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.

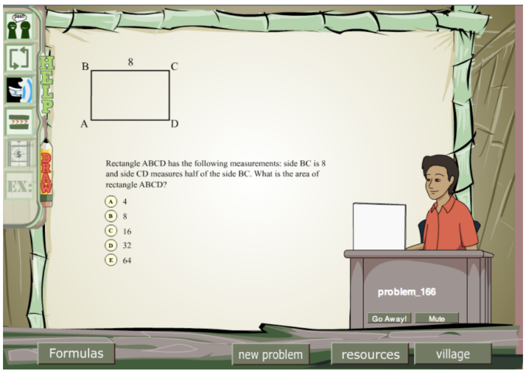

Fig. 1. The Wayang Outpost Math Tutor interface.

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.