How well does Sparse Blended Learning work? A Case Study from a Developing Country

Imran A. Zualkernan
American University of Sharjah,
Sharjah, UAE
izualkernan@aus.edu

Asad Karim
TeleTaleem
Islamabad, Pakistan
asadkarim@teletaleem.com

Abstract

Blended learning in a classroom setting serves multiple purposes of helping students and teachers learn from experts in a field. However, it is challenging to design sustainable implementations of such paradigms in developing countries. For example, most schools in developing countries do not have the technology infrastructure required to implement such approaches. This paper presents a unique method called 'sparse blended learning' that covers the complete curriculum, but in only a few blended learning experiences spread throughout the school year. In addition to blended learning, the approach also incorporates a live remote tutor, and online assessments. The technological infrastructure is brought to the students in the form of a self-contained 'learning van' that contains an electrical generator, servers, satellite connectivity to the Internet, a Wi-Fi network, and Android tablets. Content from Khan Academy was modified to use this approach for teaching Grade IV and V Mathematics in a remote semi-urban public mountain school in a developing country. The school caters mostly to the poorest children in the community. The results are that after over 6 months of intervention, there was a significant difference between control and treatment groups for both grades. Further, the effects are comparable to those expected from intelligent tutors.

1. Introduction

When it comes to deployment of learning technologies, developing countries offer unique challenges including lack of infrastructure and capital, trained teachers, educational governance, and high costs of Internet access [1]. For example, in 2011, there were 154,641 public, and 17,969 private primary schools in a developing country like Pakistan, with a total enrollment of 16,894,233 children and 436,928 teachers [2]. Many of these teachers are not trained to international standards. For example, 62.4% of public school primary teachers in Pakistan have a one year certificate of teaching (CT) and 49.6% only have a 12 year high school diploma [2]. Almost half the children of the 5-9 year age group are out of school in Pakistan [3]. While enrollment is a challenge, the quality of learning is also bleak. For example, only 27.9% of girls and 33.7% of boys of 5-9 year olds in rural areas can read. Similarly, only 25% of the girls and 32.5% of boys of this age can do simple subtraction [3]. Almost 56.6% of children in primary schools also end up repeating a grade [2]. From a governance perspective, the teacher attendance in government schools was only 83.1% which means that the teacher was 'officially' absent from the School 17% of the school year [3]. With a teacher to pupil ratio of 36.7, the class sizes are also relatively large in public primary schools [2]. Finally, in Pakistan only 38% of the primary schools have electricity [2].

One approach to problems of education in the developing world is to use a blended learning approach like the MIT Blossom's Project [4]. In this approach, videos of 'inspirational' teachers combined with associated teacher training materials are deployed in a classroom in a blended fashion. Videos are shown first, followed by problems to be posed and solved by students, and subsequent follow-up with additional videos is carried out. One important aspect of this approach is that it involves the teacher actively in addition to students. While this approach is promising in general, a number of issues arise when deploying such approaches in a developing country. First and foremost is the availability of infrastructure like a computer, and overhead projector and even electricity. While the numbers of Blossom's videos have grown over the years, another issue is the availability of videos to cover all the topics in a curriculum. The second issue can probably be resolved with time. However, the non-availability of technology infrastructure is a constraint that will remain for many decades to come.

This paper presents a learning approach called "sparse blended learning" for developing countries that actively incorporates the lack of infrastructure' constraint into account. In sparse blended learning, students are exposed to a blended learning scenario every few weeks rather than on a daily or a bi or tri-weekly basis as it would be in a developed country. In such a scenario, for example, material covered in the last two weeks of classes is covered on a single day using a blended learning approach similar to the MIT Blossoms program. This can be done by either taking the whole class of students to a computer laboratory, or by preferably bringing the technological infrastructure to them. In the research reported here, the later approach was followed where a 'learning van' was used to bring the technology infrastructure to a remote school every two weeks. The obvious question is how well such an approach works?

The rest of the paper is organized as follows. The chosen school is described next, followed by a detailed description of the sparse blended learning intervention. An evaluation of the effectiveness of such interventions is presented next. The paper ends with a conclusion.

2. The School

The school chosen for this study is situated near the town of Balakot in the Mansehra District in the Khyber-Pakhtunkhwa province of Pakistan. Balakot sits on the Balakot-Bagh Fault Line, and the school was completely demolished in the earth quake of October 2005, and was re-built using donor support. The school has a total of 11 teachers and a headmaster. Being a public school, the school caters mostly to the children of day laborers who represent the bottom economic tier of this semi-rural community. In most cases, the children are required to tend stock or engage in household work after school. Figure 1 shows a glimpse of the school environment and the children it serves.

Figure 1. The Public School chosen for this study

3. The Learning Intervention

The area targeted for the research reported here were Grade IV and Grade V Mathematics.

3.1 Content and Localization

Khan Academy (KA) [5] was chosen as the primary source of content for this intervention because it was thought that most Math content would be readily available. KA provides instructional videos tied to the *Mathematics Common Core Curriculum* of the United States [6]. KA uses traditional teach-and-test pedagogy where each topic is presented by a teacher in the form of a simple video using a blackboard. Topics are tied to multiple-choice online assessments organized according to topic maps. Students can take the assessment to test their mastery of the various topics. Tables I and Table II shows how the topics and Student Learning Objects (SLOs) from KA were mapped to the 2006 National Curriculum of Pakistan which is followed in the target school. Surprisingly, for various reasons like curriculum, cultural and pedagogical misalignment, only a small percentage of the KA videos (40% for Grade IV and 34.3% for Grade V) could be used, and 45 and 51 additional videos had to be created for Grade IV and Grade V Math respectively [7].

TABLE I. LOCALIZATION EFFORT FOR GRADE IV

Unit – 2006 National Curriculum	Sub-topics	SLO	Video/SLO	% KA Used
Numbers and Arithmetic operations	6	15	0.93	78.57
Factors and Multiples	6	13	0.77	70.00
Fractions	5	19	0.79	60.00
Decimals and Fractions	3	11	0.64	100.00
Measurements	4	17	1.12	0.00
Geometry	5	23	0.78	0.00
Information Handling	2	2	1.00	0.00
Total	31	100	0.85	40.00

TABLE II. LOCALIZATION EFFORT FOR GRADE V

Unit – 2006 National Curriculum	Sub-topics	SLOs	Video/SLO	% KA Used
Numbers and Arithmetic operations	4	11	1.00	54.55
HCF and LCM	2	5	1.20	16.67
Fractions	3	10	1.10	63.64
Decimals and Percentages	2	19	0.58	72.73
Distance, Time and Temperature	3	12	0.17	100.00
Unitary Method	2	6	0.67	0.00
Geometry	3	14	1.07	0.00
Perimeter and Area	1	6	0.67	0.00
Information Handling	2	7	0.71	0.00
Total	**22**	**90**	**0.77**	**34.78**

3.2 Technology Platform

Since a blended learning pedagogy was envisioned, a remote interactive classroom environment called *Teletaleem* (T2) coupled with the *Moodle* Learning Management System (LMS) to serve and archive the results of assessments was used. These assessments were delivered to the children using Android tablets via Wi-Fi. The original plan called for using the built-in adaptive formative assessment engine in KA. However, upon testing, the offline version of KA did not scale well to handle 30 Android tablets through the commonly available wireless routers. Consequently, Moodle was used as the assessment platform. Figure 2 shows a screenshot of the Teletaleem (T2) learning platform. The T2 platform allows a tutor to deliver remote presentations using a laptop and a standard camera and includes a whiteboard. The tutor is also able to view the video feeds of multiple classrooms in real time.

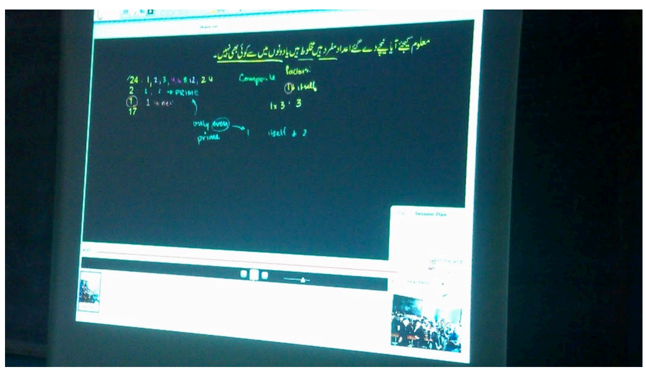

Figure 2. The Taletaleem Learning Platform

3.3 The Pedagogy

The learning intervention uses a blended learning pedagogy similar to the MIT Blossoms program [4]. However, unlike the Blossoms program, this pedagogical intervention also includes an experienced remote teacher who served as a role-model and a mentor for the in-class teacher. In addition, since the school targeted did not have computers or electricity, the learning videos were delivered through a "learning van" fashioned after the concept of a "traveling library." The learning van is a self-contained computer laboratory that contains a server, satellite access to the Internet, a power generator and UPS, wireless router, overhead projector, digital pad, teacher laptop, an external microphone and 30 Android based tablets that can connect to the server or the Internet wirelessly.

The learning van is designed to visit a school once every two weeks for a whole day and the pedagogy can be divided into three stages, pre-visit, during and post-visit. One week before the van arrives, based on the pace of each teacher, the local in-class teacher based in the school is asked to teach students a set of topics and SLOs before the van arrives. A customized online course including the SLOs and related assessments is created for each school for the day of the van visit.

Figure 3. . The learning van visit day activities

Figure 3 shows the sequence of events on the day of the visit. After the van arrives, children are shown a series of KA-type videos (Figure 3 (b)) on the topics discussed in the pre-visit phase with the in-class teacher. It is important to note that these sessions are conducted not by the in-class teacher, but by a remote tutor who is experienced and trained for Grade IV and Grade V Mathematics. This is partially done to provide mentorship for the local in-class teacher. After viewing a video on a specific SLO, the remote teacher uses the whiteboard to solve multiple sample problems to reinforce the understanding of students (Figure 3 (c)). In the next stage, the remote teacher invites students from the class to step up to the digital pad and solve additional problems. The tutor as well as the rest of the class can view how a child solves the problems (Figure 3(d)). After all the topics for day have been thus covered (about eight per day), the children are asked to check their understanding by attempting multiple-choice questions related to the topics they have just studied (Figure 3 (e)). Children use Android tablets to take formative and summative assessments. After taking assessments related to current topics, children are tested again on topics that were covered two weeks earlier to determine if any progress has been made. The results are recorded and made available through LMS reports. After attempting formative and summative assessments, the

In the post-visit phase, reports of children's assessments for the day is printed and discussed with the in-class teacher. In specific, the weaker students as well as SLOs where children did not perform well are identified. This report is then used to plan remedial session on topics for the next two weeks before the van arrives again. Finally, the results from last visit's assessments are compared with the current week to review any progress on remedial session for the last two weeks.

4. Methodology

This section describes the methodology used to evaluate the sparse blended learning intervention described in earlier sections.

4.1 Development of an Assessment instrument

Since there was no standardized instrument to measure the performance of children in Grade V and Grade IV Mathematics for the 2006 National Curriculum of Pakistan, a standardized test instrument based on Item Response Theory (IRT) [8] and classical analysis was developed. 400 items each for Grade IV and Grade V were constructed by experienced Pakistani item writers based on the 2006 National Curriculum of Pakistan. These 400 items were tested against the target population of students in Balakot and adjoining areas, and based on test results and subsequent statistical analyses, two standardized test instruments with 80 items each for Grade IV and Grade V Math were finalized.

4.2 Assignment of Control and Treatment Sections

The target school had two sections each of Grade IV and V. One section each of grade IV and grade V was randomly assigned to be either treatment or control. Table III shows the background of respective teachers for control and treatment sections. The control sections would continue with the normal teaching practices, while the treatment sections would be exposed to the learning intervention two terms. The intervention started on September 15, 2012 and was completed on March 8, 2013. At the end of the term, students in control and treatment group were tested based on the standardized test instrument developed earlier.

TABLE III. TEACHER QUALIFICATIONS IN CONTROL AND TREATMENT GROUPS

Class	Status	Teacher's Educational Qualifications	Teachers Years of Teaching Experience
IV	Treatment	12 year of schooling (HSSC), one year teaching diploma	19
	Control	12 year of schooling(HSSC), one year teaching diploma	22
V	Treatment	14 years of schooling (B.A.), one year of teaching diploma	20
	Control	16 years of schooling (M.Ed), one year of teaching diploma	15

5. Results

Results based on the final standardized test for both control and treatment groups are presented next. Since most grade distributions were found not be normally distributed (Anderson-Darlington test, $p<0.05$), non-parametric statistical analysis was used. Results of learning interventions are typically compared by reporting the effect generally given by equation (1) below by dividing the difference in estimated means by the standard deviation [9].

$$\text{Effect} = \frac{(\hat{\mu}_{treatment} - \hat{\mu}_{control})}{SD} \qquad (1)$$

However, it is difficult to compare directly the definition of 'effect' which is typically based on assumptions of normality when using non-parametric methods. Grissom and Kim [9] have proposed a method of calculating non-parametric effects by using the equation (2) given below.

$$\text{Effect} = \hat{p}_{treatment,control} = \frac{U}{n_{treatment} n_{control}} \qquad (2)$$

Where n represents the sample sizes for treatment and control groups, and *U* is the Mann-Whitney *U* statistic. The effect described in equation (2) estimates the probability that a score randomly drawn from population of treatment groups will be greater than a score randomly drawn from population of control groups.

In order to compare the non-parametric results in an informal fashion, another definition of effect shown in Equation (3) will be used in this paper as well. This notion of effect is based on medians (as opposed to means) and provides an informal comparison to effects as defined using Equation 1.

$$\text{Effect} = \frac{(median_{treatment} - median_{control})}{SD_{pooled}} \qquad (3)$$

5.1 Grade IV Results

Treatment group ($n_{treatment}$ = 35; mean = 43.73; SD=13.06; median = 43.48) for grade IV performed significantly better than control group ($n_{control}$ = 37; mean = 23.15; SD=11.09; median =23.91) (Levene's test statistic = 0.22, P = 0.641; Kruskal-Wallis H = 32.01 DF = 1 P = 0.000; Mood Median Chi-Square = 29.49 DF = 1 P = 0.000; Mann-Whitney W = 1779.0 $P<0.05$). The overall effect was 0.8872 according to equation (2). This means that a randomly chosen student from the treatment group is likely to be better than a typical student from the control group 88.72% of the time. Based on 1-Sample Wilcoxon test, control group had an estimated median of 22.83 with a 95% confidence interval of [19.57, 27.16]. Treatment group, on the other hand, had a much higher estimated median of 43.48 with a 95% confidence interval of [39.13, 47.83]. The effect based on equation (3) was 1.61.

A breakdown of the results comparing results on the various learning outcomes covered in grade IV (only first 5 units were covered in the school year) are given in Table IV below.

TABLE IV. COMPARISON OF LEARNING OUTCOMES GRADE IV

Learning Unit 2006 National Curriculum	**% KA Used**	**Comparison with Control Group** * significant	**Estimated Median [95% Confidence Interval] Wilcoxon-Sign Test**	**Effect as per Eq. (2)**	**Approx. Effect as per Eq. (3)**
Numbers and Arithmetic operations	78.57	Levene's Test statistic = 1.30, P= 0.258; Kruskal-Wallis; H = 29.18 DF = 1 P = 0.000*	Treatment: 50.0 [40.9, 54.5]	0.86641	1.21
			Control: 22.73 [18.18 27.27]		
Factors and Multiples	70.00	Levene's Test statistic = 0.00, P= 0.994; Kruskal-Wallis, H = 15.59 DF = 1 P = 0.000*	Treatment: 37.5 [31.3, 43.8]	0.7861	0.7544
			Control: 18.8 [12.5,25.0]		
Fractions	60.00	Levene's Test statistic = 1.33, P = 0.253; Kruskal-Wallis, H = 20.44 DF = 1 P = 0.000*	Treatment: 40.0 [35.0, 45.0]	0.8265	1.354
			Control: 20.00 [20.00, 25.00]		
Decimals and Fractions	100.00	Levene's Test statistic = 0.73, P= 0.395; Kruskal-Wallis, H = 22.38 DF = 1 P = 0.000*	Treatment: 43.8 [37.5, 56.3]	0.8428	1.272
			Control: 18.8 [12.5,25.0]		
Measurements	0.00	Levene's Test statistic = 1.83, P= 0.181; Kruskal-Wallis, H = 13.00 DF = 1 P = 0.000*	Treatment: 44.4 [38.9, 50.0]	0.7642	1.134
			Control: 27.8 [16.7, 33.3]		
Geometry	0.00	Not Covered	N/A	N/A	N/A
Information Handling	0.00	Not Covered	N/A	N/A	N/A

A multivariate correlation analysis based on Spearmen's ρ shows that the learning across units of learning was independent ($p < 0.05$). This means, for example, that children who did well in *Numbers and Arithmetic operations* did not necessarily perform similarly in other learning units and vice-versa.

As Table IV shows, even though there were differences, good effect sizes from 76.42% to 86.6% were achieved for the various learning outcomes showing a significant improvement over the control group.

5.2 Grade V Results

The treatment group ($n_{treatment}$=28; mean=55.24; SD=18.30; Median=57.45) for grade V outperformed the control group ($n_{control}$ = 28; mean = 43.84; SD=16.17; Median=42.55) (Levene's Test statistic = 0.28, P= 0.602; Kruskal-Wallis H = 7.69 DF = 1 P = 0.006; Mood Median Chi-Square = 4.57 DF = 1 P = 0.033; Man-Whitney W = 629.0 P=0.0058). The overall effect according to equation (2) was 0.2844. This means that the probability that a student in the treatment group would perform better than a control group student was 28.44%. The 1-Sample Wilcoxon method estimated the median of control group to be 43.6 and a 95% confidence interval of [37.2, 51.1]. Similarly, treatment group had a higher median of 57.4 and a 95% confidence interval of [50.0, 63.8]. There was an overall effect of 0.6162 based on equation (3).

TABLE V. COMPARISON OF LEARNING OUTCOMES GRADE V

Learning Unit 2006 National Curriculum	% KA Used	Comparison with Control Group * significant	Estimated Median [95% Confidence Interval] 1-Sample Wilcoxon-Sign Test	Effect as per Eq. (2)	Approx. Effect as per Eq. (3)
Numbers and Arithmetic operations	54.55	Test statistic = 0.26, p-value = 0.610; Kruskal-Wallis, H = 1.23 DF = 1 P = 0.268	Treatment: 57.1 [42.9, 64.3] Control: 50.0 [42.9, 57.1]	N/A	N/A
HCF and LCM	16.67	Levene's Test statistic = 4.58, P= 0.037; AD P<0.05. Cannot compare.	Treatment: 0.0 [0.0,50.0] Control: 0.0 [0.0, 0.0]	N/A	N/A
Fractions	63.64	Levene's Test statistic = 3.22, P = 0.078; Kruskal-Wallis, H = 4.13 DF = 1 P = 0.042*	Treatment: 28.6 [21.4,35.7] Control: 35.7 [28.6, 42.9]	0.6536	0.8467
Decimals and Percentages	72.73	Test statistic = 1.52, P = 0.223;Kruskal-Wallis, H = 1.86 DF = 1 P = 0.172	Treatment: 61.1 [50.0, 72.2] Control: 50.0 [44.4,61.1]	N/A	N/A
Distance, Time and Temperature	100.00	Levene's Test statistic = 0.19, P = 0.669;Kruskal-Wallis, H = 13.91 DF = 1 P = 0.000*	Treatment: 85.0 [60.0,90.0] Control: 45.0 [35.0, 55.0]	0.2142	0.63
Unitary Method	0.00	Levene's Test statistic = 1.20, P = 0.278; Kruskal-Wallis, H = 7.55 DF = 1 P = 0.006*	Treatment: 66.7 [58.3,75.0] Control: 41.7 [33.3, 58.3]	0.2920	0.442
Geometry	0.00	Levene's Test statistic = 0.00,P= 1.000;Kruskal-Wallis, H = 2.86 DF = 1 P = 0.091	Treatment: 42.9 [35.7,50.0] Control: 35.7 [28.6,42.9]	N/A	N/A
Perimeter and Area	0.00	Not covered	N/A	N/A	N/A
Information Handling	0.00	Not covered	N/A	N/A	N/A

Finally, a multivariate correlation analysis using Spearman's ρ shows that *Numbers and Arithmetic operations* was correlated with *Decimals and Percentages* (ρ= 0.817; Fisher's Z 95% confidence [0.410,0.953]) and *Unitary Method* (ρ= 0.804, Fisher's Z 95% confidence interval [0.378, 0.949]). Finally, *Decimal and Percentages* are correlated with *Distance, Time and Temperature* unit (ρ= 0.833, Fisher's Z 95% confidence interval [0.449, 0.957]).

So although there was an overall statistical difference between control and treatment groups in Grade V, the effects were much lower than for Grade IV. Considering that the study was conducted in the same school, it is difficult to ascertain why the effect was lower for Grade V. However, one hypothesis is that teacher qualification may have had an impact. As Table III shows, both teachers for grade IV had similar experience and training, while the teacher for the control group for grade V had an M.Ed as opposed to a simple B.A. for the treatment group teacher. This hypothesis, however, needs to be investigated further.

6. Conclusion

In are recent survey of tutoring systems, [10] argues that while it believed that answer-based, intelligent and human tutors have typical effects (given by equation (1)) of 0.3, 1.0 and 2.0 respectively, in reality, the effect is only 0.76 for intelligent tutors which is similar to human tutors where the effect is actually 0.79. If this is the case, then the sparse blended learning approach presented in this paper has performed quite well with effects (although based on medians and under assumption of normality) are overall a 1.61 for Grade IV and 0.6162 for Grade V Mathematics.

This paper explored a technology-enabled learning intervention using a 'learning van' to implement a sparse blended learning scenario in a remote mountain school of a developing country. The approach clearly worked in showing statistical gains in learning outcomes of the students. The next challenge is to scale this intervention to serve a larger number of students and teachers.

References

[1] B. Woolf , I. Arroyo, I. and I. A. Zualkernan, "Education Technology for the Developing World," *The IEEE Global Humanitarian Technology Conference*, October 30 – November 1, 2011, Seattle, Washington, USA.

[2] *Pakistan Education Statistics 2010-2011*, National Management Information System, Academy of Educational Planning and Management, Ministry of Professonal and Technical Training, Government of Pakistan, 2011.

[3] *Annual Status of Education Report (ASER) 2011*, South Asian Forum for Educational Development, January 12, 2012.

[4] R. C. Larson, and M. E. Murray, "Open educational resources for blended learning in high schools: Overcoming impediments in developing countries," *Journal of Asynchronous Learning Networks*, 12(1), 2008, 85-103.

[5] *Khan Academy*, available at https://www.khanacademy.org/ accessed January 27. 2013.

[6] *Common Core Curriculum*, available at http://www.corestandards.org/ Accessed: November 29, 2012.

[7] I. A. Zualkernan, and A. Karim "Online Content Localization for Blended Learning in Developng Countries: A Case Study using Khan;s Academy," *7th International Technology, Development and Education Conference, Valencia*, Spain, March 4-6, 2013.

[8] H. Wainer, *Computerized Adaptive Testing: A Primer*, Lawrence-Erlbaum and Associates, New Jersey, 2000.

[9] R. J. Grissom, & J. J. Kim, J. J. *Effect sizes for research: Univariate and multivariate applications*. (2nd ed.). New York, NY: Taylor & Francis, 2012.

[10] KURT VanLEHN (2011): "The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems," *Educational Psychologist*, 46(4), pp.197-221.