a variety of educational tools, teaching strategies and learning materials which were not developed with their own personal needs in mind. This has dramatic usability implications especially when the mainstream culture for which e-Learning materials are designed clashes with that of the users.

Based on the encouraging evidence established by these studies, the research discussed in this paper demonstrates a practical approach towards developing a localized web-based learning environment. By leveraging research from various fields such as Intelligent Tutoring Systems and culturally-aware instruction, this research shows how some of the complexity of localization can be managed and how aspects of intelligent tutoring can be localized. Empirical evidence indicates that localized systems perform as well as traditional tutoring systems and are potentially superior at creating relaxed, engaging learning atmospheres for the Computer Science programming domain. However care must be taken to ensure that the cultural enhancements match the tolerance level of the student users. Further refinement and improvements are planned for the systems described. A limited amount of cultural automation was undertaken, so expansion of the cultural coverage is necessary. Additional features such as deeper cultural student profiling, adjustable language density and greater tutoring flexibility are also part of the plans intended for this research.

References

1. Blanchard, E.G., Mizoguchi, R., Lajoie, S. P.: Structuring the cultural domain with an upper ontology of culture. In Blanchard, E., Allard, D. (Eds.): The Handbook of Research on Culturally-Aware Information Technology: Perspectives and Models. 179-212. Hershey, PA: IGI Global. (2011)
2. CIA The World Factbook. Available online: [https://www.cia.gov/library/publications/the-world-factbook/geos/td.html] (2012)
3. Gilbert, J.E., Arbuthnot, K., Hood, S., Grant, M. M., West, M.L., McMilllan, Y., Cross, E.V., Williams, P., Eugene, W.: Teaching Algebra Using Culturally Relevant Virtual Instructors. The International Journal of Virtual Reality 7, 1, 21-30 (2008)
4. Henderson, L. Theorizing a multiple cultures instructional design model for e-Learning and teaching. In Edmundson A. (Ed.): Globalized E-Learning Cultural Challenges. 130- 154. London, UK: Information Science Publishing. (2007)
5. McLoughlin, C.: Adapting e-Learning across cultural boundaries: A framework for quality learning, pedagogy, and interaction. In Edmundson A. (Ed.): Globalized E-Learning Cultural Challenges. 223- 238. London, UK: Information Science Publishing. (2007)
6. Mohammed, P., Mohan, P.: The design and implementation of an enculturated web-based intelligent tutoring system for Computer Science education. In I. Aedo, N-S. Chen, D.G. Sampson, J.M. Spector, Kinshuk (Eds.): 11th IEEE International Conference on Advanced Learning Technologies (ICALT). 501-505. Washington, USA: IEEE Computer Society. (2011)
7. Trinidad and Tobago. Available online: [http://en.wikipedia.org/wiki/Trinidad_and_Tobago] (2013)
8. Rehm, M.: Developing enculturated agents: Pitfalls and strategies. In Blanchard, E., Allard, D. (Eds.): The Handbook of Research on Culturally-Aware Information Technology: Perspectives and Models. 362-386. Hershey, PA: IGI Global. (2011)

Impact of a blended ICT adoption model on Chilean vulnerable schools correlates with amount of on online practice

Roberto Araya and Johan Van der Molen

Centro de investigación Avanzada en Educación (CIAE), Universidad de Chile, Chile

roberto.araya.schulz@gmail.com

Abstract. The impact of a blended ICT adoption model in 15 fourth grade classes from 11 vulnerable Chilean schools is analyzed. In this model, twice a week students attend a computer lab where the lab teacher and the class teacher select and assign on line math exercises. The platform contains approximately 2,000 exercises, but it is constantly growing with exercises introduced by teachers. During sessions the system continuously tracks students and detects students that are falling behind, allowing teachers to provide real time support and drive the progression of the entire class as a whole. Students that finish early and with good performances are assigned as part of the support team for the rest of the session. After a year of implementation, the national assessment test SIMCE math on these classes raised 0.38 standard deviations. This is more than three times the historic national improvement in 2011. The statistical analysis shows that the improvement was independent of teacher effects, and correlates with the average number of on line exercises done per student in that year.

Keywords: adoption model, ICT impact, national assessment tests, educational data mining, at risks students

1 Introduction

We study a blended Information Computer Technology (ICT) adoption model focused on ensuring intensive student practice and real time support from teacher and more advanced peers. We implemented this ICT model on several low socioeconomic status Chilean schools with high proportion of at risk students. According to a recent UN report [11], Chilean gross national income (GNI) per capita is 14,987 in 2005 PPP $, Chilean Human Development Index is the highest in Latin America and ranks 40 in the world. Even though gross national income per capita has had an important average annual growth of 3.8 from 1990 to 2012, there is huge school segregation by socioeconomic status in Chile. There is also a big educational inequality measured as

adfa, p. 1, 2011.

interschool standard deviations on a UNESCO math and language tests for third and sixth graders [14], [1], though the educational inequality is generally lower than in other Latin American countries. For example, in the UNESCO math test for third graders, interschool standard deviation was 1.97 the standard deviation obtained if students were randomly assigned, which is one of the lowest ratio of the studied countries. The educational inequality is more pronounced on the urban sector than in the rural sector. According to the last OECD PISA-2009 assessment [16], [17], science and reading performances have been improving and the performance gap in reading between high and low socioeconomic status students has being reducing. From the first UNESCO study [12] to the second UNESCO study [14], the relative position in Latin America of the math performance of Chilean third graders improved. Also the Chilean national math assessment test for fourth graders (SIMCE) shows in 2011 a significant improvement (0.12 standard deviations) and the performance gap has also been reducing (0.22 standard deviations). According to UN [11], fixed broadband internet subscription was 10.5 per 100 people in 2010 in Chile. This is one of the highest subscription rates in Latin America together with Uruguay and Mexico. Fixed and mobile telephone subscribers were 136.2 per 100 people in 2010. In schools, 90.2% of students have computers to use with access to internet [13].

Adoption of ICT generates several benefits in education [18], [19], but it also poses several challenges particularly in vulnerable schools. According to our field experience in vulnerable schools there is a weak ICT infrastructure, poor equipment maintenance, poorly prepared technical support personnel, high frequency of electric supply problems, and instable connection to internet. All of the difficulties were present on the schools where the model was implemented. Besides school infrastructure, there is a much weaker ICT infrastructure at students' homes, less availability of internet access and much weaker access to technical support. Other important difficulties in these schools are a higher proportion of at-risk or behind-grade-level students, higher percentage of students that are struggling with core math concepts, and lower attendance rates than in other schools. Additionally, students' families have less cultural capital, and therefore students are harder to teach. Teacher quality is also generally lower in these schools, since better teachers once detected they receive offers to migrate to higher socioeconomic status schools. All of these conditions pose a much higher challenge in the introduction of ICT and the possibility to have a positive impact.

2 Method

One of the main goals of the implemented model was to provide technological and human resources to increase the amount of student practice on mathematics and also to provide real time feedback for all students, and particularly for struggling students. There is abundant literature reporting the positive effect of practice in ideal laboratory conditions. There is the positive effect on memorization [6], in long term retention [7], in understanding when compared with spending time building concept maps or only studying [3], and the positive effect of practice throughout several sessions [8].

Another important aspect is feedback. There are different types of feedback: immediate versus delayed, simple positive feedback if the answer is correct or negative feedback if it is incorrect, or deeper feedback with suggestions of alternative solutions, feedback aimed at increasing student effort, etc. In a review of dozens of studies [2] concluded that the positive feedback for right answers is better than negative, and that feedback given by video, audio or computer systems is the most effective. There is also evidence of the positive effect of accumulated practice in solving mathematical problems [4]. More practice produces more learning, particularly if practice is mixed with previously learned knowledge. However there is lack of studies of impact outside of laboratory conditions [9], and particularly in low socioeconomic schools. There are several important challenges to implement a strategy based on intensive practice. One of them is motivation. There is the need to spark engagement with math problems throughout the whole year. Moreover to improve the performance of a complete class it is necessary to motivate struggling students to do intensive practice. At the core of the blended adoption model there is an early alert system. At every moment, the system lists students who are having more difficulties. In this way the teachers know in real time which students need personal attention and in what specific exercise. The early alert system also detects if there are exercises that are producing high difficulty to the whole class. This way the teachers can freeze the system and explain the required concepts. The platform is designed to drive the progression of the entire class as a whole, and not to leave students alone. It has facilities to promote the cooperation and support of students that are ahead of their peers. Students that finish early and with good performances can be assigned as part of the support team for the session. The system assigns them to help peers with difficulties. At the same time, students that are being assisted by peers assess the quality of each support event. This information helps the teachers to monitor the quality of the support and the need to teach support strategies to advanced students.

The model was implemented in 11 schools. The first year the system was implemented in the second half of the year, but most of the time was dedicated to solve infrastructure problems. In the second year (2011) the system was running in more normal conditions. All of the 469 students from the 15 fourth grade classes of the 11 municipal schools from the Lo Prado municipality were using the system. Lo Prado is a low socioeconomic district in Santiago. During the year 2011, from end of March to early October the students did 203,782 on line math exercises. After mid-October students did more exercises but they were done after the National test SIMCE, and therefore they will not be considered in the analysis. Around 20% of the exercises were assessment exercises and the rest were exercise with feedback. For every exercise done by a student, the platform records the response time as well as the performance. In the case of an assessment exercise, the system records if the student answered correctly or not, and in the case of exercises with feedback it records the number of attempts to achieve the correct answer.

3 Results

According to What Work Clearinghouse standards of the U.S. Department of Education, evidence of effectiveness [10] of educational interventions is classified into three levels: low, moderate and strong. Evidence is low when it is based only on expert opinion, which is derived from strong findings or from theories in related areas and/or expert opinion. Evidence is rated as moderate when they are made with high internal validity studies (studies whose designs support causal conclusions) but moderate external validity (studies that include a sufficiently broad range of participants to be generalizable), or vice versa, but not both validities simultaneously. At this level, belong the typical quasi-experimental studies (i.e., the control and experimental groups were not randomly assigned) that show the effectiveness of a program, or studies with small samples or groups that are not equivalent at the pretest, or correlational research with strong statistical controls to avoid selection bias or assessments that meet the Standards for Educational and Psychology Testing but the samples are not adequately representative of the population. Evidence is strong if the studies have high internal and external validity. These designs include randomly controlled trials, multisite, large scale, and no contradictory evidence.

We make two types of analysis. First we compare the SIMCE math gains of all the schools of the Lo Prado district with the result of the whole country. We also compare the Lo Prado SIMCE math gains with the results of the urban schools of a very similar neighbor district of the same socioeconomic level. Second, we compare the Lo Prado SIMCE math gains with the Lo Prado SIMCE language gains and SIMCE science gains in the same classes. Given that in each class all subjects are taught by the same teacher, this comparison aloud as to explore the teacher effect and eventually compute if the improvement obtained is independent of the teacher. Therefore, the study reported here could be classified as quasi experimental at best, and then the evidence can be classified as moderate.

One difficulty of analysis is that the SIMCEs scores are not published by student. We only have the average score of each class and the standard deviation of the country. For this reason, in what follows we use the variation of SIMCE for classes. We define the SIMCE variation for a class in a given school as the difference between the SIMCE 2011 score obtained by the class with the SIMCE 2010 average of all classes of the school. It is important to note that the students are different, since they belong to different generations. We are comparing the 2011 score of a class with the average score of the school in the previous year.

The increase of SIMCE math for the municipal schools of the Lo Prado municipality was of 16.27 points, which is greater than the variation of all schools in the country. This increase is obtained from the simple average of the SIMCE math scores of the schools. However, some schools have fewer students, so the simple average of the schools is not the most representative. Weighting proportional to the number of tested students in each course, the variation in SIMCE for math was 19.15 points. This increase is more than three times the historical rise of the country SIMCE math that was 6 points [5], and it is much higher than the increase of 8 points in SIMCE math in the schools with similar socioeconomic classification as 10 of the 11 schools analyzed.

This improvement is an in increase in 0.38 standard deviations. To test if the average improvement of the 15 classes of Lo Prado is higher than the 6 points of the average of the country student's improvements and the 8 points of the average of the country at risk student's improvements, we consider the fact that we have the standard deviation σ of the student performance in the country. From that information we can compute the standard deviation of the average of the score improvement of the 392 students that took the SIMCE test. The standard deviation σ of the student performance does not change from one year to another and σ=50. Assuming independence of the performance between students, the variance of the performance of each class i equals $\frac{\sigma^2}{n_i}$ where n_i is the number of students on the class i, and similarly the variance of the average performance of the school s(i) where the class i belongs to is equals $\frac{\sigma^2}{n_{s(i)}}$, where $n_{s(i)}$ is the number of students on fourth grade at the school s(i). The difference D_i between the score of the class i and the school average on the previous year has a variance equal to $\frac{\sigma^2}{n_i}+\frac{\sigma^2}{n_{s(i)}}=\frac{\sigma^2(n_i+n_{s(i)})}{n_i n_{s(i)}}$. Thus the weighted average score D= $\frac{1}{n}\sum_{i=1}^{k} n_i\, D_i$ of all the k=15 classes of Lo Prado has a variance Var(D)= $\frac{\sigma^2}{n^2}\sum_{i=1}^{k}\frac{n_i(n_i+n_{s(i)})}{n_{s(i)}}$, where n is the total amount of students in the k classes. The null hypothesis is H_0: $\mu = \mu_0$, where the observed mean of D is 19.15 and μ_0=8 is the mean improvement of all the at risk schools in the country (or μ_0= 6 is the mean improvement of all the students of the country). Since Var(D) = 11.1622, the cohen´s d = 3.3719, and then p=0.00074665 for μ_0 = 8. Thus the improvement obtained in Lo Prado is statistically significant. The difference between the country improvement and the average improvement of the 392 students from Lo Prado is statistically significant with a two tailed p-value < 0.0005 and the difference between the improvement of all the country at risk students and the Lo Prado students improvement is statistically significant with a two tailed p-value=0.00074665.

It is very instructive to compare the results of the schools belonging to the municipality of Pudahuel. This is a neighbor municipality and that years ago were both part of a single municipality. In Pudahuel we only consider urban schools. They are similar to those of Lo Prado which all are urban schools. Figure 1 shows the distribution of vulnerability of municipal schools in Lo Prado and the urban municipal schools of Pudahuel. The average of the 11 school´s proportions of at risk students in Lo Prado is 0.80 whereas the average of the 14 school´s proportions of at risk students in Pudahuel is 0.77. The t-test of two sample sizes and unequal variance does not reject that they have different proportions (two tailed p-value= 0.36). In addition, 10 of the 11 schools of Lo Prado are classified as “medium low” socioeconomic status (SES) by the Ministry of Education, and the remaining school is classified as "medium", which is quite similar to what occurs in Pudahuel, where 12 of the 14 schools are classified as "medium low" and the other two as "medium".

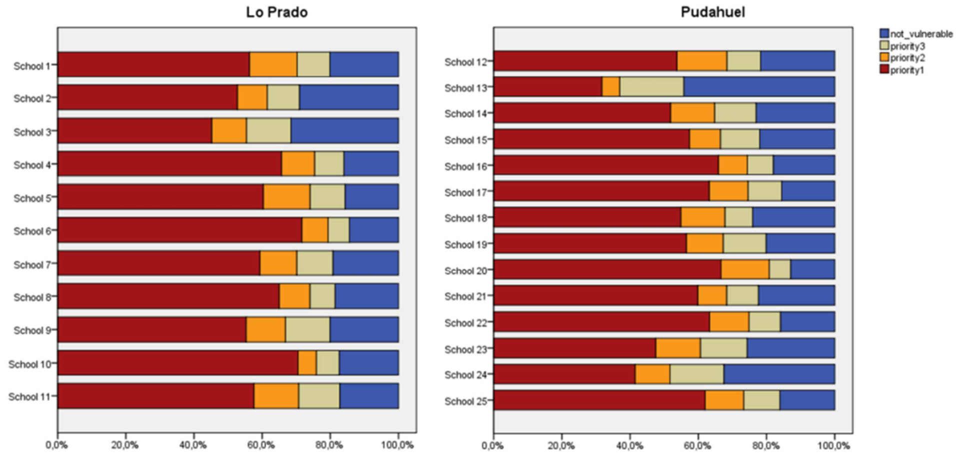


Figure 1: Proportion of at risk students in the schools of Lo Prado and Pudahuel.

In Lo Prado district the increase of SIMCE math schools was 19.15, corresponding to 392 students that took the SIMCE, whereas in Pudahuel the increase was 12.5, corresponding to 790 students. Using again the fact that the standard deviation of the average score improvement of all students on both municipalities can be computed from the standard deviation of student scores, the difference between the average scores however has a p-value = 0.064 unilateral.

There is a possibility that the increase in math SIMCE is due mainly to the contribution of teachers and/or schools. Since in all the classes the teacher teaches all subjects, if the effect is due to specific characteristics of the teacher, then one would expect that classes with increase in SIMCE math then would also have increase in SIMCE language and SIMCE science. Conversely, if there is no clear relation in the increase or decrease across subjects then the increase in mathematics is not only effect of the teacher. In Lo Prado the increase in SIMCE math was 19.15, which is much greater than the increase in SIMCE science SIMCE which was 9.66 (0.19 standard deviations) and the increase in SIMCE language that was -3.35 (0.07 standard deviations). It is important to note that the online exercise strategy was implemented only in mathematics. In language three year ago the municipality implemented a different strategy not using ICT and more recently it implemented another strategy in science but also not using ICT. Furthermore, we tested the statistical independence of changes or if there is any correlation in the 15 classes between the behavior of SIMCE math with SIMCE language and SIMCE science. As shown in Table 1, the Pearson correlation test showed positive and statistically significant correlations between the change in math and language, and between language and science, but not between mathematics and science. These tests ruled out that the possibility that the rise in SIMCE math is due entirely to teacher effect.

Table 1. Correlations between language, math and science variations.

		language_variation	math_variation	science_variation
language_variation	Pearson Correlation	1	,651**	,777**
	Sig. (2-tailed)		,009	,001
	N	15	15	14
math_variation	Pearson Correlation	,651**	1	,334
	Sig. (2-tailed)	,009		,243
	N	15	15	14
science_variation	Pearson Correlation	,777**	,334	1
	Sig. (2-tailed)	,001	,243	
	N	14	14	14

**. Correlation is significant at the 0.01 level (2-tailed).

It is also important to see the effect of the lab teacher in charge of the laboratory. Two lab teachers took charge of the lab on the 11 schools, moving from one to another to attend classes during sessions. The variation in mathematics SIMCE achieved by one lab teacher was 20.14 points which corresponds to 7 classes, while the increase achieved by the other lab teacher was 18 points, corresponding to 8 classes. The t-test for independent samples reveals that we cannot reject the equality of means (N=15 classes). That is, the SIMCE math increase is statistically similar between the two lab teachers.

While the increase in SIMCE math is not explained solely by teacher or lab teacher effect, it remains to be analyzed what may have caused it. As in all schools in the Lo Prado district the model was implemented with one or two sessions per week, it is expected that the amount of exercises attempted had an effect on learning. Figure 2 shows the variation of SIMCE math in the 15 classes of the 11 schools of the district, as function of the number of online exercises with immediate feedback between end of March and early October 2011. In total, there were 185, 413 such exercises. Exercises with errors in their statements were discounted. Also were not counted exercises answered by a student in less than three seconds and in less time than the average response time minus two standard deviations. As seen in the chart there is a positive relation between number of exercises and increase in SIMCE math. The increase is of 15.3 points in SIMCE math per 100 extra exercises per student performed online. This means that for every 12 additional exercises done per month per student, the SIMCE math increases 15.3 points (0.306 standard deviations). The result is statistically significant with a p-value of 0.029.

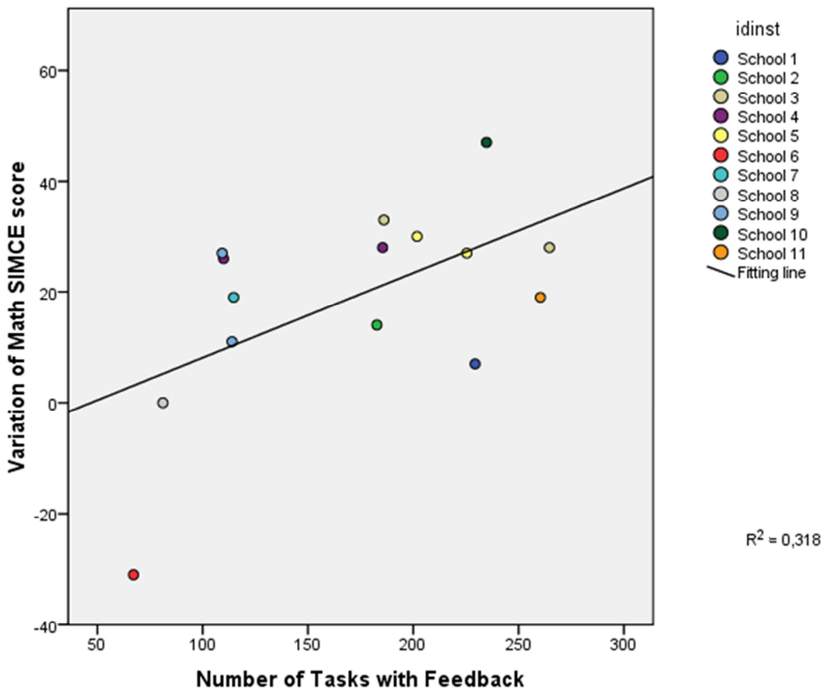


Figure 2: Variation on the 2011 SIMCE Math score of each class (N=15 classes) in function of the number of on line exercises with feedback made per student during the year from end of March to early October 2011.

4 Discussion

We have measured the impact of the adoption of a blended model focused on ensuring intensive student practice and on providing real time support from teachers and peers in all the 11 public schools in a low socioeconomic district. There are 203,782 records corresponding to the trace of on line math exercises done in 2011 by the 469 students in all the 15 fourth grades classes on these schools. The results are very promising and encouraging. First, there is a real measurable impact. The effect has been obtained not only by internal tests, but using the results of the national assessment test SIMCE math. This is a nationwide test, designed and administered by a completely independent team belonging to the Ministry of Education. The impact is very big. The increase was 0.38 standard deviations. This is more than three times the big and historic country wide increase on the SIMCE math of the year 2011. These results are in addition to differences in teachers. Moreover, there is a clear correlation between improvement and intensity of use of the ICT model. Classes that did more exercises with immediate feedback per student in the year they increased more their SIMCE math scores. This result agrees with studies in laboratory conditions, but now the results are obtained after a year of implementation in real schools and with low socioeconomic status students.

Acknowledgements: Conicyt CIE-05 Centro de Investigación Avanzada en Educación, and Fondef TIC EDU TE10I001.

References

1. Araya, R. & Gormaz, R. (2012) How come educational inequality in Cuba is much higher than in Chile? CIAE publication http://www.ciae.uchile.cl/index.php?page=view_publicacion&id_publicaciones=252
2. Hattie, J. & Timperley, H. (2007) *The Power of Feedback, Review of Educational Research. Vol 77, No. 1, pp. 81-112.*
3. Karpicke, J. & Blunt, J. (2011) Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science, 20 January.*
4. Mayfield, K. & Chase, P. (2002) The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis, 35, 105-123*
5. Ministerio de Educación (2012) *Resultados para Docentes y Directivos.* http://www.simce.cl/fileadmin/Repositorio_4B/2011/Docentes_y_Directivos/DOC_DIR_2011_4B_RBD-10089.pdf
6. Pyc, M. & Rawson, K. (2010) Why Testing Improves Memory: Mediator Effectiveness Hypothesis. *Science, 15 October 2010: 335 DOI: 10.1126/science.1191465*
7. Roediger, H. & Karpicke, J. (2006) Test-Enhanced Learning: Taking Memory Tests Improves Long Term Retention. *Psychological Science.*
8. Rohrer, D. & Pashler, H. (2007) Increasing Retention without Increasing Study Time. *Current Directions in Psychological Science*
9. U.S. Department of Education. (2009) *Effectiveness of Reading and Mathematics Software Products Findings from Two Student Cohorts.* Institute of Education Sciences National Center for Education Evaluation and Regional Assistance. February 2009
10. U.S. Department of Education. (2009) *Using Student Achievement Data to Support Instructional Decision Making.*
11. UNDP (2013). *Human Development Report 2013. The Rise of the South: Human Progress in a Diverse World.*
12. UNESCO (2001) *Primer estudio Internacional Comparativo sobre Lenguaje, Matemática y Factores asociados para alumnos de tercer y cuarto grado de Educación Básica.*
13. UNESCO (2008) *A view inside primary schools: A world education indicators (wei) cross-national study.* Technical Report.
14. UNESCO (2010) *Factores asociados al logro cognitivo de los estudiantes de América Latina y el Caribe.* Santiago.
15. UNESCO (2013) *A View Inside Primary Schools. A World Education Indicators (WEI) cross-national study.*
16. OECD (2010*) PISA 2009 Results: Overcoming Social Background: Equity in Learning Opportunities and Outcomes. Volume II*
17. Ministerio de Educación (2011) *Resultados PISA 2009 Chile*
18. Wims, P. and Lawler, M. (2007) Investing in ICTs in educational institutions in developing countries: An evaluation of their impact in Kenya. *International Journal of Education and Development using ICT* [Online], 3(1), http://ijedict.dec.uwi.edu/viewarticle.php?id=241

19. GeSCI (2011) *ICT, Education, Development, and the Knowledge Society*. Retrieved from http://www.gesci.org/assets/files/ICT,%20Education,%20Development,%20and%20the%20Knowledge%20Society(1).pdf

Children Creating Pedagogical Avatars: Cross-cultural Differences in Drawings and Language

Melissa-Sue John[1], Ivon Arroyo[1], Imran Zualkernan[2], Beverly P. Woolf [3]

[1] Social Sciences and Policy Studies, Worcester Polytechnic Institute
[2] Computer Science and Engineering, United Arab Emirates
[3] School of Computer Science, University of Massachusetts Amherst

iarroyo@wpi.edu, mjohn@wpi.edu, izualkernan@aus.edu, bev@cs.umass.edu

Abstract. This research identifies cultural differences among children's drawings especially as related to their drawings of avatars for instructional software. We invited children to draw characters and textual messages within an instructional game, as a way to establish their expectations of pedagogical avatars. We were interested in both the appearance and language of the characters of different nationalities. We describe an experiment that evaluated cultural differences in children's drawings. We analyzed drawings produced by 57 children aged 7-10 from four countries and discovered several main effects. Specifically, a significant main effect was found for a child's nationality and gender in predicting the emotion, formality of language, and use of "polite" or nice language. Girls generally expected more details in the hair, skin and facial hair of their characters and drew more emotions (positive) into their characters. Additionally, Pakistani and Argentine boys drew more details and more head coverings than did other children. Girls from Pakistan drew fantasy figures, rather than realistic figures and did not draw headscarves on their characters. The level of detail expected in the characters varied by country.

Keywords: Developing World, Pedagogical Agents, Children's Drawings, Localization

1 Introduction

Pedagogical agents used within adaptive learning environments have provided great benefit for learners as indicated by research over the last few decades (Lester, et al., 2004, Blair, Schwartz, Biswas, & Leelawong, 2006; Biswas, Schwartz, Leelawong, Vye, & TAG-V, 2005). Pedagogical agents are effective tools to support student learning; they provide motivation for learning and promote positive affective states (Arroyo, Woolf, Royer, & Tai, 2009). Results have shown that students are extremely sensitive to the appearance and gender of the characters reacting in different ways, and being more or less productive depending on the character's appearance. In a series of studies, students responded more positively when the gender of the character matched the gender of the student (Arroyo et al., in press).

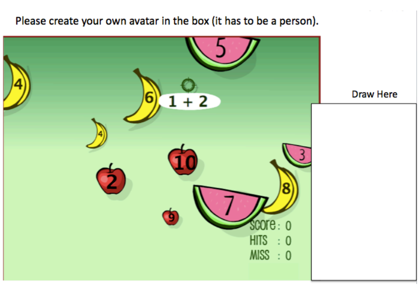
When considering the migration of educational systems and learning environments to other countries, it is unclear whether pedagogical agents would work in a similar way for students of developing countries. Should agents mimic the gestures and even dress codes of students in different countries, or is this localization effort beyond translation unnecessary? Are there differences in the style of language that pedagogical agents should use to communicate with students of different nations?

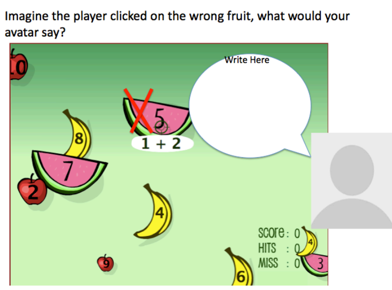
As a way to measure ecological validity, we decided to carry out an experiment that "taps into" children's minds and their expectations for pedagogical agents. We asked students to create their own pedagogical agents or avatars that would guide them through a mathematics learning game. The following article describes an experiment across four different countries in different continents, summarizes results and draws conclusions about the way to move forward to identify children's cross-cultural differences in expectations for a helpful avatar.

1.1 Background and Related Work

Having children draw as a way to mirror what is in their minds is a common technique used in psychology. Research into children's drawings has focused on three main areas: (a) the internal structure and visual realism of children's depictions (e.g., Cox, 1992); (b) the perceptual, cognitive, and motor processes involved in producing a drawing (e.g., Freeman, 1980); and (c) the reliability and validity of the interpretation of children's drawings (e.g., Hammer, 1997). Very young children produce simple scribbles, and later demonstrate representational intentions. With maturation and increased dexterity, children draw objects as they are known rather than as they are actually perceived.

Drawings of the human figure can also reflect a child's social world. La Voy and colleagues (2001) explored the idea that children from different cultural backgrounds may represent cultural differences in their drawings, because culture permeates a child's representations of people. Differences across nations indicated that American children drew more smiles than Japanese children, whom in turn drew more details as well as larger figures (La Voy et al., 2001). Similarly, Case and Okamoto (1996)





Figures1-2. A simple addition math game for younger children. Children were invited to supply a drawing for an avatar (left) and then to provide the responses the avatar might provide when the student player chose the wrong mathematics answer.

showed that there are cultural differences between Chinese and Canadian children's drawings. These findings suggest that children's drawings not only reflect representational development but a child's understanding of self and culture as well.

Having students draw characters and games, as a way to tap into their minds and establish their expectations of pedagogical characters and games is an increasingly common technique, and has particularly been implemented for learning systems/games for mathematics education. For instance, Grawemeyer and colleagues (2012) managed to have participants within the autism spectrum express and externalize their individual ideas for an educational pedagogical agent for a mathematics educational game, and to combine their individual ideas with the ideas of others in a small group. Students created their own designs and also studied other students' drawings, eventually creating a common prototype.

The outcome of one of the small groups was quite different from the norm: these children with autism designed characters, such that the student would be sitting at the back seat of a car, being able to view two avatars sitting in the front seat, from the view of the person in the back seat. Instead of showing the avatar facing forward and expressing emotions through its facial expressions, as has commonly been done in the past, the avatars (shown from the back) would have a conversation about the student's learning and progress, as children might interact with their parents when traveling at the back seat of the car. Thus these students with autism expressed their own distaste for talking directly to at people or looking into their eyes. It is assumed that an avatar designed for a typical student would promote better communication if it looked directly at the student.

Other studies have invited children to design and draw full math games, which generally included characters, human or not. For instance, Kafai (1996) invited fourth grade children to design mathematics games for younger children. Her study, identified important gender differences in the design of games. In general, boys were more likely to use fantasy locations in their games (instead of real life locations, such as a sky slope), and also were more likely to have the presence of evil characters, or the idea that an avatar would fight some evil force.

2 The study

Our study involved children invited to draw characters, avatars or pedagogical learning companions to keep student players company as they used a game to learn mathematics. The goal was not to ask for complex representations, but instead, and similar to La Voy and colleagues (2001) to explore cultural differences that are important to understand for authors of creating pedagogical avatars.

,

Children from North America Argentina, Pakistan, and Jamaica, aged 7-10 were asked to draw characters they thought would help younger support as they played a mathematics game for younger children. Children were given a printed package of 6 pages. On page 1, students were told "Help us design this math game! We are designing computer based math games for younger children. Can you help us?" On the second page, a screenshot of a simple addition math game, shown in Figure 1, where student players would click on the fruit with the right answer is shown and at the top reads "This is a picture of a math game. In this game, children will learn to add. Using the mouse, they have to click on the fruit with the right answer." The children were invited to provide a voice for their avatar by providing a response that the avatar might produce in response to a student player's incorrect answer, see Figure 2. And finally parents and teachers were instructed to complete the student demographics (age, ethnicity, nationality and gender).

We obtained drawings from 57 children from North America (14), Pakistan (11), Jamaica (18) and Argentina (14). Of these children, 30 were girls and 22 were male, mean age was 8.19 (SD = 1.42). We were interested in both the appearance and language of characters developed by these students of different nationalities.

Properties of avatar
1. Realism (Human / Fictional)
2. Gender (F / M / Unspecified)
3. Age (Child / Teen / Adult / Unspecified)
4. Details (+ 1 for each of these: body, eyes, nose, mouth, dimples/freckles, ears, teeth, hair, facial hair, head-covering, clothing, shoes, accessories, toys, skin-coloring)
5. Affect (Happy / Neutral / Sad / Angry)

Voice of avatar
7. Tone of incorrect answer (Polite/encouraging or Direct/Straightforward or rude/aggressive/discouraging)
8. Formality of incorrect answer (formal/neutral/informal)
9. Tone of correct answer (Polite/encouraging or Direct/Straightforward or rude/aggressive/discouraging)
10. Formality of correct answer (formal/neutral/informal)

Characteristics of Participant
11. Language spoken/written (English/Spanish/Pashto)
12. Videogame exposure (Do you play videogames?)
13. Have you ever **used** an avatar?
14. Have you ever **created** an avatar?
15. Age Student
16. Gender: Female (1) Male (2)
17. Ethnicity
18. Nationality

Table 1. Features of the study to analyze cultural characteristics of children's drawings. Properties of the avatar, voice of the avatar, and characteristics of the student, were analyzed to explore cultural differences.

3 Results

Although children were asked to create math avatars that looked like people, children came up with humanoid and non-humanoid images. In one study in particular, it was not clear that students had understood that we meant "characters that look like humans". Thus, for the purpose of our analyses, we only coded humanoid images (see Figure 3).

Two different human coders analyzed the pictures and messages to respond to correct/incorrect answers from student players. They coded the variables described in Table 1. Because many of these metrics might be somewhat subjective, we had two coders separately. After coding was done, we computed Kappa to analyze agreement between the coders. Whenever a variable had a Kappa value less than 0.5, we reconsidered the variable and came up with a new coding scheme. The variable was recoded and the process repeated. Variables with very low Kappas were dropped from the analysis (e.g., age of the avatar). We then carried out Analysis of Variance with the variable of interest, and nationality, gender-child as fixed factors, with age of child as a covariate. In the case of discrete variables, we ran cross-tabulations and Chi-Dquare tests. Results indicate the following.

Gender of Avatars. A significant difference was found for child's gender ($\aleph^2$ $=38.9$, $p<0.001$) and gender of the avatar, showing that most children drew characters of their same gender. No significant differences were found for nationality. Only a minority of children drew characters of unidentifiable gender.

Realism of Avatar. A significant interaction effect between gender of the student and nationality ($F=3.9$, $p<0.015$) showed that Pakistani girls drew more fantasy characters than did children from other countries, or than boys of the same country (see Figure 4).

Level of Detail. A significant main effect was found for a child's nationality in predicting level of detail of the characters ($F=3.6$, $p<0.02$). Students from Pakistan and Argentina drew more details than did children from the United States or Jamaica, regardless of their gender, two more features from Table 1 (see *4. Details*) on average. Further analyses showed differences in the amount of <u>head-coverings</u>, particularly drawn by boys in general ($F=13.6$, $p<0.001$), and for Pakistani and Argentine boys in particular ($F=13.6$, $p<0.12$), who drew more headcoverings. While we expected girls

Figure 3. A selection of avatars drawn by children in different cultures as companion for a proposed math game.

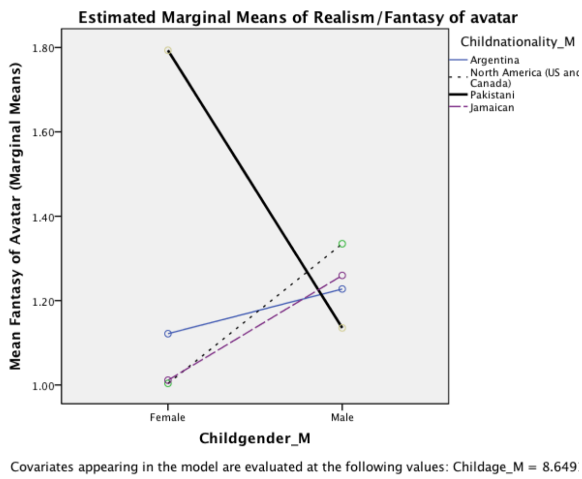


Figure 4. Girls from Pakistan drew more fantasy characters than did other children

from Pakistan to draw headscarves, they did not –in fact they tended to not draw images of real people but drew fantasy figures from other cultures such as princesses. The head accessories that boys drew were actually hats.

Another difference had to do with the drawing of clothes –children from the United States drew the least detailed clothes on their avatars (F=3.5, p<0.01). At the same time, students from Jamaica drew more hair on their characters' heads, and students from Argentina drew the most facial hair on their avatars. Meanwhile, girls in general drew sigificantly more hair on their avatars, both on the characters head (F=11.17, p<0.02) and more facial hair details (note this included eye-brows, eye-lashes, moustache, etc.) (F=8.2, p<0.001). Girls also drew more details on the skin (e.g. freckles, dimples, tatoos, etc.) (F=4.5, p<0.04). No significant differences were found in the amount of accessories used, the kind of accessories, nor in the presence of shoes, noses, eyes, nor bodies. Most students drew all of these, mostly full-bodied avatars instead of heads, and the amount of accessories did not have a consistent differential pattern across nations or genders.

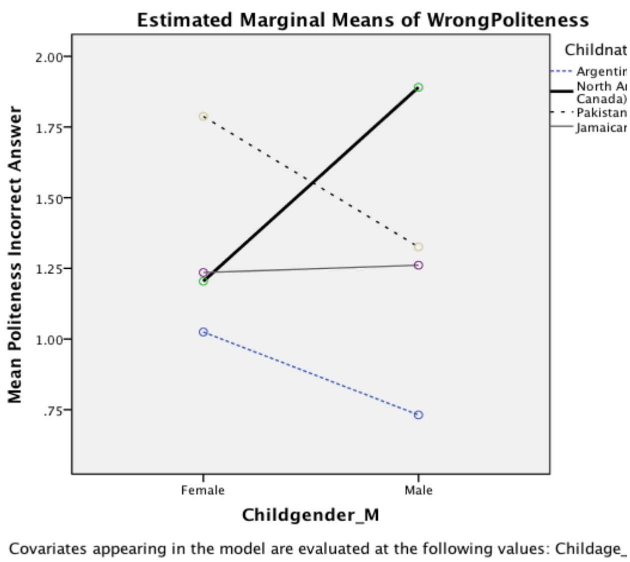


Figure 5. Gender differences in politeness of the avatar's response to incorrect answers from student player

Emotions. A significant main effect was found for emotions expressed by avatars for girls and boys. Girls across nations were more likely to draw avatars with happy faces, with boys evenly split between happy and neutral faces (gender effect, F=9.8, p<0.003) and a minority of children drew angry/agressive emotion in their characters, 5% of all students, all three were boys instead of girls.

Formality of Language. A significant main effect was found for a child's nationality predicting the formality of language for the avatars response to incorrect or correct answers from student players in the game (F=9.7, p<0.001). Students from the United States used more informal language than did students from Jamaica and Argentina (e.g. nope for "no", awesome), and children from Argentina used the most formal language (i.e. least informal language) in their answers than Jamaica, United States and Pakistan.

Tone of Language. Significant effects were found across countries for the avatars' answer after an *incorrect* answer from a student player, where as no significant differences existed for having the character express a response after a player's *correct* answer. A significant main effect for nationality (F=3.3, p<0.03) showed that students from Argentina used the least "polite" language as compared with students from other countries (e.g., least use of words such as *sorry*, *please*, *thank you* etc.), with children from the United States and Pakistan using the most polite language. Interestingly, there was an interaction effect between gender of the child and nationality (F=2.9, p<0.05), which indicated gender differences in the tone of the avatar's response to incorrect answer for children of different genders. Actually, boys' avatars from the United States used more polite language than girls' avatars from the same country, despite the fact that the appearances of U.S. boys' tended to be more aggressive than girls' (see examples in Figure 3 drawings); the reverse happened for Pakistan, where girls' avatars used more polite language than boys' (see Figure 5).

4 Discussion

Some research articles have claimed that children's drawings are a mirror to children's minds (Cherney et al, 2006). In light of this, what do these results imply in terms of the creation of pedagogical agents, and the translation of adaptive learning systems to fit new countries, after some important differences in the look and conversation of children's pedagogical agents? If we consider that what children draw is what they expect, value, and desire, the findings suggest that children, regardless of country, expect characters to be of their same gender. This is consistent with our prior findings (Arroyo et al, 2013), which indicated that matching the gender of the student with the character's gender led to improved affective, behavioral and learning outcomes, such as engagement and reduced frustration. Girls also expect more details in their character's hair, skin and facial hair. Boys might want to have more head coverings, particularly hats. Also, girls from Pakistan might prefer fantasy figures instead of figures that depict themselves. Lastly, the fact that girls in general drew more emotions (positive) on their characters could suggest an expectation of girl's avatars to emote and act affectively –however, this needs to be examined further.

It does seem important that the level of detail expected in the characters will vary by country. Children from Argentina and Pakistan might expect more level of detail in their characters than do students of the United States or Jamaica, e.g. clothes and hair. Meanwhile, differences across countries are especially marked in the *kind* of language to be used when the characters talk, specifically when student playes produce incorrect answers, with Argentine children apparently expecting the least politeness. Expectations of politeness and niceness of the language can be explained by cultural differences. People in Argentina are very straightforward in their dialog (similar to European countries such as France, Italy or Spain) and do not excuse themselves so much in their daily interactions. This is something that needs to be examined when designing characters that communicate with students, even if the communication is in the form of text and not voice. This would potentially argue against a mere translation from English to Spanish, where such polite words might show up. Differences in formality of the language between Argentina and the United States could be explained by the fact that the language might not lend itself to informal distortion of words such as “nope”.

5 Conclusions and future work

Large differences were observed in children’s design of pedagogical agents across a variety of dimensions, but probably in different areas than we had originally expected. . Differences were present across countries, across gender, and across country and gender. Main differences were in language of incorrect answer across countries, and in the look of characters, both across countries and genders. These differences span across the visual appearance of pedagogical agents as well as in the language used to communicate to student players. From a methodological point of view, having children design pedagogical agents by having the freedom to draw and create, can act as a mirror to their minds and help researchers to externalize their expectations.

The main limitation of this study has to do with the total amount of subjects available, which is not representative of different socio-economic levels of each country, as well as a lack of representation in terms of ethnicities in each country, Future work will consist on a larger study, with a much larger number of students.

Acknowledgements

This research was funded by an award from the National Science Foundation, NSF # 1109642, REESE, Personalized learning: strategies to respond to distress and promote success, Ivon Arroyo (PI), with Beverly Woolf and Winslow Burleson. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

References

Arroyo, I., Burleson, W., Tai, M., Muldner, K., Woolf, B.P. (in press) Gender Differences In the Use and Benefit of Advanced Learning Technologies for Mathematics. Journal of Educational Psychology

Arroyo, I.; Woolf, B.P., Cooper, D.G., Burleson, W., Muldner, K. (2011). The Impact of Animated Pedagogical Agents on Girls' and Boys' Emotions, Attitudes, Behaviors and Learning. ICALT 2011, IEEE's International Conference on Advanced Learning Technologies, Athens, GA.

Blair, K., Schwartz, D.L., Buswas, G., & Leelawong, K. (2006). Pedagogical agents for learning by teaching: Teachable agents. Educational Technology & Society, Special Issue on Pedagogical Agents.

Biswas, G., Schwartz, D. L., Leelawong, K., Vye, N., & TAG-V. (2005). Learning by teaching: A new paradigm for educational software. Applied Artificial Intelligence, 19(3).

Case R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. Monographs of the Society for Research in Child Development, 246(61), 1–2.

Cherney, I. D., Seiwert, C. S., Dickey, T. M. & Flichtbeil, J.D. (2006) Children's drawings: A mirror to their minds. Educational psychology: An International Journal of Experimental Educational Psychology, 26(1).

Cox, M. V. (1992). Children's drawings. Harmondsworth, UK: Penguin

Grawemeyer, B., Johnson, H., Brosnan, M., Ashwin, E., Benton, L. (2012) Developing an Embodied Pedagogical Agent with and for Young People with Autism Spectrum Disorder. Intelligent Tutoring Systems 2012. Springer.

Freeman, N. H. (1980). Strategies of representation in young children: Analysis of spatial skills and drawing processes. London: Academic Press.

Hammer, E. F. (1997). Advances in projective drawing interpretation. Springfield, IL: Thomas

Kafai, Y. B. (1996). Gender differences in children's constructions of video games. In Patricia M. Greenfield & Rodney R. Cocking (Eds.), Interacting with video (pp. 39–66). Norwood, NJ: Ablex Publishing Corporation.

La Voy, S. K., Pederson, W. C., Reitz, J. M., Brauch, A. A., Luxenberg, T. M., & Nofisnger, C. C. (2001). Children's drawings: A cross-cultural analysis from Japan and the United States. School Psychology International, 22, 53–63.

Lester, J., Branting, K., and Mott, B., (2004) Conversational agents. The Practical Handbook of Internet Computing

Comparing Paradigms for AIED in ICT4D: Classroom, Institutional, and Informal

Benjamin D. Nye

Institute for Intelligent Systems, University of Memphis
Memphis, TN 38111
benjamin.nye@gmail.com

Abstract. The landscape of technology in the developing world is changing significantly, primarily due to the rapid expansion of mobile computing devices. These changes make it important to re-evaluate practices for internet and communications technology for development (ICT4D). This paper examines three alternative paradigms for educational technology in the developing world: traditional classroom systems, institution-wide systems, and informal learning systems. The advantages and disadvantages of each paradigm are considered in terms of barriers to adoption at the student, teacher, and institutional level. Consideration is also given to educational technologies that serve as models for each type.

Keywords: Educational Technology, Barriers to Adoption, Informal Learning, Ubiquitous Learning

1 Introduction

As access to Information and Communications Technology (ICT) expands through the developing world, educational technology has the potential to play a pivotal role for supporting development. However, successful paradigms for incorporating ICT into developing world education are less clear. Educational technology in the developing world has an uneven history that includes numerous wasted investments in underutilized computers and limited learning benefits (Patra et al., 2007; Woolf et al., 2011). Moreover, the landscape of ICT in the developing world is changing drastically due to the rise of mobile handsets and wireless Internet access (International Telecommunication Union, 2012). These changes offer new opportunities, but also present new obstacles.

Research on advanced intelligence in education (AIED), such as adaptive learning systems, intelligent tutoring systems, and computer-supported collaborative systems, needs to outline the tradeoffs between different application contexts (e.g., classroom, institution-wide, and informal) to help select appropriate system designs. In this paper, these tradeoffs are framed as factors that mediate adoption of ICT, as indicated in Table 1. These factors are based on known barriers to information and communications technology that were identified from recent review papers (Gulati, 2008; Lowther et al., 2008; Bingimlas, 2009). Different paradigms have advantages and disadvantages for each factor.

Table 1. Factors Impacting Adoption of Educational Technology

System	Requirement/Possible Barrier	Description
Learner	Basic ICT skills	Computer literacy and familiarity with basic interfaces
	Independent access to ICT	Web access or computing outside of school
	Motivation to use ICT	Student interest and persistence in use
	Peer support	Peer help or collaboration
Teacher	Basic ICT skills	Computer literacy and managing applications
	Beliefs about utility of ICT	Values and expectations for an ICT design
	ICT-integrated curricula	Pre-made curricula and syllabi that incorporate an ICT intervention
	Match to pedagogical views	Match of teacher pedagogy to an ICT design
	Peer support	Communities of practice and peer views
	Time constraints	Class and preparation time available
	Training (e.g., in-service)	Training with a given ICT design
School or Institution	Administrative support	Administrative needs, reactions, and leadership toward ICT use
	Curriculum flexibility	Flexibility to modify teaching to use ICT
	ICT hardware availability	School web access and computing hardware
	Technical support	Technical staff to set up and maintain ICT
	Internet reliability	Stable, reliable internet connections

This paper considers three paradigms that have shown promise in the developing world, discussing successes and potential challenges. These paradigms will be framed in terms of the context where they are utilized: under *classroom* control, around the entire *institution* (e.g., through a central learning management system), or outside the educational system in an *informal* learning context. While these are not the only approaches (nor are they exclusive), each offers distinct strengths and challenges. Each paradigm will be briefly discussed, with attention to the barriers to sustainability noted in Table 1 and also to promising implementations that embody each approach.

2 Traditional Paradigm: ICT Under Classroom Control

The traditional paradigm for educational technology in the developing world has been classroom-centric (Gulati, 2008). The typical design sets up classroom computers or shared computer labs with educational software. In this context, educational technology is a tool that teachers use to improve learning for students. Classroom-based tools are typically tailored to domain (e.g., Algebra I) and require less flexibility than a general learning management system (LMS).

Classroom-centric ICT has many advantages when compared to other approaches. First, the classroom setting gives the teacher a significant degree of control over students to mandate and manage the use of the system by students. In a classroom setting, basic ICT skills are not typically a blocking issue as students often learn controls quickly and students with more advanced ICT skills may even help the teacher (Gulati, 2008; Ogan et al., 2012). As such, the high availability peer support mitigates deficits in basic ICT skills. Students also do not need to own personal computing devices. The motivation of students, while still important, is less critical than in other contexts. Research has found that liking a system does not necessarily correlate with learning gains, provided students still use the system as intended (Moreno et al., 2002). In a classroom, most students will do assigned work even if they do not find it interesting.

Second, the primary buy-in occurs at the teacher level. At least for initial evaluations, this mitigates many barriers related to teachers. Given that teachers have very different attitudes to technology (Lowther et al., 2008), the ability to pair up a system with technologically-receptive teachers greatly increases the likelihood of successful usage. Teacher beliefs about ICT, match to pedagogical views, and these teachers' basic ICT skills are likely to be better than average. One barrier not mitigated by this approach is peer support, as few teachers will be using the system. Additionally, scaling up to widespread use will hit these barriers once the supply of early-adopters is exhausted. Persuading uninterested teachers to adopt technology is unlikely, unless institutional entities encourage its use. So then, while this paradigm is useful for pilot testing and establishing a foothold, there may be limits to its scale.

The clear point of failure for a classroom-centric approach is institutional factors. If buy-in is primarily at the teacher level rather than the administration level, there is no assurance that the larger institutional context will offer a sustainable environment for that educational technology. If educational technology is a low priority, teachers may be pressured to focus on other matters and technical support may be unavailable. Inflexible mandatory curricula may also make it impossible to work technology into classrooms. Alternatively, curricula dedicated to computers may focus exclusively on digital literacy (e.g., learning about computers) rather than using computers to learn a broader range of topics.

Most importantly, ICT hardware depends on financial support. Investment in computers must be made at the institutional level, but developing world schools often lack the funding to support heavy investment into purchasing, managing, and replacing hardware. Low ratios of students to computers can make meaningful computing curriculum infeasible. Accessing and financing reliable Internet may also be out of the control of the school system. Many developing world areas still have unreliable electrical and Internet infrastructure, which can easily fail and derail any instructional plan relying on web connectivity (Woolf et al., 2011). So then, the primary barriers to traditional classroom ICT are at the school and institutional level. Thankfully, strong focus has been placed on overcoming hardware barriers for ICT in schools. Irregular electricity can be mitigated by using laptops, as their batteries make them immune to short power losses. Irregular Internet can be sidestepped by installing from disk media or only depending on Internet infrequently, rather than during classroom time. Pilots of Cognitive Tutor in Latin America installed software on desktops and did not note significant roadblocks due to the unreliable Internet available (Ogan et al., 2012). By implication, web-based tutoring portals are poorly-suited for the developing world classrooms. This is unfortunate, since educational technology in the developed world has moved strongly in this direction.

Two approaches have been used to overcome hardware barriers: cheaper devices and shared computing. The One Laptop Per Child program spearheaded the "cheaper hardware" approach, driving down the base cost of computers overall (Patra et al., 2007). However, this approach encountered two problems. First, even with lower costs, many schools cannot afford a laptop for every child. Second, studies on ICT interventions in the developing world find that students *prefer* to share computers (Ogan et al., 2012). As such, a number of systems have adapted to this landscape and offer one mouse or keyboard per child (Alcoholado et al., 2012; Brunskill et al., 2010), collaborative turn-taking, and other methods of individual input into a shared learning environment such as mobile devices (Kumar et al., 2012) or wireless clickers (Zualkernan, 2011). Individual inputs are inexpensive compared to computers, greatly reducing hardware costs. Additionally, these techniques complement cheaper computers since they have a multiplier effect. Computer sharing also offers greater pedagogical flexibility, since interactions with other students enable social constructivist designs that would be difficult in a single-user system.

MultiLearn+ offered one model for such a multi-input system, presenting a math game split into four quadrants on a laptop screen and supplying each student a numeric keypad (Brunskill et al., 2010). To prevent dominance by a single student, MultiLearn adapted the difficulty of questions based on student performance. This system relied on installed software, with no Internet component. At present, a laptop with educational technology designed to be shared by four or five students may be the best model for ICT in a primary or secondary school classroom. Such a system might use Internet to update the system, but cannot assume Internet will be available during a classroom session. While significant work has been done in this area, there are still many questions over the relative advantages of different presentation devices (e.g., laptop screens, projectors, voice narrative) and input devices (e.g., mice, keyboards, voice recognition, game controllers, clickers/remotes). In particular, shared mobile computing might be a transformative technology in the future. For example, Kumar et al. (2012) presented a mobile learning tutoring system based on voice recognition and suggested the potential for shared computing through voice identification. While this particular paradigm may encounter technical hurdles, computer sharing for mobiles is an important avenue that needs further research.

3 Institutional Paradigm: ICT Around the School

In a related paradigm, the institution controls a learning management system from the top down. The institution may be a school, district, or even a national system. Learning management systems (LMS) primarily provide a container and delivery platform for static media, though assessments, adaptive learning systems, collaborative systems, or tutoring systems may be incorporated. These systems can support both traditional and online classes. Worldwide, this is more common within higher education. Ubiquitous systems, which connect a variety of devices to a central system, also require an institutional paradigm.

Institutionally-centered systems have similar pros and cons with respect to student barriers, since an instructor usually guides a group of students. One advantage is that, since students interact with a shared central system, remote peer support is possible (e.g., a forum, Wiki, or social media). Unlike classroom-centered systems, institutional systems typically require each student to own a personal computing device. This is because primary use cases of LMS and ubiquitous learning are web-based homework and remote collaboration. However, teachers are the most affected by this paradigm, who will often need to redesign their curricula to fit the system. While an opt-in single classroom paradigm hides teacher barriers by excluding the most resistant or inadequately-prepared teachers, institution-wide adoption hits these barriers head-on. Institution-level barriers are also still an issue. While buy-in by the institution should increase administrative and technical support, hardware costs remain an issue. Since an LMS requires both servers and personal computing hardware, centralized institutional paradigms are more hardware intensive and more costly as a result.

A few designs have attempted to overcome these limitations. EDUCA, a ubiquitous learning platform, provides an LMS and tutoring system capabilities that can be accessed asynchronously over the web through a desktop or a mobile device (Cabada et al., 2011). Entire learning modules are downloaded to the mobile device, as well as an adaptive system for personalizing learning. Since mobile Internet is more prevalent than wired Internet in the developing world, this helps students access the system without a home computer. However, as Mexico is an "emerging market," this approach still may not translate to less developed countries with worse wireless infrastructure. An alternative approach enabled mobile devices to communicate with the school network over mobile web or through "learning pills" transferred to the student's phone during class over Bluetooth (Munoz-Organero et al., 2012). However, both approaches require the student to own a web-capable phone and passes these costs down onto students. This approach seems better suited to higher education, where students can be responsible for the costs and ownership of mobile computing devices. However, as mobile web capability becomes commonplace, ubiquitous paradigms may also become relevant for primary and secondary education. In either case, any learning management system or large-scale institutional system for the developing world must support mobiles as first class, or even primary, devices.

4 Informal Paradigm: Technology Outside the School

A precise definition for informal learning is hard to pin down, as informal learning is often described in terms of how it differs from traditional schools. Within this paper, informal learning refers to education where students have no interactive human supervision and engage with learning materials based on their own initiative. The informal paradigm is attractive in some ways. School and teacher barriers are sidestepped, learning only student barriers. Computer-based informal learning was not previously a possibility in the developing world, but the spread of web-capable mobiles is changing this drastically. However, while informal learning offers strong appeal, it is likely a case of “the grass is always greener on the other side.” First, while students’ basic ICT skills were not a major problem in other contexts, studies have found that even setting up mobile Internet on phones can be an onerous task in the developing world (Gitau et al., 2010). So then, users probably need help from community centers or user groups to get started. Independent access to ICT is also required: students need a working phone or laptop with Internet capability. Informal usage also removes the constantly-available classroom peer group, limiting collaborative work and technical help. While students may naturally form study groups, the frequency and effectiveness of such emergent groups needs further study.

However, more so than any other factor, student motivation is an imposing barrier to the success of informal learning. In a traditional classroom, students can either do their work, sit idly, or incur punishment for performing off-task behavior. By comparison, informal learning environments compete with the Internet. Students need a high motivation toward the learning content to focus on an educational technology without a societal framework. No combination of teacher or school barriers may be more formidable than competing on a level playing field against the combined forces of the online media market. A pure informal paradigm may be an uphill battle. Informal learning technologies need find or create ecological niches that learners find useful and interesting.

One way to do this is to dominate a small ecological niche. One example of this is to preload devices with educational software. This paradigm relies on users trying out default programs first, rather than installing other programs. In less-developed areas, data may also be expensive enough to impose this barrier. A multi-week study on unsupervised use of preloaded educational games on mobile phones in India found that participants averaged of 2 hours and 23 minutes per week on the game, with 46 total hours per participant on average (Kumar et al., 2010). Possibly due to the game-based delivery, students had a fairly high level of motivation to learn. While off-task use was also present (e.g., downloading music), social dominance was a larger issue. Girls were particularly vulnerable, with brothers taking their phone and parents condoning this behavior. Additionally, software must be loaded onto devices at some point, so government or industry partnerships would be required for this to scale.

A second approach is to enhance existing niches, such as informal paradigms built around emergent communities of practice. For example, Mobile-ED offered a mobile gateway to a Wiki site where users could text a term and hear the web page read to them (Ford and Leinonen, 2009). If a term did not exist, users could dictate a definition that other users could use. Integrating web communities, which tend to be based on interests, and community organizations, which tend to be based on local ties, might drive sustainable informal learning, particularly on practical subject matter such as health, economic, or vocational competencies. Community groups can provide local motivational and technical peer support, as well as form connections with other user groups. By serving the shared needs of community groups, informal systems might benefit from grassroots support.

5 Conclusions and Future Directions

Classroom, institutional, and informal paradigms can each play a valuable role in developing world education immediately and in the future. While access to ICT is expanding, most of the developing world still has little computing hardware available. With that said, in raw numbers, the developing world has a strong demand for educational software that fits its needs. Primary and secondary school classrooms can benefit from shared computing applications today, through multiple-input laptops. In the future, single-display groupware or shared voice-input mobile devices might offer cheaper and equally effective designs. To make this jump, research on user interfaces for shared computing and group learning is essential. As technology evolves, regular research on these topics will be pivotal for keeping up with shifts in access and usage patterns.

Similarly, universities can immediately benefit from ubiquitous systems focused strongly on mobile learning. In the future, ubiquitous systems should be available at earlier grade levels as mobile computing expands and data costs fall. However, creating content for inexpensive mobile learning is non-trivial and many existing open systems, such as MIT Open Courseware (Abelson, 2008), are not well-suited for low resource contexts as they rely on rich media (e.g., streaming lectures). Research on methods to quickly convert existing content designed for high-resource computers (e.g., monitors, high bandwidth) to low-resource mobiles (e.g., small screen, speakers, low bandwidth) would be valuable. Techniques for rapid language and cultural localization may also be essential.

Finally, the role of informal learning in the developing world is still taking shape. Informal learning systems must target ecological niches created by technological and societal influences. While sustained engagement has been observed, social biases and gender barriers are reproduced in informal learning contexts (Kumar et al., 2010). Game-based learning and systems designed for community groups are two areas that may offer traction for supporting self-regulated education. Research on peer support and sustaining motivation for informal learning is essential, so that informal learning is both effective and equitable.

References

Abelson, H.: The creation of opencourseware at MIT. Journal of Science Education and Technology 17(2), 164–174 (May 2008)

Alcoholado, C., Nussbaum, M., Tagle, A., Gomez, F., Denardin, F., Susaeta, H., Villalta, M., Toyama, K.: One mouse per child: Interpersonal computer for individual arithmetic practice. Journal of Computer Assisted Learning 28(4), 295–309 (2012)

Bingimlas, K.: Barriers to the successful integration of ICT in teaching and learning environments: A review of the literature. Eurasia Journal of Mathematics, Science and Technology Education 5(3), 235–245 (2009)

Brunskill, E., Garg, S., Tseng, C., Findlater, L.: Evaluating an adaptive multi-user educational tool for low-resource environments. In: ICTD 2010. London, UK (2010)

Cabada, R.Z., Estrada, M.L.B., Parra, L.E., Garcia, C.A.R.: Interpreter for the deployment of intelligent tutoring systems in mobile devices. In: ICALT 2011. pp. 339–340. IEEE Press, Washington, DC (2011)

Ford, M., Leinonen, T.: Mobile-ED: Mobile tools and services platform for formal and informal learning. In: Ally, M. (ed.) Mobile Learning: Transforming the Delivery of Education and Training, pp. 195–214. AU Press (2009)

Gitau, S., Marsden, G., Donner, J.: After access: Challenges facing mobile-only internet users in the developing world. In: SIGCHI 2010. pp. 2603–06. ACM Press, New York, NY (2010)

Gulati, S.: Technology-enhanced learning in developing nations: A review. International Review of Research in Open and Distance Learning 9(1), 1–16 (2008)

International Telecommunication Union: Measuring the Information Society. Geneva, Switzerland (2012)

Kumar, A., Reddy, P., Tewari, A., Agrawal, R., Kam, M.: Improving literacy in developing countries using speech recognition-supported games on mobile devices. In: SIGCHI 2012. pp. 1149–1158. ACM Press (2012)

Kumar, A., Tewari, A., Shroff, G., Chittamuru, D., Kam, M., Canny, J.: An exploratory study of unsupervised mobile learning in rural india. In: SIGCHI 2010. pp. 743–752. ACM Press, New York, NY (Apr 2010)

Lowther, D.L., Inan, F.A., Strahl, J.D., Ross, S.M.: Does technology integration "work" when key barriers are removed? Educational Media International 45(3), 195–213 (2008)

Moreno, K.N., Klettke, B., Nibbaragandla, K., Graesser, A.C.: Perceived characteristics and pedagogical efficacy of animated conversational agents. In: Cerri, S.A., Gouardères, G., Paraguaçu, F. (eds.) Intelligent Tutoring Systems (ITS) 2002. LNCS, vol. 2363, pp. 963–971. Springer, Berlin (Jun 2002)

Munoz-Organero, M., Munoz-Merino, P.J., Kloos, C.D.: Sending learning pills to mobile devices in class to enhance student performance and motivation in network services configuration courses. IEEE Transactions on Education 55(1), 83–87 (2012)

Ogan, A., Walker, E., Baker, R.S., Rebolledo Mendez, G., Jimenez Castro, M., Laurentino, T., de Carvalho, A.: Collaboration in cognitive tutor use in Latin America: Field study and design recommendations. In: SIGCHI 2012. pp. 1381–1390. ACM Press, New York, NY (2012)

Patra, R., Pal, J., Nedevschi, S., Plauche, M., Pawar, U.: Usage models of classroom computing in developing regions. In: ICTD 2007. pp. 158–167. IEEE (Dec 2007)

Woolf, B.P., Arroyo, I., Zualkernan, I.A.: Education technology for the developing world. In: 2011 IEEE Global Humanitarian Technology Conference. pp. 493–498. IEEE (Oct 2011)

Zualkernan, I.A.: InfoCoral: Open-source hardware for low-cost, high-density concurrent simple response ubiquitous systems. In: ICALT 2011. pp. 638–639. IEEE Press (2011)

Applying Standards and AI to Educational Technology: The IEEE Actionable Data Book Project

Avron Barr, LETSI Foundation

Tyde Richards, IEEE Learning Technology Standards Committee

Dr. Robby Robson, Eduworks Corporation

1 Introduction

In the United States and other developed countries, new products built on emerging technologies such as tablets, mobile devices, cloud-based services and eBooks have generated widespread discussion about disruptive change in education at all levels. Typical questions raised include:

- Should the classroom be flipped using online video [1]?
- Can textbooks be replaced by open educational resources [2, 3]?
- What can children learn online on their own, and how can their families help?
- Can student advancement in school be tied to competence instead of cohort?
- Can a professor effectively teach 10,000 students at once in a MOOC [4-6]?
- Are automated assessments as good as human teachers [7-9]?

Although significant change is now occurring in the United States, especially in higher education, the potential for change and innovation may be even greater in the developing world. As has been demonstrated in mobile and Internet technologies, countries with less advanced infrastructures and fewer established policies and institutions can leapfrog the West in both quality of service and speed of deployment. In addition, developing countries have requirements and constraints that can lead to disruptive innovations that would not be developed in the West. An example of this may be found in the history of radio [10]. As the story goes, post-world war II Germany was given a very limited portion of the regulated radio spectrum. They therefore started using unregulated high frequencies. AM did not work well there, so they used FM. It turned out that FM had superior sound quality and became the dominant technology for quality radio.

In this context a central question is: *How can developed and developing countries collaborate to take advantage of the strengths of each?* In this paper we argue that fruitful collaboration can take place in the area of standardization and give a concrete example of how requirements from Bali spurred innovation and how standards activities in the area of eBooks may provide solutions. But first we examine in general terms the technological and related standards landscape that is emerging in eLearning.

adfa, p. 1, 2011.

2 Changes in eLearning Infrastructure

Today, commercial eLearning sales in the United States are dominated by two product categories, "content" (e.g. course packs and supplements to textbooks) and learning management systems (LMS). According to the Campus Computing Survey, about half of higher education institutions used an LMS in 2007 [11]. By 2011 not only did virtually all universities use an LMS [12], but only 7% had not standardized on a *single* institutional LMS [13]. From an institutional, teacher and student perspective the LMS is responsible for:

- Managing student credentials and class rosters
- Tracking entitlements to publisher content that is delivered by the LMS
- Recording student activity, task completion, and assessment results
- Analyzing and reporting results for the purposes grades and institutional research
- Delivering content and managing online communication with students
- Grading (via online assessments) and reporting grades

Most of these functions save time and money. Teachers like the LMS because it alleviates the tedium of grading, students like the "anywhere, anytime" access, administrators like them because they provide data and visibility, and publishers like the LMS because it provides a method to distribute, control and monetize their digitized intellectual property. As a result, the educational technology ecosystem found in higher education today is highly LMS-centric [14]. In recent years, many K-12 schools and jurisdictions have also invested in LMS technology. Other common educational technologies, including authoring tools, learning content management systems, assessment engines, and repositories, have been heavily influenced by the need to produce content that can be delivered via an LMS. In other words, the LMS is the dominant channel for formal learning, much as television once was for video [15].

This state of affairs has been changing for several years now as newer types of learning content have become more prevalent, including mobile apps, video lectures, online meetings, social learning, eBooks, games, and simulations. The typical LMS course contains didactic content and quizzes with pre-determined answers (e.g. multiple choice, matching and fill-in-the-blank questions), whereas these newer types of content tend to be more interactive and open ended in their assessment if student outcomes. User management and tracking results are still important in formal educational settings and for publishers' business models, but app stores and sites like YouTube are more natural delivery platforms for mobile and video content. "Learning content" is being replaced by "learning applications" that are hosted as mobile apps or as web applications in the cloud. Moreover, many of the most widely used and freely available courses (MOOCS) generate their own certificates of completion and are by their nature not tied to any one institution and therefore not to any institution's LMS.

3 Emerging Standards

As a consequence of these changes, the technical standards used by eLearning systems are being updated and revised to enable distributed systems to securely exchange data across the web [16]. This trend includes the IMS Global Learning Consortium's *Learning Tools Interoperability* (LTI) and *Learning Information Services* (LIS) specifications [17, 18] and the *Experience API* (also known as "Tin Can") produced by the U.S. Advanced Distributed Learning (ADL) initiative [19]. These standards enable applications to communicate without a central broker such as an LMS. They support interoperable reporting of assessment outcomes, course completions, and additional data relevant to learning experiences.

The capabilities offered by these emerging standards are critical for the adoption of the next generation of learning applications. For example, products such as ALEKS [20, 21], Autotutor [22, 23], Brainrush [24], Carnegie Learning [21, 25], Knewton [26], Wyang Outpost [16], and many others [27] are using embedded AI and, in some cases, game dynamics to create more effective and more engaging learning experiences. Students are now using these resources (and others such as the Kahn Academy and MOOCS) because they are either more effective or more available than traditional educational offerings. However, for these products to gain market acceptance they must be able to integrate with the ambient eLearning infrastructure. At some point schools, parents, and employers will want to see *evidence* of achievement. These systems will need to communicate results to institutional LMSs, online data repositories, and a variety of personal management apps running on the mobile devices of students, teachers, and parents.

4 New Product Categories

In addition to intelligent learning applications, many other new product categories are likely to emerge. Some will be engendered by societal requirements and others by advances in educational technology.

For example, students and teachers are increasingly associated with multiple institutions at the same time [28], and many of the more innovative learning technologies (including MOOCS and most of the systems listed earlier) are typically used outside standard classroom practice. This leads to requirements to track rosters, assignments, progress, and grades across multiple institutions and multiple online learning systems and to maintain a student's preferences in a "learner model" [29-31] that can be updated and exchanged by multiple adaptive learning systems. The natural evolution of the e-portfolio will be a personal learning record store that:

- Is securely controlled by the learner;
- Is portable as the learner works with multiple schools, teachers, tutors, and publishers over the years; and

- Contains the learner's preferences and his validated and certified formal and informal learning history.

This evolution would parallel the recent evolution of Electronic Health Records and, if implemented on a global scale, would spawn a plethora of products, ranging from tools to manage learning records to learning applications that take advantage of them to deliver more personalize, culturally relevant, and educational effective learning experiences.

Similarly, advances in cognitive science, computer science and information technology are also creating both requirements and affordances for new product categories. Just as the underlying technological components of expert systems have now found their way into hundreds of products from rice cookers to mobile phones, we anticipate that the AI components of today's intelligent tutoring systems will work their way into a wide range of learning products. The same is true for automated language understanding [32], automated grading [33], affect detection [34, 35], gesture and sketch recognition [36-38], and forms of social media that enable students to collaborate with each other and with adults (e.g. "granny tutors") [39].

Returning to the theme of standards, we observe that as learning products incorporate more intelligent features, they will generate and require significantly more data about learners, learning activities, and outcomes. Their commercial success will depend in part on their ability to create value by leveraging these data across multiple systems, jurisdictions, and stages of a life. Economically, it makes sense for learning systems to share their data rather than to hoard it, which is why standardized formats for data exchange are so important.

5 The IEEE Actionable Data Book Project

As pointed out above, standards help learning technologies integrate with existing infrastructure and processes. This means that innovations developed to meet the needs of a niche market – say one dominated by relatively low bandwidth cellular access, or one in which a culture demands different levels and types of privacy – can be used in other markets as well. Tools originally created for broader (or wealthier) markets would be more easily tailored for use elsewhere. As real-world example of a project where standards, new technologies, and unique requirements from a developing country have converged, we examine the IEEE *Actionable Data Book Project for STEM Education*, or more simply the IEEE ADB project [40].

The IEEE ADB project grew out of paper presented at the IEEE Global Humanitarian Technology Conference in 2011 that discussed a broadly applicable framework for building educational applications that combined field data collection and data visualization [41]. The requirements for the system presented in that paper came from the rice ecosystem management on the Indonesian island of Bali. In 2013, the suggestions in the paper were actualized in the IEEE ADB project. The goal of this one-year

R&D collaboration is to define and demonstrate an "actionable data book" consisting of a specialized eBook based on open standards that is tailored to support STEM education and supports learner accessibility and usage preferences. The requirements for the actionable data book are that it must be able to

- Use camera and GPS data from a learner's mobile platform
- Use measurements from local lab equipment
- Exchange results of learning interactions with cloud-based LMSs, analytics engines, and other applications
- Retrieve content from cloud-based sources (e.g. content repositories)
- Store and retrieve student history and preferences in the cloud

Operationally, the project is hosted by Industry Connections, an IEEE Standards Association program that facilitates the early exploration of potential interoperability solutions [42]. Participation is free and open to interested parties. The ADB project may continue past the initial year's charter, depending upon success.

Technologically, the project anticipates the global availability of a class of mobile devices comprising smart phones and connected tablets and explores the premise that those devices, in conjunction with a new content format, may provide the first truly global platform for connected learning. The format in question is EPUB 3 [43, 44], a new eBook format defined by the International Digital Publishing Forum [45].

EBooks have emerged as a mass-market commercial success within the past few years. To date, eBooks have only replicated the static content of printed books in a digital medium, but EPUB 3 introduces interactivity to eBooks by embracing JavaScript and the HTML5 standard for web page content. These characteristics make EPUB 3 an attractive foundation for a learning delivery platform. EPUB 3 offers a complete solution for portable, interactive, connected content, and it is relatively simple to map the requirements for an interactive learning activity onto baseline EPUB 3 capabilities. Since EPUB 3 is a general-purpose technology with broad appeal outside of the education industry, it is more likely than education-specific standards to be widely adopted, supported, and have a multi-decade life span.

Although most of the technology used by the IEEE ADB project was developed for commercial purposes in the developed world, its application to learning was originally inspired by the desire to enable students in remote locations to collect field data and share their data and culture with students in the United States. The first use case to which it will be applied is the construction of an enhanced, interactive guidebook for the new UNESCO World Heritage site on Bali [46-48].

The UNESCO site covers a significant geographical area encompassing 21 communities engaged in rice production and following traditional spiritual practices. This has resulted in an enormous challenge: How does one design an interactive guidebook that promotes the conservation and preservation of the site while meeting the needs of the people who live there, the international team developing and maintaining the site, and tourists from all over the world with varying degrees of cultural sensitivity? The IEEE ADB project aims to help meet these requirements by developing onsite learning activities and guides that adapt to the local geography and culture as well as to those of the user's culture, while also providing remote connectivity to that allows students to vicariously experience the site from anywhere on the planet.

6 Conclusions

In developing economies new policies, institutions, and business models will transform the way education is delivered and managed. These efforts will take advantage of a wide range of innovative educational technologies and products to create local solutions that overcome geographical, social, and economic barriers using global infrastructure. It is easy to envision detailed student background information being securely available via the Internet and learning systems that compete with each other on the basis of how effectively they use this information.

Similarly, as more opportunities become available for students to access online video, daily lectures may become a thing of the past and expensive, classroom-based instruction may be needed less frequently or used differently, e.g. only for activities that require in-person group interactions or that use equipment not available in homes. Independent, trusted assessment services [49, 50] may allow students to progress in school based on their acquired competence, displacing today's cohort-based advancement schemes that measure progress by seat-time. The possibilities are unlimited and each educational jurisdiction will shape their solution by their specific needs and resources.

Data exchange standards and software interoperability standards are key to the flexible configuration of future systems, online services, and mobile applications. Standards-based products allow a school or a national or regional education agency to configure multiple products, including their current systems, into a stable working solution that fits local requirements and that allows new capabilities to be incorporated over time with minimal effort. The IEEE Actionable Data Book project is an example of a new model for learning delivery based on globally available, open standards that focuses on the realities of teaching and learning in the developing world.

7 REFERENCES

1. Khan, S., *The One World Schoolhouse: Education Reimagined*. 2012: Twelve.
2. Hylén, J., D.V. Damme, and F. Mulder, *Open Educational Resources*. 2012.
3. Wiley, D., et al., *A preliminary examination of the cost savings and learning impacts of using open textbooks in middle and high school science classes*. The International Review of Research in Open and Distance Learning, 2012. **13**(3): p. 262-276.

4. Anderson, T., *Promise and/or Peril: MOOCs and Open and Distance Education*. 2013.
5. Belanger, Y. and J. Thornton, *Bioelectricity: A Quantitative Approach Duke University's First MOOC*. 2013.
6. Ruth, S., *Can MOOC's and Existing E-Learning Efficiency Paradigms Help Reduce College Costs?* Available at SSRN 2086689, 2012.
7. Murray, K.W. and N. Orii, *Automatic Essay Scoring*.
8. Tsai, M.-h., *The Consistency Between Human Raters and an Automated Essay Scoring System in Grading High School Students' English Writing*. Action in Teacher Education, 2012. **34**(4): p. 328-335.
9. Vujosevic-Janicic, M., et al., *Software Verification and Graph Similarity for Automated Evaluation of Students' Assignments*. arXiv preprint arXiv:1206.7064, 2012.
10. Wikipedia. *History of Radio*. 2013 [cited 2013 May 17]; Available from: https://en.wikipedia.org/wiki/History_of_radio.
11. Green, C., *CAMPUS COMPUTING, 2007*, 2007, Claremont, CA: Claremont Graduate University.
12. Dahlstrom, E., et al., *with a foreword by Diana Oblinger*. The ECAR National Study of Undergraduate Students and Information Technology, 2011 (Research Report), 2011.
13. Green, C., *CAMPUS COMPUTING, 2011*, 2011, campuscomput.net.
14. Mott, J., *Envisioning the post-LMS era: the Open Learning Network*. Educause Quarterly, 2010. **33**(1): p. 1-9.
15. Frieden, R., *Next Generation Television and the Migration from Channels to Platforms*. Available at SSRN 2117960, 2012.
16. Alario-Hoyos, C. and S. Wilson. *Comparison of the main alternatives to the integration of external tools in different platforms*. in *Proceedings of the International Conference of Education, Research and Innovation, ICERI*. 2010. Citeseer.
17. GLC, I. *Learning Information Services*. 2012 [cited 2013 May 17]; Available from: http://www.imsglobal.org/lis/.
18. GLC, I. *Learning Tools Interoperability*. 2012 [cited 2013 May 17]; Available from: http://www.imsglobal.org/toolsinteroperability2.cfm.
19. ADL. *Training & Learning Architecture (TLA): Experience API (xAPI)*. 2013 [cited 2013 May 17]; Available from: http://www.adlnet.gov/tla/experience-api.
20. ALEKS. *ALEKS Home Page*. 2011 [cited 2011 October 20]; Available from: http://www.aleks.com.
21. Sabo, K.E., et al., *Searching for the two sigma advantage: Evaluating algebra intelligent tutors*. Computers in Human Behavior, 2013. **29**(4): p. 1833-1840.
22. Graesser, A., et al., *AutoTutor: An Intelligent Tutoring System with Mixed-Initiative Dialogue*, in *IEEE Transactions on Education*2005, IEEE. p. 612-618.
23. Hu, X., et al. *AutoTutor Lite*. 2009. IOS Press.
24. Brainrush. *Brainrush Home Page*. 2013 [cited 2013 May 17]; Available from: http://www.brainrush.com/.
25. Carnegie Learning. *Carnegie Learning Home Page*. 2013 [cited 2013 May 17]; Available from: http://www.carnegielearning.com/.
26. Knewton. *Knewton improves learning outcomes (Knewton Home Page)*. 2013 [cited 2013 May 17]; Available from: http://www.knewton.com/.

27. Institute of Education Sciences. *What Works Clearinghouse*. 2013 [cited 2013 May 17]; Available from: http://ies.ed.gov/ncee/wwc/.
28. Newbaker, P., *No Straight Path to College Graduation*, 2012.
29. Sottilare, R., et al., *Design Recommendations for Adaptive Intelligent Tutoring Systems: Learner Modeling (Vol. 1)*. 2013: Army Research Lab.
30. Woolf, B.P., *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. 2009, Burlington, MA: Morgan Kaufmann.
31. Durlach, P.J. and J.M. Ray, *Designing adaptive instructional environments: Insights from empirical evidence*, in *Army Research Institute Report*2011: Arlington, VA.
32. Robson, R. and F. Ray. *Applying Semantic Analysis to Training, Education, and Immersive Learning*. in *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)*. 2012. NTSA.
33. Valenti, S., F. Neri, and A. Cucchiarelli, *An overview of current research on automated essay grading*. Journal of Information Technology Education, 2003. **2**: p. 319-330.
34. Hussain, M., et al. *Affect detection from multichannel physiology during learning sessions with AutoTutor*. in *Artificial Intelligence in Education*. 2011. Springer.
35. Calvo, R.A. and S. D'Mello, *Affect detection: An interdisciplinary review of models, methods, and their applications*. Affective Computing, IEEE Transactions on, 2010. **1**(1): p. 18-37.
36. Weinland, D., R. Ronfard, and E. Boyer, *A survey of vision-based methods for action representation, segmentation and recognition*. Computer Vision and Image Understanding, 2011. **115**(2): p. 224-241.
37. Yin, P., et al. *Sketch worksheets: A sketch-based educational software system*. in *Proceedings of the 22nd Innovative Applications of Artificial Intelligence Conference, Atlanta*. 2010.
38. Valentine, S., et al., *Mechanix: A Sketch-Based Tutoring and Grading System for Free-Body Diagrams*. AI Magazine, 2012. **34**(1): p. 55.
39. Doctorow, C. *Deploying the British Granny Cloud to tutor poor Indian classrooms over Skype*. 2011 [cited 2013 May 17]; Available from: http://boingboing.net/2011/01/20/deploying-the-britis.html.
40. IEEE ADB Project. *Public Home Page: The IEEE Actionable Data Book for S.T.E.M. Education*. 2013 [cited 2013 May 17]; Available from: http://www.ieee-adb.org/.
41. Richards, T. and A. Barr. *Catalyzing Connected Learning through Standards*. in *Global Humanitarian Technology Conference (GHTC), 2011 IEEE*. 2011. IEEE.
42. Richards, T. *IEEE Actionable Data Book for STEM Education: Industry Connections Activity Initiation Document (ICAID)*. 2012 [cited 2013 May 17]; Available from: http://standards.ieee.org/about/sasb/iccom/IC12-006-02_IEEE_Actionable_Data_Book_for_STEM_Education.pdf.
43. Garrish, M., *What is EPUB 3?* 2011: O'Reilly Media.
44. IDPF. *EPUB 3*. 2013 [cited 2013 May 17]; Available from: http://idpf.org/epub/30.
45. IDPF. *Home Page*. 2013 [cited 2013 May 17]; Available from: http://idpf.org/.
46. Lansing, J.S. and J.N. Watson, *Guide to Bali's UNESCO World Heritage, Tri Hita Karana: Cultural Landscape of Subaks and Water Temples*, 2012.

47. UNESCO. *Cultural Landscape of Bali Province: the Subak System as a Manifestation of the Tri Hita Karana Philosophy*. 2012 [cited 2013 May 17]; Available from: http://whc.unesco.org/en/list/1194/.
48. Lansing, J.S. and J.N. Watson. *Water temples forever (Video)*. 2012 [cited 2013 May 17]; Available from: http://www.youtube.com/watch?v=jPetj4MSDdY.
49. Morgan, J. and A. Millin, *ONLINE PROCTORING PROCESS FOR DISTANCE-BASED TESTING*, 2011, Google Patents.
50. Kitahara, R., F. Westfall, and J. Mankelwicz, *New, multi-faceted hybrid approaches to ensuring academic integrity*. Journal of Academic and Business Ethics, 2011. **3**: p. 1-12.