Cost Effectiveness of a Complementary Reading Program

Misty Sailors[1] & Miguel Flores[2]

Abstract

We used a quasi-experimental, pre- post-test design to measure the cost effectiveness of a complementary reading program. Our findings indicated that the educational innovation under discussion was particularly effective in changing teacher practices and attitudes and that had the intervention been in place longer than four months, we would have seen greater reading achievement of learners in these classrooms. Our cost effective ratios and relative cost effective ratio indicated that the intervention was much more cost effective than business as usual in Malawi. Our study has implications for policy makers around the world.

Keywords: cost effectiveness, Malawi, early intervention, reading, reading materials, teacher education

Introduction

Many agree as to the importance of learning materials (Lee & Zuze, 2011) as an effective means to raising achievement (Davidson & Hobbs, 2012). So much, in fact, that textbooks and supplementary learning materials are the norm in countries like the USA. In some states in the USA, state money is provided to local school districts to purchase approved textbooks for core curriculum areas (Sailors & Hoffman, 2011). For many schools, the decision around which textbooks to purchase is made at the local level with teachers sitting on local adoption committees. For teachers, this means that textbooks align with state and local curriculum; for students, they each have their own textbooks in the various curricula areas. Many publishing companies now provide complementary reading materials (in addition to core reading texts) as part of their reading programs.

However, access to learning materials (especially reading materials) in many developing countriesis limited (Alidou et al., 2006; Hoffman, Sailors, Makalela, & Matthee, 2009; Makalela, 2005). Subsequently, book floods offset the lack of materials. Research has demonstrated the importance of book floods in improving children's writing, listening comprehension, and related language skills, which are slow to develop under traditional styles of teaching in developing countries (Elley, 2000, p. 250). Unfortunately, not only is there a lack of reading materials in developing countries, what is there is often not ofthe highest quality. Theyare often poorly designed, contain factual inaccuracies, do not support higher-level thinking or the lived experiences of the children using them, and reinforce gender stereotypes (Craig, Kraft and du Plessis, 1998; Montagnes, 2000).

[1]The University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249, Email: Misty.sailors@utsa.edu, Phone: +1 210 458 7288
[2]Tecnologico de Monterrey

Complementary reading materials alone cannot teach children to read; effective teachers are also essential. In addition to teaching and learning materials, teachers must have proper training (McGill-Franzen, Allington, Yokoi & Brooks, 1999). For example, combining materials and training in his seminal Book Flood study, Elley (2000) demonstrated that it is possible to double the rate of reading achievement of primary age students in developing countries by flooding their classrooms with approximately 100 high-interest books per class *and* providing teachers with instruction on how to use the books.

For teachers to be able to fully implement a reading program, they must have support for that implementation, including teachers' guides (Buckler, 2011). The prevalence of teachers' guides appears to assist teachers in knowing what to do with learning materials (Craig et al., 1998). For example, McGinn and his colleagues (McGinn, Warwick, & Reimers, 1989, as cited in Nannyonjo, 2007) suggested that guides provide teachers with pedagogical suggestions, recommend activities for classroom use and offer diagnostic tests to inform instruction. Farrell and Heyneman (1989) also noted that a well-designed teacher's guide could be a very effective form of in-service teacher training.

Even though research has demonstrated that access to reading materials and properly trained teachers contribute to the effectiveness of early grade reading programs (Sailors et al., 2014), lingering questions exist around the cost effectiveness of programs as an efficient way of using educational resources (Levin & McEwan, 2001).

The study under discussion in this paper was guided by the following research questions: "What are the outcomes of a complementary reading program on the reading achievement of young readers and the instructional practices of their teachers? What are the costs of that complementary reading program? What are the relative costs and outcomes of the program?" Moreover, "To what extent can the program be regarded as providing value for money?"

The Context of Malawi

Although it is one of the most peaceful countries on the Sub-Sahara African continent, Malawi faces many challenges. Narrow and land-locked, Malawi is home to more than 14 million people. The vast majority of people are subsistence farmers who live on less than two dollars a day. In a system where schooling became free under the Free Primary Education Policy, the Government supports almost three million students in primary schools, made up of Grades 1-8 (DoE, 2011). Since education in Malawi is not compulsory (at the time of this study), 12% of primary age children do not attend school. In urban areas, 91% of children attend school, while only 80% of rural area children attend school. Large class sizes exacerbate the situation. With a described student to teacher ratio of 80:1, it is common to see 125 gradeone children in a room no larger than 400 square feet (Sailors et al., 2014). Furthermore, there is a scarcity of reading materials in classrooms(Macro International, 2008).

Recognizing the need for greater change, the Malawi Ministry of Education, Science and Technology (MOEST), with the Malawi Institute of Education (MIE) and support of donor partners designed and instituted the National Primary Curriculum (NPC). The MIE designed, printed and distributed textbooks in each of the core subject areas for grades one to eight, including Chichewa (national) and English (official) language textbooks. At the time of this study, the language of instruction for the first four years of primary education was Chichewa with English as a subject (Chilora, 2007). Although the government has a one textbook per student policy, the reality is that there are approximately 2.25 students per Chichewa textbook in the primary schools (DoE, 2011). Teachers do not appear to use their textbooks consistently or adequately and are reluctant to let children use the books for fear of destroying the books. Head teachers (principals) keep the books in theiroffice to prevent those books from deterioration and to offset the possibility of a future shortage (World Bank, 2010).

That said, the Malawi government recognized need for additional materials if young and beginning readers were to not only learn to read, but also grow to love reading.

Thus was the beginning of the Textbooks and Learning Materials Program, known locally as Read Malawi. The United States Agency for International Development (USAID) funded this holistic reading program; it was implemented through a partnership by the MOEST, the University of Texas at San Antonio and key implementing collaborators, namely the Malawi Institute of Education (MIE).

The Read Malawi Program

The Read Malawi programwas a holistic intervention aimed at resourcing schools and classrooms with complementary reading materials and training for the teachers and school leadership on the proper implementation of the program. We describe each in the sections that follow.

Learning materials. Because complementary reading materials are important to young readers, the Read Malawi program designed and developed 180 titlesas a way of supporting reading skill development, love for reading, and, ultimately, a culture of readers (Sailors et al., 2012). The 180 titles(books)developed under the programwere authored at the MIE and at variousteacher-training colleges in Malawi. Authors of these local books were pre-service and in-service teachers, lecturers, and government officials. Authors wrote, revised, and field-tested their books with students. Curriculum specialists revised and edited the books for clarity and to make them of the highest quality as to provide for quality instructional materials. The books were representative of the distinctiveness of Malawi and were based on experiences of young children. The books covered concepts in life skills, mathematics, and sciences. The books supported personal, moral, and social development; demonstrated the importance of cultural practices and traditional folklore; and supportedgirl empowerment.

The books drew from several teaching formats that would support teachers in pedagogical approaches. Included in these materials were Guided Reading books (written at students' instructional levels for students to read in a small group setting with a more skilled reader, namely their teacher) and Read Aloud books (books that teachers read orally to students in a whole group setting). Additionally, the curriculum development team designed two alphabet books and alphabet posters. All materials were written in Chichewa and English.

Teaching materials. Because teaching materials can support teachers in improving their instruction, the Read Malawi program created several supporting materials for teachers, including 90 teachers'guides; each of the guides connected one guided reading and one read aloud book. The guides were linked directly to the national curriculum and provided teachers with innovative ways of developing literacy and subject area skills. The teachers' guides were developed in both English and Chichewa; the English teachers' guides coincided with the portions of classroom time spent on English, while the Chichewa guides coincided with the portion of classroom time spent on Chichewa. The guides for English complementary books were written with close attention to support second language practices. They were also written to support teachers in the teaching of reading strategies, including word identification/knowledge strategies, fluency strategies, and comprehension strategies. The guides used a patterned and systematic approach; once teachers grew accustomed to the patterns in the guides, then they could implement them in ways that were helpful to their students. The overview guidewas designed as an overview of the program and spelled out the key components of the program. In all, each participating school was resourced with over 4,100 teaching and learning materials.

Support for implementation. In order for an innovative program to be fully implemented, teachers need support; the team created and implemented a model of school-based support for teachers. Specifically, the team employed directive coaching (Deussen, Coskie, Robinson & Autio, 2007) to support the implementation of the program in classrooms.

Often referred to as an "outside in" approach (Sheridan, Edwards, Marvin & Knoche, 2009), directive coaching has been shown to improve teachers' practices (Pianta, Mashburn, Downer, Hamre & Justice, 2008). The support offered through directive coaching also made it possible for us to monitor, support and document the fidelity of the program. Our model of directive coaching drew upon local structures within the Malawi educational system: the staff at MIE served as coaches as did the head teachers (principals) and Primary Education Advisors (PEAs), although coaching was a new construct for many of our colleagues.

While the program met or exceeded all goals associated with our deliverables, we were interested in the cost effectiveness associated with the program. Namely, we were interested in the cost effectiveness related to the replication of the program at other schools in the country.

To that end, we employed a cost analysis, comparing the cost effectiveness of this innovativecomplementary reading program to that of the business-as-usual program in Malawi to determine a cost effective ratio.

Methods

In today's economic climate, given the limitation of resources for education, research methods that allow for informed decisions around how best to spend education monies are crucial. While traditional effectiveness studies are important for demonstrating feasibility, those studies do not take into consideration the cost of the intervention in relation to the outcomes it produces. Likewise, decision-makers need evidence related to cost as part of the evaluation of educational interventions (King, 1997). Cost effectiveness studies are one such means of analysis that offer both an evaluation of the effectiveness of an intervention as well as an evaluation of the ratio of cost to that of effectiveness.

However, one cannot make an overall determination of whether a program is worthwhile in an absolute sense; one can only infer that a program is "relatively more cost effective than other alternatives" (Karoly et al., 2001, p. 11).The results from the cost effectiveness analysis do not identify the best alternatives, but provide details to make an informed decision (Hummel-Rossi & Ashdown, 2002, p. 9). Finally, these studies are scarce inside educational research because of the complexities of data collection related to costs. In the next sections, we address the issue of identifying costs and ourcost effective analysis of the Read Malawi program.

Identifying Costs

Several different approaches to identifying costs are discussed in the literature; we elected to follow the model set forth by Levin and his colleagues (Levin and McEwan, 2001). Dubbed the "ingredients approach," thisapproach requires researchers to identifyelements required for any particular intervention, placing monetary values on each of those ingredients (Levin & McEwan, 2001). The program (as it was funded) was a three-year, development, and research activity with multiple phases and various activities. These activities included a development phase (in which we developed materials), an implementation phase, and evaluation and replication phase. At the time of this study, we had completed the first two. This multi-phased approach meant that some costs were spread across the various phases.

We isolated the costs associated with these two phases and then worked to further isolate the costs into activities within each phase.

Traditional cost effectiveness studies typically categorize costs into personnel, facilities, equipment and materials, other program inputs, and regular client input (Levin & McEwan, 2001). Thus, we began this process with an investigation of our itemized expenditures. Theseactualized expenditures allowed us to be much more effective in representing the true cost of the program, as compared to the use of a budget, which only provides for anticipated expenses (Levin & McEwan, 2001, p. 79). We then identified a comprehensive list of direct, indirect and contributed (in-kind) costs associated with the program; we were unable to gather costs not directly associated with the program, including costs of facilities and contributions.

Utilizing an inductive approach to our cost data, our team worked together to classify the cost data into activities, organized around: development; printing and delivery; implementation (training of teachers, head teachers and community); administrative; overhead; research and evaluation; general capacity and team building; and other. For this analysis, we used only the costs associated with potentially replicating the program, which excluded all costs related to development.

Identifying Effectiveness Measures

We used a variety of instruments to measure the effectiveness of the program. These instruments were used for both pre-implementation and post-implementation data. All instruments have been used in previous research studies (see Sailors et al., 2012; Sailors et al., 2014, for example) and are described in full in technical reports. An overview of the instruments follows:

Teacher questionnaire. This instrument probed teachers' attitudes and perceptions of their students and the teacher's views on the learning capacity of their students. It also probed for teachers' attitudes and perceptions about their own language and pedagogical skills and about the materials available to them. Finally, the teacher questionnaire probed teachers' attitudes and perceptions about their school and the communities from which their students drew.

Documenting instruction. This instrument documentedthe use of pedagogy associated with active teaching principles during literacy instruction.

Active learning, which stems from active teaching, is important to creating independent students who are empowered to continue learning outside of formal education. Active learning is the opposite of rote learning, and is rooted in the construction of knowledge rather than the memorization of knowledge. That is, students are in a constant state of understanding their world through the constructions they have and make during a lesson. There are several tenets that guide the principles of active learning, including (a) learning is social and is shared by a community; (b) learning moves from the known to the unknown; and (c) active students are risk-takers. Teachers who understand active learning understand that they must structure their instruction and teach in active ways that promote this type of learning for students. The active teaching instrument documented the instruction of the teacher and indicates whether (or not) there were elements of active teaching in the observed instruction.

The entire lesson was considered during data collection related to the rubrics. Data collectors read each statement listed in the active teaching rubric and made a decision whether they agreed (or not) that the teacher demonstrated elements listed in each of the statements on the rubric. A data collector could: Strongly Disagree; Disagree; Agree or Strongly Agree with each statement. The difference between Agree and Strongly Agree were degree-driven. That is, if an observer simply agreed that the teacher demonstrated the displayed behavior (just noted the behavior), the observer would mark "agree." However, if the observer felt strongly that the teacher was clearly demonstrating the behavior, the observer would mark, "strongly agree." The same held for "strongly disagree" and "disagree."

Studentliteracy assessment. The final instrument we used was the student literacy assessment. The instrument was designed to measure literacy constructs in Standard 1, 2 and 3 through an individual assessment including knowledge of vowels, letter combinations, word identification as well as sentence and passage reading. The instrument was designed to measure literacy constructs in Standard 3 through a group assessment including dictation and story writing.

Examining the role of directive coaching on implementation visits. Together with our colleagues at the MIE, we designed and validated monitoring and evaluation rubrics, following the work Hall & Hord (1987). These instruments documented the degree to which the innovation was being implemented.

These instruments included the (a) Artifact Analysis; (b) Student Interview; (c) Observation Checklist; (d) Levels of Use Interview; and (e) Concerns Assessment instruments. This data assisted us in shaping the program during the development phases as the tools were used by head teachers during in the ongoing assessment of implementation. Finally, the monitoring and evaluation data allowed us to examine the degree of implementation of the program as a moderating variable when examining instructional practices and student achievement. Monitoring and evaluation visits were once per term.

Training and reliability. Inter-rater reliability was established at various points throughout observer training. Fifteen observers were trained over the course of five days. The two days of training occurred at the Malawi Institute of Education in a classroom-based environment; observers were introduced to the theoretical framework of the study and to the use of the instruments. On the third day of training, observers were paired and placed in classrooms. Members of the research team (trainers) accompanied each of the pairs into classrooms and answered questions that the pairs may have had. At the end of each day, the teams met to compare collected data and to further clarify any questions.On the fourth day, observers were placed in non-study classrooms with the trainer. To evaluate interrater reliability, data points were compared between each pair of points of agreement between observers and the trainers. Each trainer was placed with two observers in a classroom, marking data independent of each other. Point-by-point agreement using Cohen's kappa statistic revealed an interrater reliability of .80 or better for each observer. Results of the interrater reliability analyses using the kappa coefficient were acceptable, with an estimate of $r \geq .80$ (Fleiss, 1981).The fifth day of training concerned the logistics of data collection. All data collectors were curriculum specialists and/or educational researchers, all were fluent in Chichewa, and all were former classroom teachers in Malawi.

Creating composite variables. We aggregated scores on the individual measures within each assessment. Below is an explanation of those aggregate scores.

Student achievement.We created composite scores for student achievement, teacher questionnaire, and observation by taking the row mean of all the items included in each composite.

We were, in essence, effectively imputing missing values by assigning the row mean based on the data available for that row (e.g., if one individual had three scores for that row out of a total possible of six, the mean was computed based on the six available scores).

Active teaching composite. To create this composite we took the standardized row mean of select questions from the teacher observation survey of the items. We then summed this variable with the standardized overall active teaching rating and re-standardized the total to create the active teaching composite variable.

Monitoring and evaluation data.For the purposes of this study, we used data from the final monitoring and evaluation visits. We ended up with final implementation ratings on 35 teachers (and 531 students) in this pilot post-test dataset. Each of these teachers was assigned a holistic rating for project implementation: "below expectations", "met expectations", or "exceeded expectations".

Establishing equivalency. We used preliminary regression analyses of our pretest to ensure that the schools randomly assigned to project and control treatments were equivalent at the outset in terms of the distribution of classroom practices and the perceptions. In a second set of analyses, using the posttest data set, we used regression analyses to examine the impact of the program on teachers' beliefs and perceptions, their instructional practices, and the engagement of their students. The teacher questionnaire and observation composite variables were regressed on a dummy variable for group (i.e., 1 = treatment group; 0 = delayed control group). To account for the potentially nested structure of the data (teachers nested in schools and schools nested in zones) on variance estimates, we used likelihood ratio tests to compare various multilevel and ordinary regression models to determine the best fitting models to the data.

Two- level models with random intercepts for schools emerged as the best fits to the data, and teachers' scores within the same zone did not appear to be correlated. To simplify the analysis, we used ordinary regression with robust standard errors throughout specifying schools as the clustering variable.

Analyzing the Cost Effectiveness

Cost Effective Analysis (CEA) assesses the effectiveness (*E*) of a particular program or intervention relative to the cost (C) of implementation.By combining information on effectiveness and costs, it is possible to determine which program provides a given level of effectiveness at the lowest/highest cost (Levin & McEwan, 2001). In other words, *E* represents the impact evaluation in which effects are expressed as an average treatment in a sample of individuals.The costs of an intervention, *C*, reflects the costs of additional resources used in the intervention, which include any resource with their appropriate estimate of the "opportunity cost".In other words, "using resources in one way, we are giving up the ability to use them in another way, so a cost has been incurred" (Levin & McEwan, 2001).

The ingredient approach, as previously discussed, consists of identifying all resources or ingredients consumed in an intervention and the valuation of each ingredient.In the specification of ingredients, four categories are included in the analysis: (1) personnel, (2) facilities, (3) equipment and materials, and (4) other program inputs.Hence, the calculation of total cost of the intervention includes those associated with administration, planning, training, monitoring,and supervision of the program.

The total cost estimates are adjusted for price inflation, time-value, and currency exchange when appropriate.Furthermore, when prices of the ingredients are expressed in nominal terms, the cost analysis considers different measures such as the consumer price index or the GDP deflator.In order to make the cost-effectiveness analysis comparable across years, it is necessary to convert future costs to their present value using standard procedures for these purposes (see for example Levin & McEwan 2001, pp. 92).

Once estimates on both the cost (*C*) and effectiveness (*E*) are available, the assessment is done in the form of cost-effectiveness ratios (CER) for the different alternatives resulted from the intervention (Levin, 1998).This relationship can be denoted in the following equation:

$$CER = \frac{C}{E}$$

This ratio indicates the cost required to obtain a single unit of effectiveness, as defined by the evaluator. Those ratios with smaller values are relatively more cost-effective which in turn indicates the best candidates for new investments.

Results

Obtaining Effect Sizes

Table1 shows the descriptive statistics and effect size calculations for each of the student's effectiveness measures by grade.In looking for significant differences of means between the treatment and control groups, we included a *t*-test statistic along with its significance level for each of the variables. To facilitate the comparisons of means between groups and across standards, we follow Cohen's standard interpretation (Cohen, 1988, pp. 25) of effect sizes.Thecorresponding classic t-test for the difference between two population means assumes that the two population variances are equal. This assumption is not always valid in the case of two independent samples where each of the sample variances differs. In order to make up for this variation, we used a pooled standard error, which accounts for one group of participants contributing more weight to the standard deviation than another does. Hence, our estimates account for sample variances as we aimed to obtain a better estimate of the variances in our target populations.

Table 1: Descriptive Statistics, Effect size and Significance Tests of Group Differences for Student Variables

Variable Name	Standard	Treatment Mean	SD	Control Mean	SD	Effect size	t-test
Chichewa Letter ID 1	1	0.47	(0.3445)	0.50	(0.3242)	-0.07	1.03
Chichewa Picture ID 1	1	0.91	(0.1519)	0.92	(0.1519)	-0.07	1.01
Chichewa Word Recognition 1	1	0.18	(0.1463)	0.20	(0.1622)	-0.11	1.60
Chichewa Dictation 1	1	0.10	(0.3478)	0.10	(0.3687)	-0.02	0.27
Chichewa Writing Accuracy 1	1	0.19	(0.6640)	0.18	(0.5687)	0.01	-0.12
Standard 1 Chichewa Literacy	1	-0.02	(1.0639)	0.02	(0.9321)	-0.06	0.83
Chichewa Letter ID 2	2	0.69	(0.3210)	0.72	(0.3134)	-0.09	1.34
Chichewa Picture ID 2	2	0.779	(0.1792)	0.790	(0.1669)	-0.06	0.90
Chichewa Word Recognition 2	2	0.24	(0.3414)	0.26	(0.3584)	0.35	1.03
Chichewa Dictation 2	2	0.48	(0.3478)	0.53	(0.3687)	-0.13	0.85
Chichewa Writing Accuracy 2	2	0.74	(1.4986)	0.85	(1.6918)	-0.07	1.02
Chichewa Syllable ID 2	2	0.36	(0.6705)	0.42	(0.7051)	-0.09	1.25
Chichewa Sentence 2	2	1.30	(2.0430)	1.08	(2.0248)	0.11	-1.61
Chichewa Word Recognition 2	2	4.59	(12.6421)	5.05	(13.3392)	-0.03	0.50
English Sentence 2	2	0.35	(1.4070)	0.30	(1.3351)	0.03	-0.49
English Word Reading 2	2	0.66	(3.2142)	0.80	(3.8415)	-0.04	0.58
English Picture ID 2	2	4.95	(0.4047)	4.93	(0.4505)	0.03	-0.41
English Dictation 2	2	0.20	(0.5120)	0.17	(0.4864)	0.07	-0.95
English Writing Accuracy 2	2	0.68	(1.8173)	0.52	(1.2895)	0.10	-2.08**
Standard 2 Chichewa Literacy	2	-0.04	(0.9729)	0.04	(1.0252)	-0.07	1.04
Standard 2 English Literacy	2	0.04	(1.0117)	-0.04	(0.9880)	0.07	-1.01
Chichewa Picture ID 3	3	1.00	(0.0690)	1.00	(0.0097)	-0.09	1.27
Chichewa Dictation 3	3	1.33	(1.1554)	1.38	(1.1278)	-0.04	0.58
Chichewa Mechanics 3	3	0.43	(0.5845)	0.46	(0.5945)	-0.05	0.71
Chichewa Composing 3	3	0.60	(0.8514)	0.62	(0.8419)	-0.02	0.32
Chichewa Sentence 3	3	5.35	(5.6913)	5.70	(5.7570)	-0.06	0.88
Chichewa Word Reading 3	3	16.97	(21.4742)	18.07	(21.3747)	-0.05	0.74
Chichewa Total Writing 3	3	1.03	(1.0787)	1.07	(1.4094)	-0.03	0.48
English Sentence 3	3	3.99	(4.9289)	4.25	(4.7488)	-0.05	0.77
English Word Reading 3	3	5.77	(11.9563)	4.54	(10.2993)	0.11	-1.59
English Dictation 3	3	0.70	(0.9067)	0.80	(0.9200)	-0.10	0.92
Standard 3 Chichewa Literacy	3	-0.03	(1.0415)	0.03	(0.9567)	-0.07	0.99
Standard 3 English Literacy	3	0.03	(1.0117)	-0.03	(0.9880)	0.07	-1.02

*p<0.10; ** p<0.05; ***p<0.001. Standard deviation in brackets.

The results suggested no significant differences between groups for the student variables, except for *English Writing Accuracy 2* in which a positive small effect size of 0.10 (p<0.05) was exhibited in favor of the treatment group.

When exploring the teacher questionnaire data, most of the variables demonstrated significant differences between groups, with an average effect size of 0.48 for the treatment group for all grades, see Table 2. The highest effect sizes were found for the variables *Availability* and *Readability of Complementary Books* with mean differences between treatment and control group of approximately 1.07 (p<0.001) and 1.03 (p<0.001), respectively.

Comparisons across grades are somewhat consistent although few variables showed no statistical significant differences in means, dependent upon which grade was observed. For example, in the case of the variables *Attitudes towards Other Teachers*, significant for grades one (p<0.001) and 2 (p<0.001); *Gender Attitudes*, significant for grades 2 (p<0.001) and 3 (p<0.001); and *Chichewa Capacity*, *English Capacity* and *General Attitudes*, which showed no significance at grade 3.

While teachers in grade 1 tended to exhibit the largest effects on *Chichewa Capacity* (p<0.001), and *Textbooks Use* (p<0.001), the largest effects for grade two teachers were found in *Availability of Complementary Books* (p<0.001) *and Readability of Complementary Books* (p<0.001). Grade three teachers exhibited the largest effects on two variables: *Text Books Quality* (p<0.001)and *Internal Professionalism* (p<0.001).

Table 2: Descriptive Statistics, Effect Size and Significance Tests of Group Differences for Teacher Variables

Variable Name	All Standards				STANDARD 1				STANDARD 2				STANDARD 3			
	Proj Mean	Control	Effect size	t-test	Proj Mean	Control	Effect size	t-test	Proj Mean	Control	Effect size	t-test	Proj Mean	Control	Effect size	t-test
AvailCompBooks	2.91	2.05	1.07	-26.02***	2.59	2.04	0.81	-11.53***	3.16	1.96	1.59	-22.23***	2.90	2.15	0.82	-12.81***
	(0.81)	(0.80)			(0.66)	(0.70)			(0.67)	(0.82)			(0.96)	(0.86)		
ReadCompBooks	2.88	2.05	1.03	-25.10***	2.57	2.01	0.81	-11.49***	3.15	2.01	1.51	-21.10***	2.95	2.13	0.89	-12.43***
	(0.84)	(0.78)			(0.64)	(0.72)			(0.68)	(0.80)			(1.02)	(0.81)		
Textbooks Quality	3.18	2.72	0.85	-20.69***	3.12	2.72	0.80	-11.43***	3.13	2.65	0.86	-11.99***	3.27	2.80	0.89	-12.37***
	(0.53)	(0.54)			(0.46)	(0.51)			(0.56)	(0.56)			(0.55)	(0.51)		
Self Efficacy	3.56	3.37	0.47	-11.49***	3.54	3.35	0.55	-7.78****	3.60	3.55	0.13	-7.76***	3.557	3.419	0.34	-4.56***
	(0.39)	(0.41)			(0.37)	(0.32)			(0.38)	(0.39)			(0.38)	(0.43)		
My English	3.40	3.16	0.42	-10.13***	3.34	3.06	0.47	-6.60***	3.36	3.23	0.23	-3.27***	3.50	3.19	0.55	-7.67***
	(0.53)	(0.63)			(0.56)	(0.62)			(0.55)	(0.61)			(0.47)	(0.64)		
Community	2.42	2.17	0.39	-9.50****	2.24	2.34	-0.18	2.54**	2.54	2.11	0.69	-9.72***	2.416	2.051	0.54	-8.76***
	(0.72)	(0.53)			(0.58)	(0.53)			(0.73)	(0.48)			(0.79)	(0.54)		
Chichewa Capacity	3.50	3.27	0.39	-9.40***	3.52	3.07	0.77	-10.91***	k	3.25	0.34	-4.73***	3.50	3.49	0.02	-0.25
	(0.52)	(0.66)			(0.45)	(0.68)			(0.55)	(0.71)			(0.55)	(0.49)		
General Attitudes	2.90	2.85	0.11	-2.80***	2.95	2.89	0.18	-2.50**	2.88	2.83	0.15	-2.08**	2.85	2.83	0.05	-0.70
	(0.39)	(0.41)			(0.30)	(0.38)			(0.36)	(0.38)			(0.48)	(0.47)		
Gender Attitudes	2.66	2.57	0.15	-3.41***	2.70	2.51	0.33	-4.55***	2.58	2.57	0.02	-0.23	2.72	2.67	0.07	-0.91
	(0.59)	(0.63)			(0.66)	(0.50)			(0.45)	(0.69)			(0.64)	(0.69)		
English Capacity	2.25	2.14	0.19	-4.72***	2.19	2.08	0.19	-2.69***	2.40	2.20	0.33	-4.61***	2.19	2.16	0.05	-0.69
	(0.56)	(0.58)			(0.57)	(0.56)			(0.56)	(0.66)			(0.53)	(0.50)		
Chichewa Capacity	3.16	3.01	0.24	-5.93***	3.10	3.00	0.19	-2.72***	3.21	2.93	0.50	-7.02***	3.18	3.13	0.07	-0.96
	(0.52)	(0.65)			(0.41)	(0.61)			(0.39)	(0.67)			(0.68)	(0.66)		
Attitude toward Other Tch	3.50	3.34	0.29	-7.01***	3.46	3.32	0.26	-3.68	3.49	3.36	0.25	3.46	3.55	3.34	0.35	-4.86***
	(0.59)	(0.51)			(0.59)	(0.48)			(0.56)	(0.47)			(0.61)	(0.57)		
Textbooks Quantity	2.85	2.67	0.32	-7.84***	2.80	2.79	0.03	-0.43	2.92	2.837321	0.17	-2.36**	2.84	2.40	0.80	-11.16***
	(0.52)	(0.55)			(0.46)	(0.57)			(0.51)	(0.44)			(0.59)	(0.52)		
Text Books Use	3.41	3.16	0.54	-13.20***	3.45	3.16	0.75	-10.84***	3.45	3.30	0.33	-4.53***	3.34	3.04	0.63	-8.77***
	(0.44)	(0.46)			(0.36)	(0.40)			(0.41)	(0.49)			(0.51)	(0.47)		
Internal Professionalism (Prof)	3.19	2.91	0.49	-11.89***	3.21	3.04	0.37	-5.23***	3.20	2.92	0.47	-6.60***	3.17	2.77	0.75	-8.76***
	(0.44)	(0.67)			(0.34)	(0.56)			(0.46)	(0.67)			(0.04)	(0.75)		
External Prof	3.40	3.21	0.28	-6.89***	3.33	3.28	0.08	-1.12	3.45	3.17	0.42	-5.91***	3.42	3.20	0.33	-4.59***
	(0.62)	(0.66)			(0.64)	(0.52)			(0.66)	(0.66)			(0.56)	(0.79)		
Prof	3.36	3.07	0.50	-12.15***	3.31	3.16	0.31	-12.15***	3.40	3.06	0.57	-8.04***	3.37	3.00	0.61	-8.28***
	(0.50)	(0.62)			(0.49)	(0.47)			(0.53)	(0.63)			(0.41)	(0.73)		

*p<0.10; ** p<0.05; ***p<0.001. Standard deviation in brackets.

The meandifferences between groups of classroom observations are shown in Table 3. All variables were statistical significance across grades, except for *Text Variety* in grade 1 and 3, *Curricular* in grade 2, and *Writing* in grade 3.

Table 3: Descriptive Statistics, Effect Size and Significance Tests of Group Differences for Observation Variables

Variable Name	All Standards Group				STANDARD 1 Group				STANDARD 2 Group				STANDARD 3 Group			
	Project Mean	Control	Effect size	t-test	Project Mean	Control	Effect size	t-test	Project Mean	Control	Effect size	t-test	Project Mean	Control	Effect size	t-test
	(1.02)	(1.08)			(1.02)	(1.08)			(0.88)	(1.04)			(1.15)	(1.11)		
Class size	122.4	96.38	0.53	-13.47***	124.3	99.76	0.55	-7.92***	105.8	83.57	0.67	-7.3***	137.12	99.89	0.60	-8.85***
	(55.31)	(41.69)			(48.88)	(40.64)			(36.65)	(29.52)			(70.52)	(51.22)		
Hands on Print	35.94	35.11	0.03	-0.77	28.74	37.58	-0.32	4.61***	41.88	37.18	0.18	-2.35**	37.74	30.25	0.30	-4.22***
	(26.09)	(26.48)			(24.30)	(29.62)			(25.62)	(26.06)			(26.71)	(22.45)		
Text Variety	2.59	2.53	0.08	-1.93*	2.66	2.61	0.07	-1.00	2.73	2.60	0.21	-2.9***	2.40	2.40	0.01	-0.14
	(0.645)	(0.70)			(0.69)	(0.65)			(0.55)	(0.65)			(0.62)	(0.79)		
Curricular	2.95	2.94	0.04	-0.86	2.94	2.81	0.29	-4.15***	3.04	3.05	-0.03	0.37	2.90	2.96	-0.14	1.95*
	(0.400)	(0.528)			(0.45)	(0.43)			(0.36)	(0.61)			(0.35)	(0.49)		
Writing	12.25	15.44	-0.19	4.53***	14.25	18.36	-0.20	2.88***	13.63	18.60	-0.33	4.65***	8.88	8.95	0.00	0.06
	(15.59)	(18.45)			(17.15)	(22.53)			(12.01)	(17.35)			(16.53)	(11.94)		

*p<0.10; ** p<0.05; ***p<0.001. Standard deviation in brackets

Effect Sizes by Optimal Implementation Rating

By incorporating implementation rating in the calculation of effect size, an improvement in the mean of the treatment group is observed for some of the variables.As the intervention program gets closer of meeting or exceeding the expectations, these improvements are significant and in certain cases higher positive effects are found for variables that otherwise were not showing significant effect sizes between groups.Such effects are identified of greater importance for the set of 'student' variables.

The effect size estimates for student variables as the program was implemented (by optimal implementation rating) is displayed in Table 4.It also includes the respective statistical test for mean differences across implementation levels.The findings indicate that 16 out of 33 variables now exhibit statistical significant effects resulting in higher means for the treatment as compared with the control group. Specifically, positive and statistical significant effects of higher implementation ratings on literacy assessment variables are present in *Chichewa Letter ID* 1 (p<0.10), *Chichewa Word Recognition* 1 (p<0.05), *Chichewa Writing Accuracy* 1 (p<0.001), *Grade 1Chichewa Literacy* (p<0.001), *English Writing Accuracy* 2 (p<0.10), and for the complete list of variables in the grade 3.Among them, *English Word Reading* 3 (p<0.001) exhibits the largest effect size between groups.

Table 4: Effect Size Estimates by Optimal Implementation Rating for Student Variables

| | As Implemented | Optimal Implementation Rating | | | |
		1	2	3	F-Statistic
Chichewa Letter ID 1	-0.07	-0.17	0.24	0.23	2.93*
Chichewa Picture ID 1	-0.07	0.01	-0.21	0.25	1.74
Chichewa Word Recognition 1	-0.11	-0.04	-0.08	0.42	4.39**
Chichewa Dictation 1	-0.02	-0.02	0.19	-0.03	0.76
Chichewa Writing Accuracy 1	0.01	-0.15	-0.15	0.70	5.58***
Standard 1 Chichewa Literacy	-0.06	-0.12	-0.02	0.60	4.27***
Chichewa Letter ID 2	-0.09	0.22	-0.25	-0.54	6.53***
Chichewa Picture ID 2	-0.06	-0.01	-0.17	-0.05	0.87
Chichewa Word Recognition 2	-0.07	-0.23	-0.16	-0.17	0.13
Chichewa Dictation 2	-0.13	0.30	-0.50	-0.41	3.03*
Chichewa Writing Accuracy 2	-0.07	-0.27	-0.03	0.04	1.6
Chichewa Syllable ID 2	-0.09	-0.05	-0.22	-0.13	0.7
Chichewa Sentence 2	0.11	0.06	0.02	-0.03	0.1
Chichewa Word Recognition 2	-0.03	-0.04	-0.16	-0.08	0.34
English Sentence 2	0.03	0.00	0.06	0.08	0.09
English Word Reading 2	-0.04	0.03	-0.07	0.61	0.22
English Picture ID 2	0.03	-0.04	0.06	-0.01	0.29
English Dictation 2	0.07	0.03	0.12	-0.03	0.35
English Writing Accuracy 2	0.10	0.13	-0.08	0.17	2.36*
Standard 2 Chichewa Literacy	-0.07	-0.08	-0.19	-0.17	0.24
Standard 2 English Literacy	0.07	0.02	0.65	-0.24	0.17
English Word Recognition 2	0.11	0.09	0.06	-0.10	0.52
Chichewa Dictation 3	-0.04	-0.06	-0.23	0.48	5.22***
Chichewa Mechanics 3	-0.05	-0.08	-0.10	0.56	5.36***
Chichewa Composing 3	-0.02	-0.18	-0.08	0.52	5.58***
Chichewa Sentence 3	-0.06	-0.03	-0.15	0.43	3.76***
Chichewa Word reading 3	-0.05	-0.05	-0.11	0.36	2.52*
Chichewa Total Writing 3	-0.03	-0.14	-0.09	0.55	5.63***
English Sentence 3	-0.05	-0.06	-0.16	0.50	4.84***
English Word Reading 3	0.11	0.03	-0.03	0.79	6.58***
English Dictation 3	-0.10	-0.05	-0.30	0.36	5.66***
Standard 3 Chichewa Literacy	-0.07	-0.06	-0.16	0.48	4.57**
Standard 3 English Literacy	0.07	0.12	-0.15	0.50	4.58**

*$p < 0.10$; ** $p < 0.05$; ***$p < 0.001$

Teacher variables also exhibited significant mean differences by optimal implementation rating, as noted in Table 5. Note that while only two significant large effects sizes are present when excluding implementation ratings, once the program "meets the expected" or "exceeds" optimal implementation, significant large effects are found for the treatment group in four variables. These are *My English* ($p < 0.001$), *Text Book Use* ($p < 0.001$), *Readability of Complementary Books* ($p < 0.001$), and *Community* ($p < 0.001$).

Table 5: Effect size by Degree of Implementation for Teacher Variables

	As Implemented	Optimal Implementation Rating			
		1	2	3	F-Statistic
My English	0.42	0.32	0.52	0.81	9.82***
My Chichewa	0.39	0.23	0.42	0.37	2.72*
Self Efficacy	0.47	0.26	0.71	0.44	13.51***
General Attitudes	0.12	0.29	0.05	0.55	15.45***
Gender Attitudes	0.15	0.03	0.23	0.65	13.26***
English Capacity	0.19	0.41	0.32	0.38	0.40
Chichewa Capacity	0.24	0.15	0.51	0.15	19.66***
Attitudes towards Other Teachers	0.29	0.05	0.65	0.43	18.59***
Textbooks Quantity	0.32	0.30	0.45	0.49	1.77
Text Books Use	0.54	0.52	0.90	0.24	17.52***
Text Books Quality	0.85	0.99	1.08	1.04	0.37
Availability of Complementary Books	1.07	1.06	1.15	1.19	0.74
Readability of Complementary Books	1.03	0.83	1.22	1.23	6.56***
Internal Professionalism	0.49	0.48	0.57	0.67	2.3
External Professionalism	0.28	0.40	0.46	0.69	3.57**
Professionalism	0.50	0.58	0.66	0.75	1.4
Community	0.39	0.59	0.51	1.00	3.34**

$*p<0.10; ** p<0.05; ***p<0.001$

We obtained similar findings when we explored effect size by implementation levels in the observation variables (see Table 6). Higher levels of implementation resulted in significant positive effects in almost all the variables with exception of *Text Variety*.

Table 6: Effect size by Degree of Implementation for Teacher Variables

	As Implemented	Optimal Implementation Rating			
		1	2	3	F-Statistic
Class size	0.53	0.93	0.57	0.47	4.6**
Hands on Print	0.03	0.00	0.45	0.19	10.34***
Dispositional	0.10	0.35	0.08	0.50	12.96***
Text Variety	0.08	0.28	0.22	-0.06	3.39**
Curricular	0.04	0.48	-0.12	0.05	29.36***
Writing	-0.19	-0.16	-0.13	0.19	6.99***
Active	0.13	0.38	0.21	0.46	5.05***

$*p<0.10; ** p<0.05; ***p<0.001$

Cost Effectiveness Results

As described earlier, the cost effectiveness of any intervention can be measured by assessing the relationship between costs (C) and effects (E). This is done by calculating cost-effectiveness ratio (CER) previously defined. A decision rule for the more cost-effective alternative is to choose those with the lowest cost per unit of effectiveness. The present analysis focuses on identifying the most cost-effective alternatives under the argument that the Read Malawi Program met (optimal implementation 2) or exceeded expectations (optimal implementation 3) regarding the implementation of the project.

We also offer an alternative interpretation of the results by calculating effectiveness-cost ratio aiming to facilitate the exposition of findings. In other words, we include estimates of the effect sizes for each $10 of cost per student, computed by dividing each effect size (E) by the pertinent cost per student (C) and multiplying by 10. The $10 figure will then serve as standard unit of expenditure, which allows us to compare the derived cost-effectiveness ratios across the intervention.

In this case, the larger the effect sizes on this expenditure level, the greater the educational impact of resources on achievement; if properly interpreted, there is no difference in the conclusions produced by calculating cost-effectiveness (C/E) or effectiveness-cost (E/C) ratios (Levin & McEwan, 2001).

Table 7 provides the CER ratios for those student variables with small or above positive effect sizes.Among the different effectiveness measures with an optimal implementation rating of two, *Grade two English Literacy* resulted as the more cost-effective, implying that for one standard deviation increase there is an associated cost of $8.20. Alternatively, an expenditure of 10 dollars brings a rise of 1.2 standard deviations on this variable.*Chichewa Letter ID 1* shows the second lowest CER, suggesting a cost of $22.20 per standard deviation, or what is traduced in 0.5 standard deviation increases for each $10 expenditure cost.

Table 7. Cost Effectiveness Ratio Estimates by Optimal Implementation for Student Variables

	Effect Size	Cost per Student	Cost-Effectiveness Ratio (C/E)	Effectiveness-Cost Ratio (E/C)	Effectiveness-Cost Ratio (E/C)[a]
A. Optimal Implementation 2					
Chichewa Letter ID 1	0.24	5.36	22.2	0.05	0.5
Chichewa Dictation 1	0.19	5.36	28.9	0.03	0.3
Chichewa Sentence 2	0.02	5.37	240.3	0	0
English Sentence 2	0.06	5.37	90.8	0.01	0.1
English Picture ID 2	0.06	5.37	87.7	0.01	0.1
English Dictation 2	0.12	5.37	46	0.02	0.2
Standard 2 English Literacy	0.65	5.37	8.2	0.12	1.2
English Word Recognition 2	0.06	5.37	90.8	0.01	0.1
B. Optimal Implementation 3					
Chichewa Letter ID 1	0.23	5.36	23.72	0.04	0.42
Chichewa Picture ID 1	0.25	5.36	21.31	0.05	0.47
Chichewa Word Recognition 1	0.42	5.36	12.91	0.08	0.77
Chichewa Writing Accuracy 1	0.7	5.36	7.68	0.13	1.3
Standard 1 Chichewa Literacy	0.6	5.36	8.87	0.11	1.13
English Sentence 2	0.08	5.37	67.67	0.01	0.15
English Word Reading 2	0.61	5.37	8.85	0.11	1.13
English Writing Accuracy 2	0.17	5.37	32.36	0.03	0.31
Chichewa Dictation 3	0.48	5.37	11.23	0.09	0.89
Chichewa Mechanics 3	0.56	5.37	9.57	0.1	1.05
Chichewa Composing 3	0.52	5.37	10.34	0.1	0.97
Chichewa Sentence 3	0.43	5.37	12.51	0.08	0.8
Chichewa Word Reading 3	0.36	5.37	14.75	0.07	0.68
Chichewa Total Writing 3	0.55	5.37	9.83	0.1	1.02
English Sentence 3	0.5	5.37	10.64	0.09	0.94
English Word Reading 3	0.79	5.37	6.83	0.15	1.46
English Dictation 3	0.36	5.37	14.86	0.07	0.67
Standard 3 Chichewa Literacy	0.48	5.37	11.22	0.09	0.89
Standard 3 English Literacy	0.5	5.37	10.74	0.09	0.93

[a] For each $10 cost per Student

CER calculations when the program "*exceeded expectations*" in relation to implementation rating (optimal 3), not only more effectiveness measures exhibits positive effects derived from the intervention, but also their magnitude is relatively higher as compared to other implementation ratings.This in turn is associated with lower cost effective ratios, as for example, the associated cost for one standard deviation increase of *English Word Reading 3* isapproximately $6.83, which is among the lowest of all effectiveness measures.In other words, for each $10 dollars spent it would be possible to raise students'*English Word Reading 3* about 1.46 standard deviations.For the same $10 expenditure cost, four additional variables were found which effectiveness-ratio would increase more than one standard deviation, including *Chichewa Writing Accuracy 1, Grade 1 Chichewa Literacy, English Word Reading 2, Chichewa Mechanics 3*.

Cost effectiveness ratios by project implementation ratings are also calculated for teacher variables, see Table 8. The calculations point out the lowest ratio for *Readability of Complementary Books* with cost of $40.40 per one standard deviation increase when the program implementation "meets" (optimal 2) or "exceeds" (optimal 3) expectations.

This also indicates that for each $10 of expenditure, a gain of approximately one-quarter standard deviation is reported. Somewhat similar results are found with optimal implementation 3 for the variables *Availability of Complementary Books* and *Text Books Quality* with a cost per standard deviation of $43.00 and $45.00, respectively.

Table 8. Cost Effectiveness Ratio Estimate by Optimal Implementation For Teacher Variables

	Effect Size	Cost per Teacher	Cost-Effectiveness Ratio (C/E)	Effectiveness-Cost Ratio (E/C)	Effectiveness-Cost Ratio (E/C)[a]
A. Optimal Implementation 2					
My English	0.52	49.3	95.6	0.01	0.10
My Chichewa	0.42	49.3	118.2	0.01	0.08
Self Efficacy	0.71	49.3	69.7	0.01	0.14
General Attitudes	0.05	49.3	1087.8	0.00	0.01
Gender Attitudes	0.23	49.3	210.3	0.00	0.05
English Capacity	0.32	49.3	154.4	0.01	0.06
Chichewa Capacity	0.51	49.3	96.0	0.01	0.10
Attitudes towards Other Teachers	0.65	49.3	76.0	0.01	0.13
Textbooks Quantity	0.45	49.3	108.7	0.01	0.09
Text Books Use	0.90	49.3	54.6	0.02	0.18
Text Books Quality	1.08	49.3	45.5	0.02	0.22
Availability of Complementary Books	1.15	49.3	43.0	0.02	0.23
Readability of Complementary Books	1.22	49.3	40.4	0.02	0.25
Internal Professionalism	0.57	49.3	86.1	0.01	0.12
External Professionalism	0.46	49.3	107.8	0.01	0.09
Professionalism	0.66	49.3	74.9	0.01	0.13
Community	0.51	49.3	96.5	0.01	0.10
B. Optimal Implementation 3					
My English	0.81	49.3	60.9	0.02	0.16
My Chichewa	0.37	49.3	132.0	0.01	0.08
Self Efficacy	0.44	49.3	111.6	0.01	0.09
General Attitudes	0.55	49.3	90.5	0.01	0.11
Gender Attitudes	0.65	49.3	76.3	0.01	0.13
English Capacity	0.38	49.3	129.0	0.01	0.08
Chichewa Capacity	0.15	49.3	336.2	0.00	0.03
Attitudes towards Other Teachers	0.43	49.3	115.7	0.01	0.09
Textbooks Quantity	0.49	49.3	100.9	0.01	0.10
Text Books Use	0.24	49.3	208.6	0.00	0.05
Text Books Quality	1.04	49.3	47.5	0.02	0.21
Availability of Complementary Books	1.19	49.3	41.6	0.02	0.24
Readability of Complementary Books	1.23	49.3	40.1	0.02	0.25
Internal Professionalism	0.67	49.3	73.3	0.01	0.14
External Professionalism	0.69	49.3	71.3	0.01	0.14
Professionalism	0.75	49.3	66.0	0.02	0.15
Community	1.00	49.3	49.2	0.02	0.20

[a] For each $10 cost per teacher

Projecting CER Results

The next set of results involves projecting CE ratios of variables when increasing the length of the intervention. As discussed before, the program was evaluated within the first 4 months of the intervention limiting, to some extent, the total effects for the outcomes variables considered in the analysis.In order to address this issue, we created a projection of CER for each of the students, teacher, and observational variables.In doing this, we first estimate the effect size for each additional month of intervention. This effect size is then extrapolated to predict the associated effects of additional 4 and 8 months of intervention.This type of procedure has been found in different studies, see for example Thompson (1999), Seng (2005), and Camilli et al. (2010). Cost effective ratios are then computed by using the appropriate estimated cost per outcome for these different scenarios.

The projected cost-effectiveness ratios for student variables by optimal implementation rating two and three are showed in Table 9.

It displays information of effect sizes and cost-effectiveness ratios obtained as: the program is implemented; as well as effect size projections by 1, 4, and 8 additional months of intervention; and finally, projected cost-effectiveness and effectiveness cost-ratio related for 8 additional months of intervention.

In analyzing effect size projections for optimal implementation 2, a projected increase of almost 2 standard deviations is observed in *Grade 2 English Literacy* if implementation of the program continued for an additional 8 months.Furthermore, considerable reductions with respect to cost-effective ratios are obtained for the same additional months of intervention.

The cost associated with one standard deviation increase for *Grade 2 English Literacy* dropped from $8.20 to $2.73, while *Chichewa Letter ID 1* costs was reduced from $22.20 to $7.39 and *Chichewa Dictation 1* decreased from $28.90 to $9.64.

For optimal implementation 3, the results indicate that 15 out of 19 student variables show projected cost ratios lower than $5.00 per one standard increase on the respective variable, two variables with a cost range from $5.00 - $10.00, and two variables with greater than $10.Among all variables included in this Table, *English Word Reading 3* appears as the most effective-cost measure as by $10 expenditure cost it is projected a positive effect of approximately 4.4 standard deviations.

Table 9. Projected Cost Effective Ratio by Optimal Implementation for Student Variables

			As Implemented		Effect Size Projections by Additional Months of Intervention		Projected Cost Ratios	
	Effect Size	Cost-Effectiveness Ratio (C/E)	Effectiveness Cost Ratio (E/C)[a]	1	4	8	Cost Effectiveness Ratio (C/E) of 8 months intervention	Effectiveness-Cost Ratio (C/E) of 8 months of intervention
A. Optimal Implementation 2								
Chichewa Letter ID 1	0.24	22.2	0.45	0.30	0.73	0.89	6.02	1.66
Chichewa Dictation 1	0.19	28.9	0.35	0.23	0.35	0.40	13.40	0.75
Chichewa Sentence 2	0.02	240.3	0.04	0.03	0.19	0.25	21.48	0.47
English Sentence 2	0.06	90.8	0.11	0.07	0.08	0.09	59.67	0.17
English Picture ID 2	0.06	87.7	0.11	0.08	0.10	0.11	48.82	0.20
English Dictation 2	0.12	46	0.22	0.15	0.27	0.32	16.78	0.60
Standard 2 English Literacy	0.65	8.2	1.22	0.82	0.79	0.84	6.39	1.56
EnglishWord Recognition 2	0.06	90.8	0.11	0.07	0.09	0.10	53.70	0.19
B. Optimal Implementation 3								
Chichewa Letter ID 1	0.23	23.72	0.42	0.28	0.72	0.88	6.09	1.64
Chichewa Picture ID 1	0.25	21.31	0.47	0.31	1.01	1.16	4.62	2.16
Chichewa Word Recognition 1	0.42	12.91	0.77	0.52	0.61	0.68	7.88	1.27
Chichewa Writing Accuracy 1	0.70	7.68	1.3	0.87	0.79	0.82	6.54	1.53
Standard 1 Chichewa Literacy	0.60	8.87	1.13	0.76	0.77	0.83	6.46	1.55
English Sentence 2	0.08	67.67	0.15	0.10	0.13	0.14	38.36	0.26
English Word Reading 2	0.61	8.85	1.13	0.76	0.66	0.67	8.01	1.25
English Writing Accuracy 2	0.17	32.36	0.31	0.21	0.30	0.35	15.34	0.65
Chichewa Dictation 3	0.48	11.23	0.89	0.60	0.82	0.94	5.71	1.75
Chichewa Mechanics 3	0.56	9.57	1.05	0.70	0.73	0.78	6.88	1.45
Chichewa Composing 3	0.52	10.34	0.97	0.65	0.73	0.80	6.71	1.49
Chichewa Sentence 3	0.43	12.51	0.8	0.54	1.19	1.40	3.84	2.61
Chichewa Word Reading 3	0.36	14.75	0.68	0.46	1.08	1.39	3.86	2.59
Chichewa Total Writing 3	0.55	9.83	1.02	0.68	0.92	1.05	5.11	1.96
English Sentence 3	0.50	10.64	0.94	0.63	1.45	1.76	3.05	3.28
English Word Reading 3	0.79	6.83	1.46	0.98	2.05	2.47	2.17	4.60
English Dictation 3	0.36	14.86	0.67	0.45	0.56	0.62	8.66	1.15
Standard 3 Chichewa Literacy	0.48	11.22	0.89	0.60	0.87	1.01	5.32	1.88
Standard 3 English Literacy	0.50	10.74	0.93	0.62	0.97	1.13	4.75	2.10

[a] For each $10 cost per student

Teacher variable projections are displayed in Table 10. After cost-effectiveness ratios are projected for eight additional months of intervention, among all variables in optimal implementation ratings 2 and 3, the two more cost-effective are *Readability of Complementary Books* and *Availability of Complementary Books.* The projections suggest that by any increase of one standard deviation on each of these variables, the cost associated with it approximates $13.36 and $13.86, respectively. These would represent a reduction of approximately $26.70 and $33.60, respectively, when comparing those cost-effectiveness ratios obtained from the first 4 months of intervention. Furthermore, the interpretation of projected effectiveness-cost measures suggests that a $10 cost per teacher would result in a rise of these two variables in almost three-quarters of standard deviations.

Table 10: Cost Effective Ratio Estimates by Optimal Implementation for Teacher Variables

	As Implemented			Effect Size Projections by Additional Months of Intervention			Projected Cost Ratios	
	Effect Size	Cost-Effectiveness Ratio (C/E)	Effectiveness Cost Ratio (E/C)[a]	1	4	8	Cost Effectiveness Ratio (C/E) of 8 months intervention	Effectiveness-Cost Ratio (C/E) of 8 months intervention
A. Optimal Implementation 2								
My English	0.52	95.6	0.10	0.64	1.03	1.55	31.88	0.31
My Chichewa	0.42	118.2	0.08	0.52	0.83	1.25	39.38	0.25
Self Efficacy	0.71	69.7	0.14	0.88	1.42	2.12	23.23	0.43
General Attitudes	0.05	1087.8	0.01	0.06	0.09	0.14	362.59	0.03
Gender Attitudes	0.23	210.3	0.05	0.29	0.47	0.7	70.09	0.14
English Capacity	0.32	154.4	0.06	0.4	0.64	0.96	51.49	0.19
Chichewa Capacity	0.51	96.0	0.10	0.64	1.03	1.54	31.99	0.31
Attitudes towards Other Teachers	0.65	76.0	0.13	0.81	1.3	1.95	25.35	0.39
Textbooks Quantity	0.45	108.7	0.09	0.57	0.91	1.36	36.22	0.28
Text Books Use	0.90	54.6	0.18	1.13	1.81	2.71	18.2	0.55
Text Books Quality	1.08	45.5	0.22	1.36	2.17	3.25	15.15	0.66
Availability of Complementary Books	1.15	43.0	0.23	1.44	2.3	3.44	14.32	0.70
Readability of Complementary Books	1.22	40.4	0.25	1.53	2.44	3.66	13.47	0.74
Internal Professionalism	0.57	86.1	0.12	0.72	1.14	1.72	28.72	0.35
External Professionalism	0.46	107.8	0.09	0.57	0.92	1.37	35.93	0.28
Professionalism	0.66	74.9	0.13	0.82	1.32	1.98	24.96	0.40
Community	0.51	96.5	0.10	0.64	1.02	1.53	32.18	0.31
B. Optimal Implementation 3								
My English	0.81	60.9	0.16	1.01	1.62	2.43	20.31	0.49
My Chichewa	0.37	132.0	0.08	0.47	0.75	1.12	43.99	0.23
Self Efficacy	0.44	111.6	0.09	0.55	0.88	1.33	37.18	0.27
General Attitudes	0.55	90.5	0.11	0.68	1.09	1.64	30.16	0.33
Gender Attitudes	0.65	76.3	0.13	0.81	1.29	1.94	25.43	0.39
English Capacity	0.38	129.0	0.08	0.48	0.76	1.15	43.00	0.23
Chichewa Capacity	0.15	336.2	0.03	0.18	0.29	0.44	112.06	0.09
Attitudes towards Other Teachers	0.43	115.7	0.09	0.53	0.85	1.25	38.56	0.26
Textbooks Quantity	0.49	100.9	0.1	0.61	0.98	1.47	33.63	0.30
Text Books Use	0.24	208.6	0.05	0.3	0.47	0.71	69.53	0.14
Text Books Quality	1.04	47.5	0.21	1.3	2.08	3.11	15.84	0.63
Availability of Complementary Books	1.19	41.6	0.24	1.48	2.37	3.56	13.86	0.72
Readability of Complementary Books	1.23	40.1	0.25	1.54	2.46	3.69	13.36	0.75
Internal Professionalism	0.67	73.3	0.14	0.84	1.35	2.02	24.43	0.41
External Professionalism	0.69	71.3	0.14	0.86	1.38	2.07	23.77	0.42
Professionalism	0.75	66.0	0.15	0.93	1.49	2.24	22.00	0.45
Community	1.00	49.2	0.2	1.25	2.01	3.01	16.39	0.61

[a] For each $10 cost per teacher

With respect of observational variables, the respective projections are showed in Table 11. In general, all variables experienced significant reductions upon cost-effectiveness ratio after eight additional months.

Table 11: Cost Effective Ratio Estimates by Optimal Implementation for Observational Variables

	As Implemented			Effect Size Projections by Additional Months of Intervention			Projected Cost Ratios	
	Effect Size	Cost-Effectiveness Ratio (C/E)	Effectiveness Cost Ratio (E/C)	1	4	8	Cost Effectiveness Ratio (C/E) of 8 months intervention	Effectiveness-Cost Ratio (C/E) of 8 months intervention
A. Optimal Implementation 2								
Class size	0.57	97	0.06	0.71	1.13	1.8	32.2	0.31
Hands on Print	0.45	122	0.04	0.56	0.9	1.35	40.5	0.25
Dispositional	0.08	715	0.00	0.10	0.15	0.23	238.2	0.04
Text Variety	0.22	251	0.01	0.27	0.4	0.65	83.6	0.12
Active	0.21	266	0.01	0.26	0.41	0.62	88.7	0.11
B. Optimal Implementation 3								
Class size	0.47	115	0.09	0.59	0.95	1.42	38.4	0.26
Hands on Print	0.19	282	0.04	0.24	0.39	0.58	93.8	0.11
Dispositional	0.50	109	0.09	0.63	1.01	1.51	36.3	0.28
Curricular	0.05	1076	0.01	0.06	0.1	0.15	358.5	0.03
Writing	0.19	295	0.03	0.23	0.37	0.56	98.3	0.10
Active	0.46	118	0.08	0.58	0.93	1.39	39.4	0.25

Discussion: Implementation Matters

The purpose of this study was to provide evidence related to the cost effectiveness of the educational innovation under discussion. That is, we set out to investigate if the effectivenessof the Read Malawi program justified the costs of the same program. We used a quasi-experimental, pre- post-test design to measure the effectiveness of the program. We also created a cost effectiveness ratio and a relative cost-effective ratio of the innovation and compared it to the business-as-usual model as a way of comparing the results of the innovative program. In this section, we discuss the possible meanings of our findings as well as the implications for education in Malawi (and in other countries like Malawi).

We also discuss the implications of our findings for policy makers, in Malawi and countries like Malawi, and in other countries who sponsor work like ours.

While there were very little student effects in the short four-month intervention period (with only one variable being statistically significant), our explorations around implementation by teachers and the ways in which the levels of implementation influence student achievement are promising. Our findings suggest that after only a four-month intervention, students in classrooms where teachers are implementing at the highest levels (rating of optimal) outperform students in classrooms where teachers are implementing at "below" or "expected" levels on multiple measures of reading achievement. Three of the five variables we tested were statistically significant at the "optimal" level rating when compared to the other two ratings with medium to large effect sizes for each in grade one. Our findings were very similar in grade three with all but one variable for the "optimal" level showing statistical significance when compared to the other two levels, again with medium to large effect sizes for most of the variables. Clearly, the more a teacher implements an innovation such as the one under scrutiny, the higher the reading achievement of her students.

Similarly, implementation levels appear to make a difference in teacher practices as well. While the traditional comparison of groups demonstrated statistically significant differences in teachers' attitudes and values, there were marked differences in practices when looking at the implementation levels of the treatment teachers. Teachers who simply implemented at an "expected" level showed statistically significant differences on several survey measures (including self-efficacy; attitude toward other teachers; use and valuing of core textbooks; as well as their perception of the language abilities of their students) when compared to other teachers.

Teachers who implemented at an "optimal" level showed statistically significant differences on several measures over those teachers who implemented at "expected" levels. Teachers who participated in the intervention improved their perception of their students, including their attitudes toward gender (expectations of gender around success). These same teachers also changed their attitude toward their abilities in English. Similarly, these teachers recognized that the readability of the books in their classroom was more aligned with the reading abilities of their students—presumably, this meant they were using the books with their students.

These teachers showed a greater disposition toward active teaching. Finally, these teachers reported more interactions with their head teachers and their primary education advisors than did teachers in the "expected" group.

Our cost effectiveness ratios (using projected data) associated with student outcomes when nested within teachers' implementation levels were insightful, when compared to the cost effectiveness ratios of business-as-usual. Ouranalysis indicted that "expected" implementation of the program by classrooms teachers would yield an intervention thatis nearly six times less the cost (than traditional instruction) for the improvement in overall English performance by grade two students.

Similarly, our analysis indicated that "optimal" implementation of the program would yield an intervention that is nearly three times less the cost (of traditional instruction) for the improvement in overall Chichewa performance by grade one students. Likewise, our analysis indicted that "optimal" implementation of the program by classrooms teachers would yield an intervention that is nearly 64 times less the cost (than traditional instruction) for the improvement in overall Chichewa performance by grade three students. Finally, our analysis indicted that "optimal" implementation of the program by classroom teachers would yield an intervention that is nearly 155 times less the cost (than traditional instruction) for the improvement in overall English performance by grade three students.

Conclusion: The Easy the Difficultand, the Ambitious

From a traditional perspective, our intervention had several relatively strong effects on teaching practices. These centered on the books. Teachers reported that they were displaying the books for students to use and that they found the materials accessible for their students. They also reported a valuing of their existing core textbooks and that they used those textbooks more than did teachers in the control group. These are important findings as they represent that the teachers were actually making the books available for their students, which is often not the case in book floods. Books that come into schools in countries like Malawi are often stored in the principal's office behind a locked cabinet (Hunt, 2007) and not used with students. Additionally, our findings may indicate that with the flooding of additional books in classrooms, teachers may have become more cognizant and valuing of the textbooks that had been provided to them in years past. Finally, the intervention demonstrated to teachers how to schedule their week so that there was time for both their existing program and the intervention.

This is important as teachers often face bandwagons of innovations and those same teachers believe that the innovations must usurp any existing practices they have in place (Jampole, 2009). It was perhaps due to the linking of their existing core textbooks with the complementary books that treatment teachers reported using the materials in ways that were statistically and significantly different than control teachers.

Perhaps one explanation of these findings is that some components of innovations such as the one under examination are simply easier to implement. Drawing from other studies that have attempted to implement innovations with teachers (Rubin, Sutterby, & Sailors, 2010), our findings would indicate that the books were perhaps the most straightforward component of the program. We saw improvements within teachers' practices and attitudes and across the two groups. Most likely, this was due to the fact that books were such a pervasive part of the intervention: Teachers received them—lots of them! In making the books available for students, teachers were more willing (or able) to implement this component of the intervention. Perhaps the fact that thebooks were available to teachers and that making them available to students was uncomplicated was yet another reason teachers implemented them so readily. Finally, much focus was placed on using books during the monitoring visits paid to the participating teachers as part of this intervention.

The findings discussed in this study suggest the importance of implementation levels. Presumably, most interventions in education are much focused on "what may work." Our findings may indicate that "what may work" may be misleading. Rather, a focus on "what may work, for whom" and "under what conditions" may be a more interesting question.

References

Alidou, H., Boly, A., Brock-Utne, B., Diallo, Y. S., Heugh, K., & Wolff, H. E. (2006). Optimizing Learning and Education in Africa—the Language Factor. Paris, France: Association for the Development of Education in Africa (ADEA).

Buckler, A. (2011). Reconsidering the evidence base, considering the rural: Aiming for a better understanding of the education and training needs of Sub-Saharan African teachers, International Journal of Educational Development, 31, 244-250.

Camilli, G., Vargas, S., Ryan, S., & Barnett, S. (2010). Meta-Analysis of the Effects of Early Education Interventions on Cognitive and Social Development. Teachers College Record 112, 579-620.

Chilora, H. G. (2007). Use of a learner's language of play as a medium of instruction and as the language for initial literacy development in lower primary school classes in Malawi: A report on the Literacy Across the curriculum longitudinal study, 2004 – 2006. Domasi: Malawi Institute of Education.

Cohen, J. (1988). Statistical power for the behavioral sciences. New York: Erlbaum.

Craig, H. J., Kraft, R. J., & du Plessis, J. (1998). Teacher development: Making an impact. Washington, DC: Academy for Educational Development.

Davidson, M., & Hobbs, J. (2012). Delivering reading intervention to the poorest children: The case of Liberia and EGRA-Plus, a primary grade reading assessment and intervention. International Journal for Educational Development.
http://dx.doi.org/10.1016/ j.ijedudev.2012.09.005.

Department of Education Planning, Ministry of Education Science and Technology (DoE) (2011). Education statistics 2011. Lilongwe, Malawi: Author.

Deussen, T., Coskie, T., Robinson, L., & Autio, E. (2007). "Coach" can mean many things: five categories of literacy coaches in Reading First (Issues & Answers Report, REL 2007-No. 005). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest.

Elley, W.B. (2000). The potential of book floods for raising literacy levels. International Journal of Educational Research, 46, 233-255.

Farrell, J. P, and Heyneman, S. P. (1989). Textbooks in the Developing World: Economic and Educational Choices. Washington, DC: The World Bank.

Fleiss, J. (1981). Statistical methods for rates and proportions (2nd ed.). New York: Wiley.

Hall, G. E., & Hord, S. M. (1987). Change in schools: Facilitating the process. Albany: State University of New York Press.

Hoffman, J. V., Sailors, M. W., Makalela, L., & Matthee, B. (2009). Language Policy and Literacy Instruction: The view from South Africa to South Texas. In J. Hoffman & Y. Goodman. Changing Literacies for Changing Times (pp. 233-248). Newark: International Reading Association.

Hummel-Rossi, B., & Ashdown, J. (2002). The state of cost benefit and cost-effectiveness analyses in education, Review of Educational Research 72, 1-30.

Hunt, G. (2007). Failure to Thrive? The Community Literacy Strand of the Additive Bilingual Project at an Eastern Cape Community School, South Africa. Journal of Research in Reading, 30, 80-96.

Jampole, E. (2009). Traditions, storytelling and crossroads. In F. Falk-Ross, S. Szabo, M. B. Sampson, & M. M. Foote (Eds.). Literacy issues during changing times: A call to action. Commerce, TX: Texas A&M University-Commerce.

Karoly, L. A., Kilburn, M. R., Bigelow, J. H., Caulkins, J. P., Cannon, J. S., & Chiesa, J. R. (2001). Assessing Costs and Benefits of Early Childhood Interventions: Overview and Application to the Starting Early Starting Smart Program. Executive Summary. Seattle, WA: Casey Family Programs; Santa Monica: RAND.

King, J. (1997). Cost analysis in education: Paradox and possibility. Educational Evaluation and Policy Analysis, 19, 309-317.

Lee, V., & Zuze, T. (2011). School resources and academic performance. Comparative Education Review, 55, 369-397.

Levin, H. (1998). Cost-Effectiveness and educational policy. Educational Evaluation and Policy Analysis, 10, 51-69.

Levin, H., & McEwan, P. (2001). Cost effectiveness analysis: Methods and applications (2nd edition). New York Sage Publications.

Macro International. (2008). Malawi Education Assessment Report. Calverton, Maryland, USA: Author.

Makalela, L. (2005). We Speak Eleven Tongues: Reconstructing Multilingualism in South Africa. In B. Brock-Utne and R. K. Hopson. (Eds.). Languages of Instruction for African Emancipation: Focus on Postcolonial Contexts and Considerations (pp.147-74). Cape Town: CASAS.

McGill- Franzen, A., Allington, R.L., Yokoi, L., & Brooks, G. (1999). Putting books in the classroom seems necessary but not sufficient. Journal ofEducational Research 93, 67-74.

Montagnes, I. (2000). Textbooks and learning materials 1990-1999: A global survey. Paris, France: UNESCO.

Nannyonjo, H. (2007). Education inputs in Uganda: An analysis of factors influencing learning achievement in grade six. Washington, DC: The World Bank.

Pianta, R., Mashburn, A., Downer, J., Hamre, B., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher–child interactions in pre-kindergarten classrooms. Early Childhood Research Quarterly, 23(4), 431–451.

Rubin, R., Sutterby, J., & Sailors, M. (2010). The easy, the difficult and the almost impossible. In J. Cassidy, S. Garrett & M. Sailors (Eds.). Literacy Coaching: Research and Practice (pp. 95-112). Corpus Christi, TX: Texas A & M University Corpus Christi.

Sailors, M., &Hoffman, J. V. (2011). Establishing a Print-rich Classroom and School Environment. Best Practices of Literacy Leaders in Schools (pp. 184-205). Guilford Press.

Sailors, M., Hoffman, J. V., Pearson, P. D., McClung, N., Shin, J., Phiri, L., & Saka, T. (2014). Supporting change in literacy instruction in Malawi. Reading Research Quarterly, 49(2), 209-231.

Sailors, M., Hoffman, J.V., Kaambankadzanja, D., Chilora, H., Mapondera, A., & Betha, F. (2012). Complementary reading material development: Accomplishments, successes, and challenges of the Read Malawi programme. Pula: Botswana Journal of African Studies, 26, 171–186.

Seng, T. (2005). The impact of ICT on learning: A review of research. International Educational Journal 6(5), 635-650.

Sheridan, S. M., Edwards, C. P., Marvin, C. A., & Knoche, L. L. (2009). Professional development in early childhood programs: Process issues and research needs. Early Education and Development, 20(3), 377–401.

Thompson, B. (1999). If Statistical Tests Are Broken/Misused, What Practices Should Supplement or Replace Them? Theory & Psychology 92, 165-181.

World Bank. (2010). The education system in Malawi. Washington, DC: Author.