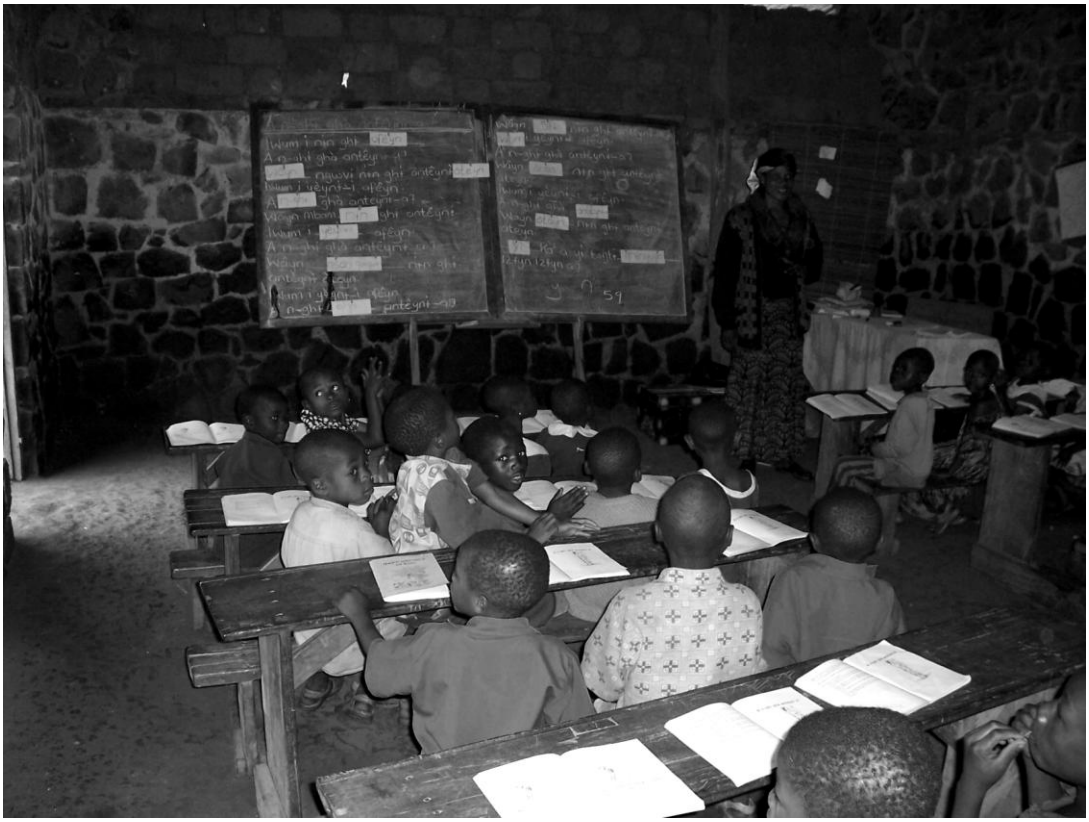


The Kom (Cameroon) Multilingual Education Project (KEP)

Status Report – July 2009

Stephen L. Walter (GIAL and SIL)
Kristine R. Trammell (SIL Cameroon)

Bamenda, North West Province, Cameroon

Table of Contents

The Kom (Cameroon) Multilingual Education Project (KEP)

Executive Summary

The KEP project is an educational innovation being piloted in the Boyo Division of North West Province in Cameroon. The key element of the innovation is the use of the local language as the primary language of instruction in Classes 1 to 3. The innovation also includes supporting curriculum materials in the local language and the training of teachers to teach via the local language.

For purposes of research and documentation, the KEP innovation has been implemented in 12 schools—both government and private—scattered around the Division. Twelve English-medium schools have been selected as comparison schools based on similarity of size, location, and previous educational history. All of this has been done in consultation with the educational authorities of the area.

The KEP program or experiment has now been in operation for two years. At the end of each year, standardized testing has been done in both KEP and comparison schools to document the impact of the KEP innovation. The primary metric of comparison is assessment of learning outcomes in Reading, Math, and Oral English, but data is also being collected in order to measure impact measures of participation such as enrollment, dropouts, and rates of repetition. The intent of the research component of the project is to track participating students through the full six years of basic education to better assess the extent to which the innovation impacts the long term educational potential of the students who were a part of the innovation.

Testing done after the first year (2008) of the innovation revealed that children in the KEP program outscored their counterparts in the Standard schools by a large margin—52.32 percent to 16.12 percent on a basic test of the skills specified as learning outcomes for Class 1. Testing of Class 1 students at the end of 2009 school year showed rather similar results—47.09 percent versus 20.93 percent. The achievement advantage was seen across the entire curriculum even including development of Oral English though the advantage was less in this subject than in the other two.

The testing of Class 2 students (2009 only) reveals that the effectiveness of the innovation extends to these students as well though the differential was somewhat reduced from Class 1—54.12 percent to 33.76 percent. To the surprise of many the Class 2 students have continued to show more progress in learning English than the children in the English-medium schools.

In addition to measuring overall program achievement, the research work has examined the results of the innovation on a number of variables of typical interest to educators and policy makers—the achievement and trajectory of individuals schools, the variables of age, gender, differential achievement in the various subjects, and the performance of children from ethnic backgrounds other than Kom. The following bullet points summarize findings to date on these factors.

- The achievement of individual schools varies immensely in both KEP and Standard schools by as much as 265 percent with the greatest variation seen in the Standard schools.
- In general, the weakest of the experimental schools performed at about the same level as the best of the control schools.
- The **best** of the experimental schools outperformed the best of the Standard schools by 93 percent in Class 1 and by 32 percent in Class 2.

- The **worst** of the experimental schools outperformed the worst of the Standard schools by 207 percent in Class 1 and by 91 percent in Class 2.
- The presence of under-age children in both programs is substantial ranging from 27 in the Standard program to 39.3 percent in the KEP program.
- Under-age children, as a group, consistently underperform in direct proportion to their age (i.e., the youngest children have the lowest scores) in both programs.
- Overage children generally outperform standard age children in direct proportion to their age. Exceptions are those who have evident learning disabilities.
- Gender appears to be a slight factor in participation and a non-factor in achievement in both programs. Female children generally outperform their male peers by a slight margin which is NOT statistically significant.
- Class 1 children in the KEP schools (learning to read in Kom) show a huge advantage in reading achievement (about 500 percent).
- Class 1 children in the KEP schools show a similar but slightly lower advantage in math achievement.
- Class 2 children in the KEP schools continue to show a large advantage in both reading and math though less than that seen in Class 1—97 percent in reading and 116 percent in math.
- Class 1 children in the KEP schools show an advantage in learning oral English though the advantage is much less than that seen in Reading and Math.
- Class 2 children in the KEP schools continue to show an advantage in learning oral English but the advantage is less than that seen in Class 1.

The impact of the innovation on the variables of promotion, repetition, and abandonment are still quite 'soft' due to a problem encountered in consistently identifying students by name. Tentatively (we urge extreme caution in how these findings are treated), the KEP program increased the success rate of children moving on to and completing Class 2 by approximately 48 percent. The repetition rate was reduced by 55 percent and the rate of abandonment after Class 1 by 45 percent.

While the KEP program has led to dramatic improvement in educational outcomes in all areas measured, in our judgment, the KEP schools are still not as strong educationally as they should be. In general, one expects schools providing good instruction to score in the 65-80 percent range on criterion-referenced tests of learning (and several of the KEP schools achieved at this level). Overall, however, the mean performance level of the schools in the KEP program in 2009 was 47.1 percent and 54.12 percent in Classes 1 and 2 respectively. The primary factors appeared to be instructional time lost (maternity leave, sickness, transfers) and low motivation on the part of some teachers.

The KEP program is now entering its third full year of implementation. Given that this is the last year for which full external funding is being provided, the coming challenge will be that of working out whether and to what extent the innovation can be either generalized to other schools or maintained by the existing educational infrastructure of the region. Discussions and negotiations on this subject will be held during the coming year in an effort to find solutions to these questions.

The basic report is contained in the first 30 pages. In addition, several Appendices have been attached to the end of the report. For the most part, these address more technical questions so will not be of interest to everyone. Not all of these have been completed as of the first version of the report (July 23, 2009). The final version (late July) will contain the completed Appendices.

Introduction

The Kom language is spoken by the Kom people and is the primary language spoken in the Boyo Division of North West Province in Cameroon. The Boyo Division is very mountainous and less than 50 kilometers from one end to the other along the paved road which transects the area. The result is that most available space which is even somewhat flat is either used for living or for farming.

The Kom people, numbering an estimated 250,000, are primarily horticultural with some small businesses located in the larger settlements. In general, the Kom would be considered a rather aggressive people so pursue education and business opportunities both inside and outside the Kom area.

There are approximately 150 primary schools scattered throughout the area. The majority of these are government schools but private schools—especially Catholic schools—are still prominent. Several secondary schools are found in the area along with a teacher training institution in the largest town of the area—Fundong.

The primary schools in the Kom area are typical of rural Africa—menially constructed and poorly equipped with no running water, no electricity, no latrines, and often no windows, doors, lights, or books). Teachers are not well trained and often poorly paid. Supervision is minimal and technical support almost non-existent. The instructional year (and day) is shortened by frequent disruptions, lack of discipline, teacher absences, many holidays, and all sorts of unplanned "activities" like marching, singing contests, gardening, policing the school grounds, etc. Effective education under such circumstances is difficult at best. With no school books, overpopulated classrooms, and the teacher speaking a language the children do not understand, it is reasonable to ask whether learning is possible.

The Kom Education Project (KEP) is an experimental program designed to test whether and to what extent beginning instruction in the first language enhances educational outcomes among participants. The KEP project is one component of a larger research project being managed by SIL Asia having the same purpose. Research in the West (Thomas and Collier, 1997; Linholm-Leary, 2004; Cummins, 2005) has provided convincing evidence that early instruction in a child's first language (in the case of those living in a context where their first language is not the majority language) allows such children to "catch up" with their peers speaking the majority language by the beginning of middle school. Conversely, when minority language children are educated only via the majority language—English in the case of the US—these children finish their education far behind all others—as low as the 11th percentile.

It is not safe to assume, however, that program effects observed in the West will hold when such programs are transferred to developing country contexts. The KEP project is one component of a larger research project being managed by SIL Asia designed to test just this question. Specifically, *will the educational outcomes of children entering rural schools in a developing-country context speaking only a local language be significantly enhanced IF these children received significant early instruction in the local language before moving on to L2 medium instruction?*

A Brief Recent History of MLE in Kom

The PROPELCA project, a joint initiative of the linguistics department of the University of Yaounde and SIL Cameroon sponsored an early experiment in mother tongue education in the Kom area beginning in the late ‘80s. This program eventually extended to some 30 primary schools in the Kom area. The PROPELCA initiative primarily provided training for teachers and some basic instructional material. With no on-the-ground supervisory or

administrative structure, actual use of Kom as a language of instruction in the schools was somewhat erratic. Some teachers were enthusiastic in their use of Kom in the classroom while others were quite indifferent. Furthermore, it was frequently the case that Kom might be used as the medium of instruction in Class 1, not in Class 2, and then perhaps again in Class 3 depending on the movement of teachers between schools and their level of interest in the project.

In 2001, a major national overhaul of the basic curriculum was implemented in the Kom area. Among other things, this curriculum overhaul added several new subjects to the curriculum including French. The addition of this new content meant that Kom-medium instruction was squeezed out.

In 2006, the authors began exploring the potential for a re-introduction of mother tongue based instruction using the existing curriculum. After an extensive period of discussion and negotiation, educational officials agreed to the introduction of an experimental project in the area in a subset of primary schools in the area. For experimental purposes, it was agreed that 12 schools would be set up as experimental schools using a Kom-based instructional program and 12 matching schools would be identified as control schools. Nothing changed in the control schools other than the fact that formal year-end testing was done in these schools to measure educational progress using the standard curriculum.

Curriculum development for the program commenced in early 2007. Some instructional material already existed as a result of the earlier PROPELCA project. Some of this material was used as is and some was adapted for the new project. A teacher training workshop was organized during the summer of 2007 to prepare teachers to use the new curriculum. Instruction began in the Fall[1] of 2007. KEP project staff, with the permission and encouragement of the local education inspectors, made fairly regular supervisory visits to the experimental schools to monitor their operation and to deal with problems. The first round of testing was carried out in the Spring of 2008. Instruction was extended to Class 2 in the Fall of 2008 with a second round of testing in the Spring of 2009. Instruction in Class 3 in Kom is scheduled for launch in 2009 and a third round of testing will be done in the Spring of 2010.

Structure of the Project.

The KEP project is, at once, a demonstration of feasibility and a research project. The project has its own technical administrator but is set within the existing framework of educational administration already in place in the Kom area. While the KEP administrator is responsible for curriculum and instruction in the experimental classes, teacher selection and payment is handled through the infrastructures already in place (whether the Ministry of Education or the Private School administrations operating in the area).

As noted above, there are twelve KEP (experimental) schools and twelve control[2] schools. The experimental schools are scattered among the three sub-divisions that make up the Boyo Division. School enrollment is open and is managed by the head masters of the respective schools. The KEP project has insisted that children not be enrolled in Classes 2 or 3 who did not complete earlier Classes in the same program but this has not always been observed. The project also specified class size limits but these have not always been observed either.

[1] For those not familiar with these terms, 'Fall' refers to the latter months of the year, and 'Spring' refers to the early months of the year.

[2] Reports on this project have used several difference terms o refer to the pre-existing English-medium schools—the standard strategy for providing basic education in North West Cameroon. These include: English schools, English-medium schools, control schools, traditional schools, or Standard schools. In this report we will use primarily *Standard schools* to refer to the 12 schools which serve as the basis of comparison for the Kom-medium schools. The other terms may be used occasionally but should be understood as referring to the same set of 12 schools. The participating schools are listed in Appendix D.

The KEP project reports to all levels of educational administration in the Province and makes reports (mostly oral) to these on a somewhat regular basis.

Methodology

Test instruments

The means of comparison between the two models is performance on standardized tests administered to all students in both experimental and control schools. The content of the tests is based on the learning outcomes specified in the national curriculum for each class. Since the two curricula are not precisely the same, the test was based on outcomes shared by the two programs and excluded those unique to one program or the other. The test instrument did not seek to be comprehensive with respect to all learning outcomes as it was not meant to be a full-scale assessment of instructional mastery. Rather, it tested mastery of a sample of knowledge and skills in the major subject areas of reading, math, and English-language learning.

The two tests had the following structure:

Table 1. Structure and content of the year-end assessments administered in 2009.

Class 1 Test	Class 2 Test
Reading	Reading
Word recognition – 5 items	Basic grammar – 6 items
Grammar – 4 items	Reading comprehension -4 items
Reading comprehension – 4 items	Math
Math	Addition – 3 items
Counting – 2 items	Subtraction – 3 items
Place value – 2 items	Problem recognition – 3 items
Addition – 3 items	English language learning
Subtraction – 3 items	Social interaction – 3 items
English language learning	Factual response – 4 items
Social interaction – 3 items	Free response – 3 items
Factual response – 4 items	
Free response – 3 items	

The 2009 version of the test was different from the 2008 version in several regards. First, we eliminated one entire section which was not, in our opinion, well-constructed and thus a questionable measure of children's knowledge—the section on science. Second, we shortened each of the sections to reduce the time of test administration and because we felt the shortened version would still be an adequate measure of assessment. Third, we added some additional practice items to help children understand how to take the test. The end result was that the 2008 Class test contained 41 points, while the 2009 version of the same test contained 32.75[3] points—about a 20 percent reduction in length.

Language of Assessment

All children were tested in the language in which they were being instructed. Obviously, this means that the reading sections of the test looked different in that one test contained questions about English and a sample text in reading comprehension in English while the counterparts in the other test were in the Kom language. The tests sought to establish

[3] The odd number of points is due to the fact that the oral English section was weighted slightly due to the fact that 3 of the items were evaluated on a multipoint scale.

comparability by taking texts of equal length from similar lessons in the yearly curriculum and by asking similar types of questions. For example, in the Class 2 test of reading comprehension, three questions were factual and one question asked the test taker to identify the subject of the story ("What was the story about?"). The other two sections of the test—math and Oral English—were identical for both control and experimental programs.

Test administration

Before the final test was given, a practice test was circulated to all schools in both programs to help both schools and students prepare for the assessment. (Informally, it is reported that some teachers made good use of these practice tests while others ignored them all together.) In all cases, the tests were administered by a small group (a team of 6 people in all) of examiners selected and trained for this purpose. Arrangements for testing were made through the school headmaster or head teacher. Teachers and usually either the head teacher or headmaster were present during testing. A general explanation of the test and test procedure was given to all children in both programs in Kom to maximize their understanding of the process. This introduction included an explanation of the multiple choice format (which was used only for the reading section). Each written section of the test included additional practice items to further assist the test taker in understanding how to take the test.

For the actual taking of the test, each item in the reading section was read twice orally by the test examiner and then the test takers were instructed to select the best answer. The oral reading of items did not apply to the section on reading comprehension which children had to read for themselves along with the comprehension questions.

The items in the math section were simple computation so children were given basic instructions on how to proceed and left on their own to complete the problems. The one exception in the math section was the word problems in Class 2. These were read to the children with the children then given the opportunity to select the correct mathematical expression of the word problem from the three options available in a multiple choice format.

The reading and math sections of the testing were done corporately in the classroom. The oral English testing was done one-on-one with each student tested. The examiner read or expressed a fixed question ("What is your name?") in English and gave the child an opportunity to respond appropriately. Responses were evaluated as acceptable or not with no partial credit. For the free response items, children were shown a picture or an illustration and asked to say anything they could about the content of the item. Credit was based on a 5 point scale depending on the completeness and adequacy of response.

The tests were scored by the examiners with this scoring reviewed by project personnel and corrected when necessary. Data entry was done by project personnel using EXCEL and analysis was done using the commercial software program MINITAB.

The total number of children tested in Class 1 was 687 (24 schools) and in Class 2 624 (24 schools).

Instrument reliability.

The notion of *reliability* is used by researchers to describe several aspects of an instrument's performance in a testing situation. One, does the instrument measure what it is designed to measure? Two, are the individual items in the instrument well constructed so that they measure accurately what they are supposed to measure? Three, was the instrument administered in such a way that it elicited an adequate and accurate measure of what it was designed to measure? In recent years increasingly more sophisticated and rigorous methods have been developed to measure instrument reliability. In this case, we applied a very common and widely-used measure called Cronbach's alpha. Cronbach's alpha is a statistic which depends upon the relationship between response on individual items to overall test

performance. The logic is that those who are more capable would tend to answer more items correctly while those who are less able would tend to not be able to answer correctly the more difficult items. Variation in ability to answer individual items should have a consistent relationship with the overall variation in test performance.

We used Cronbach's alpha to estimate the reliability of the instruments used for this research study. For standardized educational assessments, one normally expects to find values of alpha between .70 and . 90 with a higher value indicating a higher level of instrument reliability. The instrument for Class 1 had an overall Cronbach's alpha of .89—quite high for this type of an instrument. Divided into parts by language, the instrument(s) had an alpha of .87 for the Kom version and .81 for the English version of the test. We were very satisfied with these measures of reliability.

Cronbach's alpha for the Class 2 test overall was .82. If we treat the test as two separate tests by language we get a Cronbach's alpha of .79 for the Kom version of the test administered in the KEP schools and .76 for the English version of the test administered in the Standard schools. All of these are very acceptable measures of test reliability indicating the general validity of the test(s) although slightly lower than those of the Class 1 instrument. There are a variety of possible reasons for these slightly reduced measures of reliability but we lack the data to investigate these at this time.

Results - General

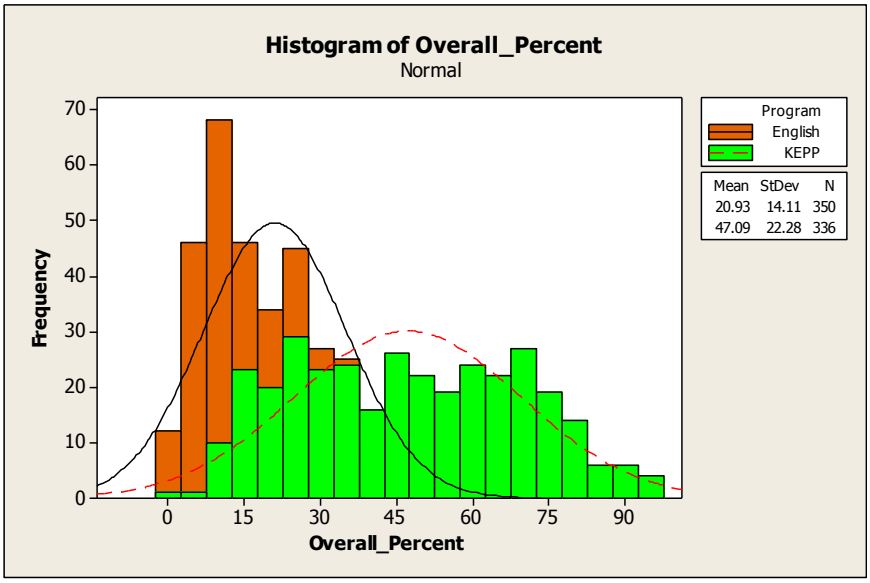
The benefits or potential benefits of an educational innovation can be assessed in a number of ways. At the macro level one examines education indicators such as gross national enrollment and efficiency rates and development indicators like social returns to investment, national manpower development, international competitiveness, etc. At the micro level, the indicators tend to be somewhat more concrete and localized. These may be cognitive, "Are children learning more and remembering it longer?"; participative, "Are more children going to school, staying in school, and graduating?"; or sociopolitical, "Is the community happy with the impact of the innovation on their lives, their values, and their families?"

Because this innovation—multilingual education using the Kom language as a medium of instruction—is just in its second full year of operation, the report will focus primarily on cognitive outcomes (learning) and secondarily on efficiency outcomes (repetition, dropout, and promotion rates). As we get deeper into the innovation, more information will be available to examine other impacts.

Overall Results – Class 1

First, we will look at some general descriptive statistics detailing the performance of children in the two programs in Class 1. Figure 1 is an overlaid histogram showing the distribution of scores in Class 1 for the two programs in 2009.

Figure 1. Distribution of Class 1 students by program on the year end assessment.



The orange bars in the histogram (partly hidden) show the distribution of students in the Standard program. Notice that the performance band having the greatest number of students is that of 10 to 15 percent. Nearly 70 students—almost a quarter of the total—scored in this range. Notice also that this distribution approaches zero at about 65 percent with a peak or mean at about 20 percent.

The green bars reflect the distribution of performance of the Class 1 students in the KEP program. The distribution is pretty normally distributed at the extremes though rather flat in the middle. The distribution centers at approximately 50 percent.

Table 2 contains some of the descriptive statistics summarizing the results of the testing for Class 1.

Table 2. General statistics comparing Class 1 students from the two programs in 2009.

	Standard Program	**KEP Program**
Number of participating students	350	336
Highest score	66.8 percent	97 percent
Lowest score	0 percent	2 percent
Mean and Standard Deviation	20.93; SD = 14.11	47.09; SD = 22.28
First Quartile	9.9	27.7
Median score (Second Quartile)	17.6	47.14
Third Quartile	28.7	65.6
Number scoring above 75 percent	0	42
Number scoring above 50 percent	16	153
Number scoring above 25 percent	123	269
Number scoring **below** 10 percent	95	7

Some of the contrasts noted in Table 2 are dramatic! For example not a single student from a Standard school scored above 75% on the year-end test while 42 children from the KEP program scored above this level. Similarly, we note that just 16 out of 350 children scored above 50 percent in the Standard schools while 153—almost half of the KEP children—scored above this level. These are certainly startling comparisons.

Tables 3 and 4 add both depth and statistical information to the comparisons for Class 1. Looking at scores in the 'Percentage' column we observe that the mean score for all children

in KEP Class 1 was 47.1 percent while the mean score of those in the Standard schools was 20.9. By this broadest of measures, the children in the KEP program had a 125 percent advantage[4] over the children in the Standard program. The large value of T highlights the size of the difference—a value of 2.5 is usually sufficient to indicate that one sample is statistically different from a second sample. A T value of 18.29 is not often seen in education or social science research thus emphasizing the enormity of the difference in performance on this test by the two groups of students.

Table 3. Overall performance of KEP and Standard schools on the year-end test, Class 1, 2009.

	Total Test (32.75 total points)		
Program	**Raw Score**	**Percentage**	**Statistics**
Kom-medium[5]	15.42	47.1	T = -18.29; P = 0.000
English-medium	6.88	20.9	

Table 4 and Figure 2 provide a more detailed comparison of Class 1 students by program and by content areas. Children in the experimental schools demonstrate a strong advantage in reading and mathematics and a lesser advantage in Oral English (55 to 42.8 percent). The relative advantage in reading and math is over 200 percent while that of oral English is just 28.5 percent.

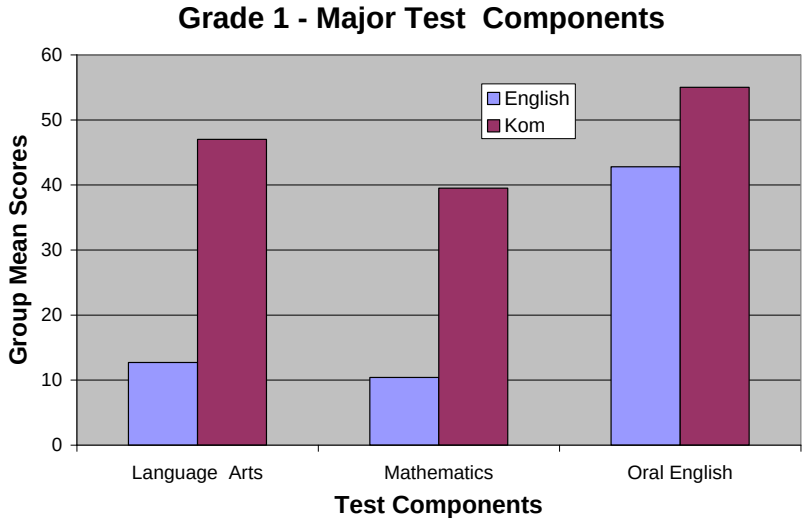
Table 4. Comparison of students by program and content area in Class 1, 2009.

	Reading (13 points)		
Program	**Raw Score**	**Percentage**	**Statistics**
Kom-medium	6.11	47.0	T = -20.09; P < 0.000
English-medium	1.65	12.7	
	Mathematics (10 points)		
Program	**Raw Score**	**Percentage**	**Statistics**
Kom-medium	3.95	39.5	T = 14.10; P < 0.000
English-medium	1.04	10.4	
	Oral English (9.75 points)		
Program	**Raw Score**	**Percentage**	**Statistics**
Kom-medium	5.36	55.0	T = -7.16; P < 0.000
English-medium	4.17	42.8	

[4] In this paper we use several mechanisms for describing the difference between the impacts of the two programs. One of these is *simple difference*. Simple difference, as the term suggests, refers to the mathematical difference in measures between two programs or program units. For example, if one person scores 55 percent on a test and a classmate scores 65 percent, the simple difference is 10 percent. We may occasionally refer to this difference as *absolute advantage*. Another way of comparing two scores is in terms of *relative advantage* or *percentage of gain*. This term is used as a measure of improvement relative to the original (usually lowest) score. Percentage gain is computed by subtracting the original or lowest score from the second or highest score and dividing the difference by the original or lowest score. The use of the T or F statistic is a formal statistical means for making a similar comparison but we will not try to explain how these statistics are computed and what they mean in this paper.

[5] Throughout the presentation, the term, 'Kom-medium' refers to children in the experimental schools (KEP program) while 'English-medium' refers to those in the traditional schools in which English is used as the primary or only language of instruction.

Figure 2. Comparative performance on the major components of the Class 1 assessment.



The advantage demonstrated in oral English on the part of the KEP program is more surprising given that children in the Standard classes are receiving ALL of their instruction in English and thus hearing English being spoken at least four hours a day. Children in the KEP program receive instruction in oral English for just 5 hours a week yet appear to be getting a stronger foundation in English—at least oral English—via this modality. This finding (similar to that reported in 2008) appears to run contrary to the common popular perception that the best strategy for ensuring development in English is to immerse children in English-medium classrooms.

Table 5 provides a more detailed comparison of performance by skill areas. This information is being included in the interest of completeness and for administrators who want a closer look at the learning implications of the innovation being reported here.

Table 5. Comparative performance by test subsections, KEP and Standard, Class 1, 2009.

	Standard	KEP	Performance Advantage	Statistics
Language Arts				
Word Recognition	14.2	47.6	235 percent	T = 13.35; p < 0.000
Grammar	15.9	46.0	189 percent	T = 15.88; p < 0.000
Reading Comprehension	6.9	47.3	586 percent	T = 21.15; p < 0.000
Mathematics				
Counting	15.2	40.6	167 percent	T = 9.95; p < 0.000
Place Value	6.2	33.3	437 percent	T = 10.91; p < 0.000
Addition	10.9	46.7	328 percent	T = 13.81; p < 0.000
Subtraction	9.5	35.7	276 percent	T = 10.70; p < 0.000
Oral English				
Social Interaction	79.8	82.8	4 percent	T = 1.57; p = .117
Factual Response Items	51.0	79.8	57 percent	T = 10.31; p < 0.000
Free Response Items	18.7	24.5	31 percent	T = 2.78; p = 0.006

Looking through Table 5, we note only one skill set for which the two groups are statistically equivalent—Social Interaction under Oral English. Performance on this item was high for both groups at a level which indicates general overall mastery. It would appear that this section—the most basic in all of language learning—has evidently been well-learned by both groups.

The largest difference or advantage is in reading comprehension where the KEP students demonstrated an overwhelming 586 percent advantage over students in the Standard program. The mean score for the KEP students indicates that they are reading and interacting knowledgeably with a text after just one year of instruction. The very low level for children in the Standard program—just 6.9 percent—indicates this task was totally beyond their capability, a finding consistent with the baseline study done in 2006 and with research in other parts of Africa.

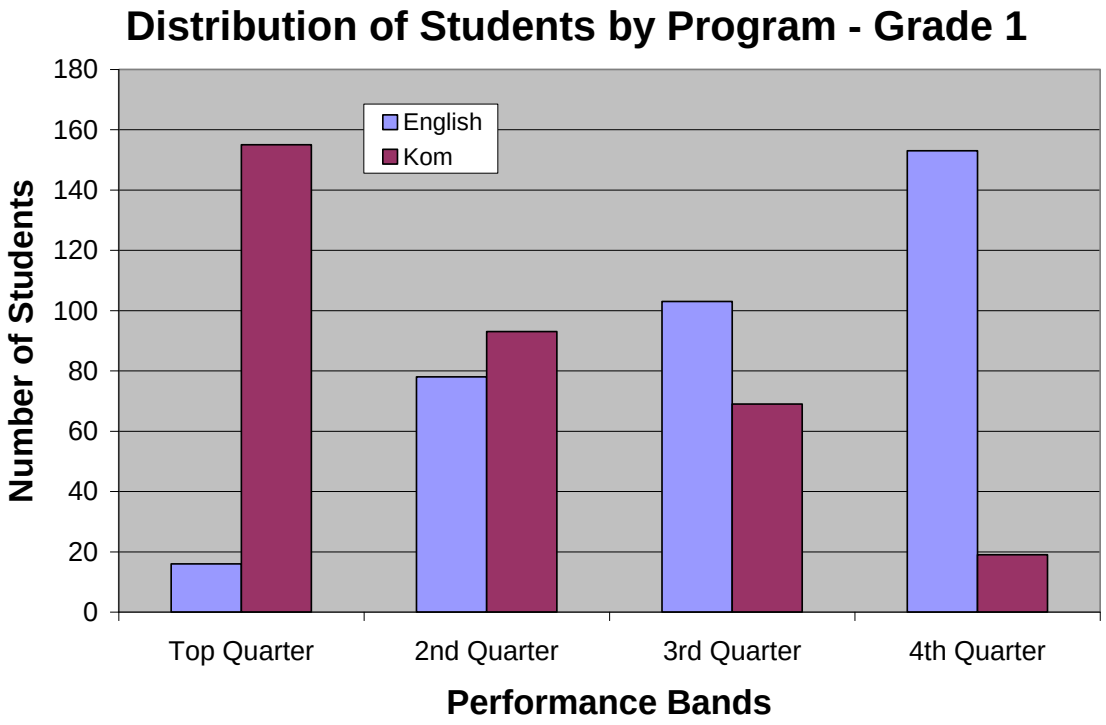
The large differences observed with respect to mathematics are noteworthy as well. Even though most of the items in this section required simple computation with no language involved, the KEP program manifests a large advantage. We interpret this finding to be due to one of two (or maybe both) factors. First, to a greater extent than commonly realized, mastering the facts and processes of basic math cannot be divorced from language in that they have to be explained—they cannot be learned from simple observation as one can learn how to make a basket or to plant a row of beans. Secondly, we suspect that the teachers in the KEP program were able to more efficiently use the available instruction time to teach subjects like math because of their use of the language of the children. This use of a common language of instruction meant that they did not have to spend large blocks of time trying to make a point or to clarify an issue; the children understood them right away and could devote their time to skill-building, not trying to understand what the teacher was saying.

Comparison by student ranks. In Cameroon, as in many other countries, it is common to rank students by their level of achievement from highest to lowest. This type of comparison is easy to do and easy to understand though such rankings lack formal statistical properties. Accordingly, we have included a figure which reflects this ranking though it does not give names and scores. Rather, it divides the population of all Class 1 test takers into four equal groups.[6]

Figure 3 (below) probably highlights as well as any figure can the educational strength of the Kom-medium program. Of the 171 Class 1 students in the top performing 25 percent, only 17 were from Standard schools. Not shown in the figure is the fact that the top-ranked student form the Standard program ranked number 79 out of 686 students tested. At the other end of the performance range, we find almost the exact opposite situation. Of 171 students making up the bottom 25 percent of all Class 1 students, 152 are from Standard schools and 19 are from Kom-medium schools.

[6] This kind of division is similar to the more formal notion of a quartile. Technically, a quartile is the dividing line or value between two of four equal sections (in number) of a population. Quartile values are given in the tables of descriptive statistics. In the case of this figure (and a similar one for Class 2), however, we will be focusing primarily on the groups between the quartiles rather than the quartile points themselves. For this purpose, we will talk about the top quarter, the second quarter, the third quarter, the bottom quarter, etc., referring to the blocks of students which fall into each of these groupings. To be more explicit, the top quarter refers to the 25 percent highest scoring students in a defined population. Conversely, the bottom quarter refers to the 25 percent lowest scoring students in a defined population.

Figure 3. A comparison of the two programs in terms of the rankings of students by performance on the year-end assessment for Class 1.



Clearly, the innovation under review had a dramatic overall impact on the performance of children in Class 1. Another way to characterize the difference is to compare the performance of an average child in the KEP to the range of performance in the Standard program. In technical terms, the performance of an average child in the KEP program (47.09 percent) places that child at the 94th percentile among students in the Standard program. In ordinary terms, this means that the performance of an average child in the KEP program would place that child among the very brightest of all children in the Standard program measured in terms of performance on the year end test.

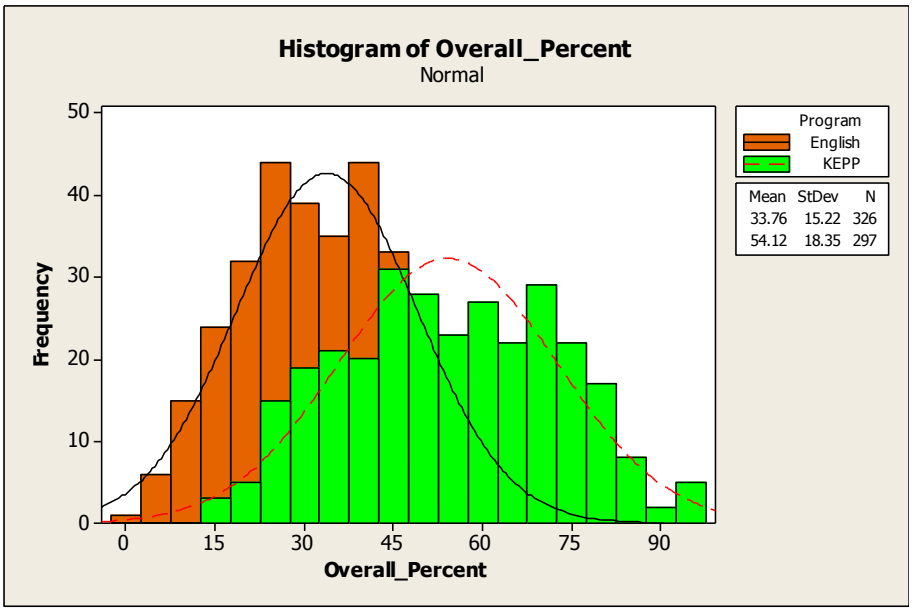
Overall results - Class 2

The Class 2 results being reported here represents those students who were in Class 1 the year before (2008) and who moved on to Class 2 plus any students from any other source who joined those promoted from Class 1 the year before. As will be noted in a subsequent section, the makeup of this group is not as straightforward as we had expected. It appears that there were fewer students who moved on (in both programs) than we expected and more "others" in the group that we are still working to identify. The analysis which follows immediately will treat the Class 2 students as though they had all moved up from Class 1 in 2008. Later, we will look more carefully at the 'real' makeup of Class 2 to better present and interpret the results of the testing which was done.

In the experimental program, instruction continues in Class 2 in Kom, the local language. In the practical outworking of the program, it was decided to have teachers follow cohorts of students from Class 1 through Class 3. This reduced the training load somewhat and provided a greater level of continuity between teachers and students. It should be noted that this practice also occurs in the Standard schools but is not the prevailing model.

As we did with Class 1, we start with some general descriptive statistics describing the performance of children in the two programs in Class 2. Figure 3 is an overlaid histogram showing the distribution of scores in Class 2 for the two programs.

Figure 3. Distribution of all scores in Class 2, 2009.



The distribution of scores of both groups of students is approximately normal[7]. The primary difference is that the mean performance of students from the Standard program is 20 percentage points less than that of the experimental group. We also note that the distribution curve for the experimental group is also a little "flatter" than that of the Standard group meaning that their performance is spread over a broader range than that of the Standard students. This will be reflected in the table of descriptive statistics which follows.

Table 6. Descriptive statistics of students in Class 2, 2009.

	Standard Program	**KEP Program**
Number of participating students	326	297
Highest score	83.0 percent	94.8 percent
Lowest score	0.0 percent	13.0 percent
Mean and Standard Deviation	33.76; SD = 15.22	54.12; SD = 18.35
First Quartile	22.6	40.7
Median score (Second Quartile)	32.8	53.9
Third Quartile	43.2	68.7
Number scoring above 75 percent	3	45
Number scoring above 50 percent	43	171
Number scoring above 25 percent	220	284
Number scoring **below** 10 percent	17	0

As the histogram above suggests, the data in the table show that Class 2 students from the Standard program have narrowed the gap somewhat between themselves and those in Class 2 of the KEP program. The difference between the programs is more marked at the range above

[7] The word 'normal' in this context is a technical term referring to what is more commonly known as a bell curve. A bell curve is symmetrical peaking in the middle and dropping off on both sides. This kind of distribution is very common and often seen when one gets a good measure of some trait or capability.

50 percent. Only 43 children in the Standard program or 13.2 percent scored above 50 percent while 171 or 57.6 percent of those in the KEP program scored above 50 percent. Based on the mean of the two groups, the students in the KEP program show a 60 percent achievement advantage over those in the Standard program.

Table 7 reports the same comparison between control and experimental children in Class 2.

Table 7. Overall comparative performance of experimental and control schools on the year-end test, Class 2, 2009.

	Overall (28.75 points)		
Program	**Raw Score**	**Percentage**	**Statistics**
Kom-medium	15.6	54.1	T = -15.00; P < 0.000
English-medium	9.7	33.7	

The results tend to parallel those of Class 1. The primary observable difference is that the gap between children in the control and experimental programs is smaller being only 60.5 percent rather than 125 percent.

As was done for Class 1, results of the Class 2 test are also presented broken down by major content areas—Reading, Mathematics, and Oral English. Looking at the data in Table 8 we note disparities between the two programs similar to those seen in Class 1 in the two content areas of Reading and Mathematics. However, in the case of Oral English, children in the two programs were quite similar with only an 8.7 percentage point differential between the two. Children in the KEP program are still outpacing those in the English-medium program but not by much.

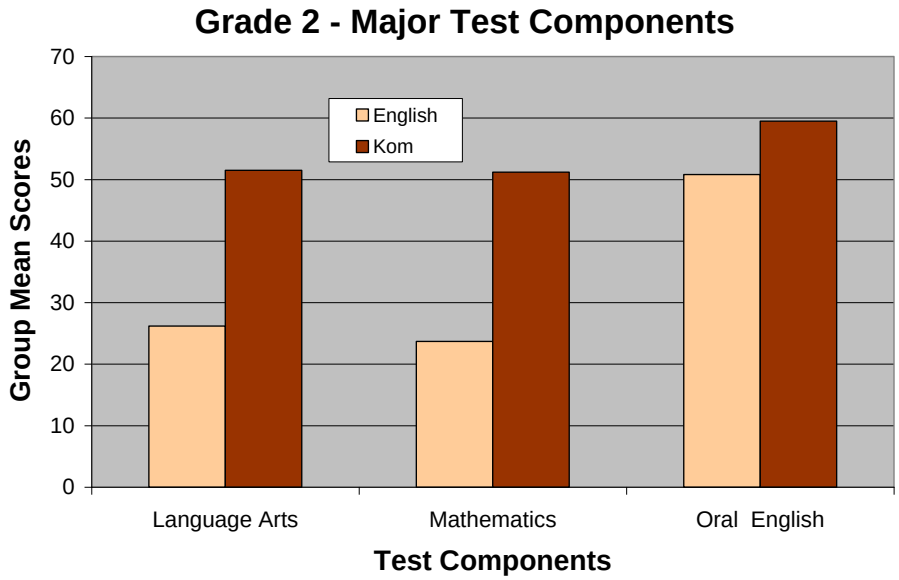
Table 8. Comparison of children in control and experimental schools by content area in Class 2, 2009.

	Reading (10 points)		
Program	**Raw Score**	**Percentage**	**Statistics**
Kom-medium	5.15	51.5	T = -14.76; P < 0.000
English-medium	2.62	26.2	
	Mathematics (9 points)		
Program	**Raw Score**	**Percentage**	**Statistics**
Kom-medium	4.61	51.2	T = -14.13; P < 0.000
English-medium	2.13	23.7	
	Oral English (9.75 points)		
Program	**Raw Score**	**Percentage**	**Statistics**
Kom-medium	5.80	59.5	T = -5.26; P < 0.000
English-medium	4.95	50.8	

More commentary of an analytical or theoretical nature will be added on the interpretation of these data in a later section. For the present, it is probably sufficient to note that the large disparity continues between the two programs in the content areas while it is decreasing in the area of oral English. The trend with respect to oral English suggests that, by the end of Class 3, the English advantage may switch to the English-medium classes. It will be interesting to see whether or not this happens when testing is done at year-end in 2010.

Figure 4 provides a graphic rendition of the data in Table 8. It is easy to see that there is less of a gap between students in the Standard and KEP programs compared to that seen in Class 1. The profile in Class 2 is actually closer than that of Class 1 to what we would expect for mother-tongue instruction.

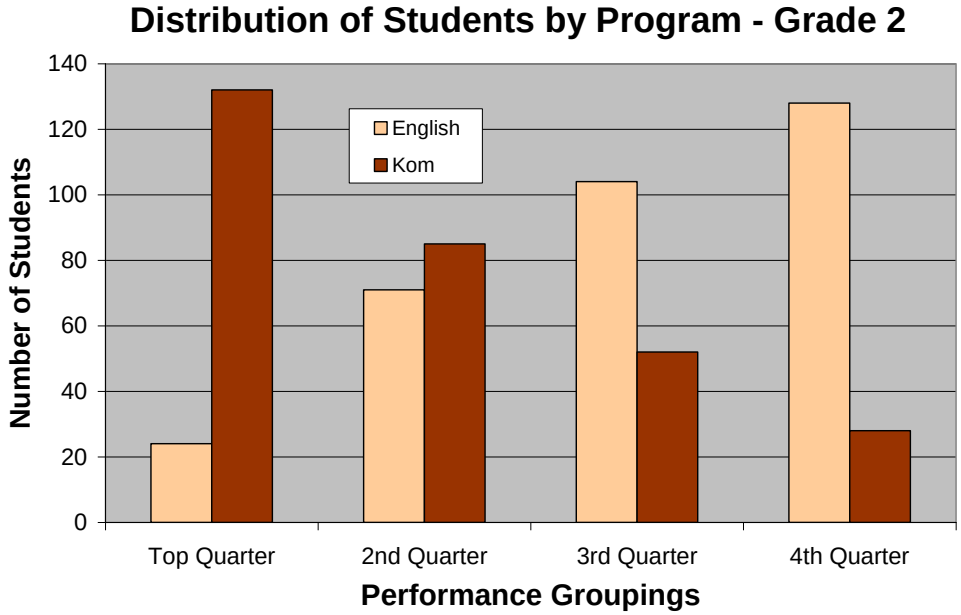
Figure 4. Comparison of Standard and KEP students in Class 2.



As indicated by the data in Table 8, the performance gap between the two programs is smaller in Class 2 than it was in Class 1 (either in 2008 or in 2009).

Comparison by student ranks. As in the case of Class 1, a ranking of students was done by performance to see how these compared when divided into four equal groups. Figure 4 contains the results of this ranking.

Figure 2. A comparison of the two programs in terms of the rankings of students by performance on the year-end assessment for Class 2.



The profile is very similar to that seen in Class 1. There are two slight differences. First, there are slightly more children from the Standard program in the Top Quarter and slightly more children from the KEP program in the bottom or 4th Quarter. The second difference is that the ratio of students in the 3rd Quarter (next to the bottom) is biased towards children from the Standard schools (not a good thing in this case) meaning that a greater proportion of English-medium children are in this next-to-the-lowest group than we found in Class 1.

Results – Specific Skills

Reading Comprehension

The ability to read is probably the most fundamental skill learned in basic education. Without this skill, the student is largely cut off from the opportunity to learn other content as much of this comes from reading whether reading the blackboard or reading from a book.

In developed countries, education is predicated on the assumption (and established experience) that children will be generally proficient readers by the end of Class 1 (reading appropriately graded text at 60 words a minute with good comprehension) and will be very competent readers by the end of Class 2 (90-100 words per minute with excellent comprehension of appropriately graded material). Researchers and educators working in the African context have commonly observed that five years of schooling are required before African children become passable readers. Even then, reading skills are weak and some have still not mastered the process. A recent EGRA[8] study done in Mali showed that by the end of Class 4, only 18 percent of students tested in the traditional Malian model—instruction via French—had reached even a minimal level of reading competency[9] and only 8 percent had reached an acceptable level (40 words per minute). At the end of Class 6, a total of 54 percent—just over half—had reached the most minimal level of reading proficiency and only 36[10] percent had achieved the standard of 40 words per minute.

This widely attested delay in developing reading skill has been attributed to such factors as lack of materials, poor instruction, inadequate funding, lack of parental support, etc. As we argued in the 2008 Kom report, we believe much of the delay can be attributed to language barriers—instruction in a second language and learning to read in a language which is poorly or inadequately understood. We present as evidence the results of testing on reading comprehension done this year in the Kom multilingual education project. In the common test administered in Class 2 to children in the control and experimental schools of the study, children were presented with a short text consisting of four simple sentences and a total of 22 words (three sentences and 22 words in the Kom version of the test). All children were asked to respond to four comprehension questions. Three of the questions were factual with the answer found directly in the text and one was a more general question asking what the story was about. A multiple choice format was employed with three choices for each question. Table 9 (on the next page) summarizes the results of performance on this task.

[8] EGRA is an acronym for Early Grade Reading Assessment. These assessments are currently being carried out by the Research Triangle Institute of Raleigh, North Carolina under contract to USAID and the World Bank. The intent is to solidly document the rate of development of reading skills in educational systems outside of the West. In the case of Africa, this typically means the *slow or delayed* development of such skills.

[9] In the EGRA research, a reading rate of 25-39 words per minute is considered minimum basic proficiency and a rate of 40 words per minute is considered acceptable proficiency. By contrast, in the US, 60 words per minute is considered average for Class 1 children and 90 words per minute is considered average or normal for Class 2 children.

[10] In the case of both Class 4 and Class 6 the 18 percent and 54 percent respectively **include**s those who achieved at the higher level—8 and 36 percent respectively.

Table 9. Comparison of performance on the reading comprehension task for children in the control and experimental programs.

	English-medium schools			**Kom-medium schools**		
Questions answered correctly	N	Percent of Total	Cumulative Percent[11]	N	Percent of Total	Cumulative Percent
0	107	32.8	100	35	11.8	99.9
1	114	35.0	67.2	41	13.8	88.1
2	78	23.9	32.2	53	17.8	74.3
3	26	8.0	8.3	69	23.2	56.5
4	1	0.3	0.3	99	33.3	33.3
	326			297		

In the case of the KEP program, it is very evident that reading comprehension is strong by the end of Class 2 in that 56.5 percent of children read the included text at 75 percent or higher comprehension and almost 75 percent correctly answered two or more of the comprehension questions. This level of performance reflects clear evidence of reading skill at the end of Class 2, a level of performance which would appear to place such children between Class 4 and Class 6 in the Malian results and between Class 4 and 5 in the baseline study on reading done in the Kom area before the project began.

The performance of the children in the English-medium schools is a little harder to interpret. Taken at face value, the data show that 8.3 percent of children read with considerable comprehension and 23.9 percent appeared to read with moderate comprehension. Both the Malian EGRA data and our own baseline survey on reading skill development in the Kom area suggest that we should expect little to no reading skill in English at the end of Class 2. Are the children in the English-medium schools in the Kom area doing better than other research would predict? The answer to this question probably lies in the testing format used—a three-item multiple choice format. In this format, the test-taker has one chance out of three of correctly answering a question simply by randomly guessing from among the available choices.

Fortunately, it is possible to model the effect of random guessing in an effort to judge whether the observed scores might be a product of pure guessing. Table 10 compares the results of a random guessing model with those observed for the reading comprehension task on the part of children in the English-medium schools.

Table 10. Comparison of observed test performance with that predicted by a random guessing model.

Modeling Scores on the Reading Comprehension Task of Children in Class 2 of Standard Schools.			
Questions answered correctly out of 4 with 3-item multiple choice	Performance predicted by a random guessing model	Observed performance from the test	Difference between observed and predicted
0	65	107	+42
1	129	114	-15
2	95	78	-17
3	31	26	-5
4	3	1	-2

By comparing columns 2 and 3, we observe that the performance of the English-medium children on the reading comprehension task was actually **lower** than predicted by the strategy

[11] In this case, cumulative percentage has been computed from high-performing to low-performing because this method facilitated discussion about the item.

of random guessing (more scored zero than predicted, and fewer answered some questions correctly than predicted by the model). The most likely explanation for this result would seem to be that approximately 40 children failed to even attempt the task and almost all of the remainder could do nothing other than guess at the correct answer. This result would coincide with the previous baseline testing and with other results such as those from Mali.

A comparison of results for children in Classes 1 and 2 in the Kom-medium program is instructive in terms of being able to "observe" progress in learning to read in the early classes. Table 11 compares performance between KEP[12] Classes 1 and 2 on the measure of reading comprehension for the two classes. Figure 5 contains exactly the same information as Table 11 but is included for those who find it easier to interpret a graph than a table.

Table 11. Growth in reading comprehension for KEP students between Classes 1 and 2.

	KEP Class 1		**KEP Class 2**		
Items answered correctly	N	Percent of Total	N	Percent of Total	Change
0	60	17.9	35	11.8	- 6.1
1	79	23.5	41	13.8	- 9.7
2	78	23.2	53	17.8	- 5.4
3	77	22.9	69	23.2	0.3
4	42	12.5	99	33.3	20.8
	336		297		

Figure 5. Development of skill in reading comprehension between Classes 1 and 2 in Kom-medium schools.

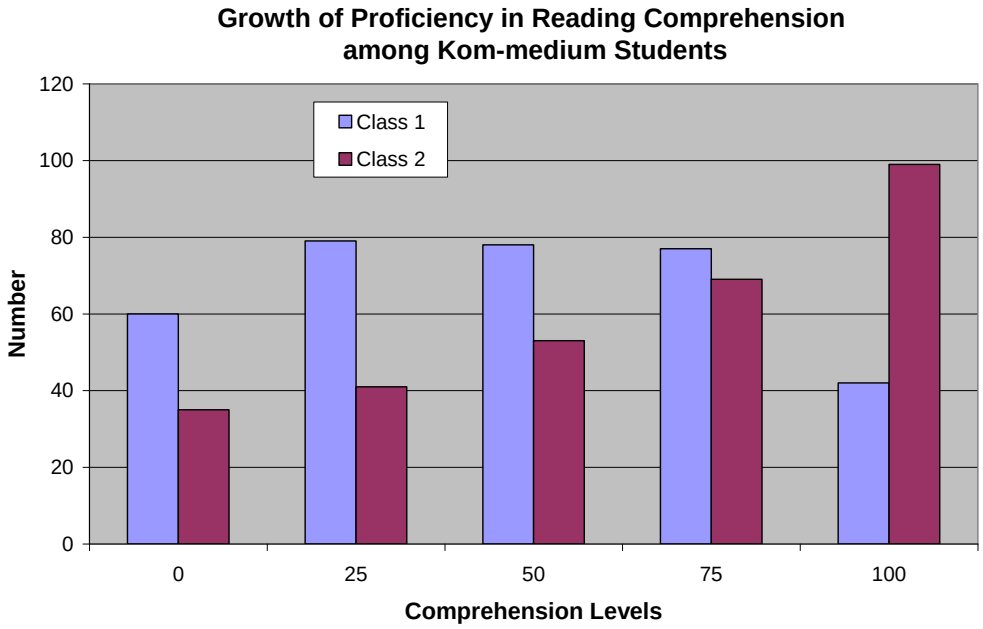


Figure 5 shows Class 1 students being approximately normally distributed in each of the comprehension categories with a mean close to the 50 percent category. The most plausible interpretation of this profile is one which says that substantial learning has taken place but the skill in question has not yet been mastered. By contrast, Class 2 students show a marked

[12] The performance of children in the English-medium classes is not being included in this comparison as performance was so low as to not be helpful with more than 75 percent either not attempting any of the questions or failing to answer any of them correctly.

increase in reading proficiency as one moves from low to high levels of comprehension. This profile is consistent with a 'mastery' profile though there are still more children in the lower categories than educators would like to see.

Perhaps most significant of all is the fact that 35.4 percent of Class 1 students in the KEP program were reading with a high level of comprehension after just one year of instruction—a number which jumps to 56 percent after Class 2. Clearly, we see that children in African schools can learn to read after just one or two years of instruction IF appropriate instructional strategies and materials are being employed.

Learning of Mathematics

Math is one of the core subjects of early basic education and one which is known to be troublesome for the children of Cameroon. Therefore, it is worth taking a close look at how children did at learning basic math skills in the two programs.

	Class 1 - 2009			**Class 2 - 2009**		
	Standard	**KEP**	**Statistics**	**Standard**	**KEP**	**Statistics**
Counting	15.2	40.6	P < 0.000	Not tested in Class 2		
Place Value	6.2	33.4	P < 0.000	Not tested in Class 2		
Addition	10.9	46.7	P < 0.000	18.8	54.7	P < 0.000
Subtraction	9.5	35.7	P < 0.000	10.4	32.2	P < 0.000
Problem recogn.	Not tested in Class 1			31.3	66.8	P < 0.000
Overall	10.4	39.5	P < 0.000	23.7	51.2	P < 0.000

The Class 1 results from the Standard schools certainly lend support to those who assert that children in Cameroon struggle to learn math. Performance on even the most basic task of counting was poor—only 15.2 percent. With an overall mean score of just 10.4 percent on the math section of the test, we would be led to conclude that there is something fundamentally flawed about how math is being taught in the schools of the country. However, performance of children in the KEP schools—a mean score almost 4 times as high as that in the Standard schools—suggests that the problem is not Cameroonian children nor basic instruction in math. Rather, the data suggest, there is some other problem. It is possible that children are doing poorly in math because they do not understand what the teacher is saying when the subject matter is being presented. It is also possible that there is not enough time on task in the Standard schools given the very crowded instructional calendar that specifies what is to be taught and how much time is to be given to each subject. It is also possible that math is being taught in such a way—choral repetition with little individual interaction with math problems—that it is not actually being learned in the Standard schools.

The overall scores are higher in Class 2 than in Class 1 suggesting that children in both programs have made progress in learning their math. However, the reality may not be as favorable as it appears. The tasks under Problem Recognition were all multiple choice whereas the others required actual computation. It would appear the multiple choice format may have pushed the scores up somewhat artificially on that task. Performance on the tasks of addition and subtraction—both basic math content in Class 2—was quite similar to that of Class 1 (the actual problems to be solved, of course, were more difficult and suited to the curriculum of Class 2).

What do we learn from the Math results? First, the large advantage in math for the KEP schools parallels that seen in reading. While one might argue that the higher scores seen in reading were due to the language of the test (being a second language for those in the Standard program and in the first language for those in the KEP program), it is harder to make this

claim with respect to math since the items in this section generally required only the doing of basic computation with no language involved.

The data suggest that math skills are not well taught and learned in Cameroon even in the first language. This may be the result of any number of variables; inadequate time on task, lack of adequate drill and practice, poor instruction and/or instructional methodology, inadequate or overly accelerated curriculum, etc. At the same time, however, the data suggest that language of instruction may well be a major factor in math learning since there is such a large differential between those in the Standard and KEP programs. A learning advantage of 300 or 400 percent is enormous making it hard to avoid the conclusion that math instruction benefits enormously from language of instruction (in the first language).

Oral English

We included a test of oral English in the year-end assessment especially in an effort to monitor how the KEP children were doing relative to those in the Standard program in learning English. In Appendix B there is a more extended (and technical) discussion of whether an assessment of oral English is the best way to monitor English language learning on the part of the children in the region AND whether progress in learning oral English correlates with general mastery of curricular content.

Worldwide, parents tend to object to the notion of instruction via the first language fearing that this will compromise their children's educational future. They are especially concerned that their children learn the language of business and government. In the case of the Kom area, that language is English. In our early negotiations in the area about this pilot project, parents and educators alike expressed the concern that beginning instruction in the first language would hinder mastery of English, the language of higher education as well as business.

Researchers have found that, contrary to what one would expect, children receiving a strong early foundation in their first language with a more gradual introduction to the second language actually progressed further than those receiving all of their instruction in the second language (Thomas and Collier, 1997). For this reason, we have paid close attention to the issue of second language mastery in an effort to allay the concerns of both parents and educators that the KEP model is going to compromise English development.

The test of oral English includes sections on social interaction, ability to respond to requests for factual information or to obey specific instructions, and a section testing children's ability to use the language to comment on the content of a picture or illustration. The exact same test was administered to children in both programs in the same way. Table 13 contains the results of that testing for 2009 in both Classes 1 and 2 in both control and experimental schools. In looking at the data, it is important to keep in mind that children in the control (Standard) schools receive ALL of their instruction in English so are exposed to the language for the entire school day. The children in the KEP program receive approximately an hour a day of instruction in English as a second language.

Table 13. Comparison of performance on a test of oral English in both programs in Classes 1 and 2, 2009.

	Class 1 - 2009			**Class 2 – 2009**		
	Standard	**KEP**	**Statistics**	**Standard**	**KEP**	**Statistics**
Social Exchange	77.9	82.8	$p = 0.117$	80.1	86.0	$p = 0.002$
Factual Response	51.0	80.0	$p < 0.000$	56.5	69.8	$p < 0.000$
Free Response	18.5	24.5	$p = 0.006$	32.3	39.3	$p = 0.003$
Overall	42.8	55.0	$p < 0.000$	50.8	59.5	$p < 0.000$

The data in Table 13 provide reasonably strong evidence that being in the KEP program is not compromising those students' development of proficiency in oral English. The only one of the comparisons which was not statistically significant in Class 1—Social Exchange—became significant in Class 2 meaning that those students actually improved relative to their peers in the Standard schools. Apart from the lower performance on the Factual Response task in the KEP schools in Class 2, the KEP students generally maintained their advantage in oral English over those in the Standard schools.

(The interested reader is referred to Appendix B for a more detailed discussion of this finding.)

Participative Impacts: Enrollment, repetition, and attrition

So far our analysis has focused on differential cognitive achievement between the two programs. Thomas and Collier (1997) found that attrition and repetition rates were lower when children received early instruction in their first language, but that was in the west. Is the same effect going to be true in a developing country? The argument has been that if children do better in school (because of their L1 instruction), they will feel more successful and thus be less likely to drop out. Now that we have two years worth of data, we are in an initial position to see whether there might be a similar pattern in the Kom area. Table 14 presents the relevant data.

Table 14. Data on enrollment, promotion, dropouts, and repetition in the research project.

	Standard Schools	**KEP Schools**
# Enrolled in Class 1 in 2008	336	325
# Enrolled in Class 2 in 2009	326	297
# from Class 1 2008 completing Class 2 in 2009	130	192
# who are recorded as repeating Class 1 in 2009	22	10
# who enrolled in Class 1 in 2008 for whom there is no further record	177	122
# in Class 2 2009 not enrolled in Class 1 2008	181	102

After we generated this initial set of data from out database, we immediately were led to question what we had found. How was it possible that only 130 out of 336 students in the Standard schools and 192 out of 325 students in the KEP schools from 2008 moved on to and completed Class 2 in 2009? We did identify 22 and 10 students in the Standard and KEP programs respectively who repeated Class 1 in 2009. This still leaves over a hundred students in both programs unaccounted for. As of this writing, we consider these data soft and unreliable. One of the known sources of discrepancy is the unreliability of names entered on the test booklets by children, teachers, and/or examiners. Careful scrutiny of these lists has identified about 16 percent of cases where the names are close enough to make corrections. That still leaves 84 percent of the cases without any explanation.

If these data hold up (we are working on trying to resolve some issues regarding student names), then we have the following rather grim statistics:

- Those who moved from Class 1 (2008) to Class 2 and completed the year are only 38.7 percent in the Standard program and 59.1 percent in the KEP program. (a relative improvement rate of 52.7 percent because of KEP)

- Those who simply abandoned school after Class 1 are 53.9 percent in the Standard program and 37.5 percent in the KEP program. (43.7 percent improvement (decrease) in the rate of abandonment because of KEP)
- Those who formally repeated Class 1 are 6.5 percent in the Standard program and 3.1 percent in the KEP program. (110 percent reduction in this statistic in the KEP program).
- The percent of repeaters in Class 2 (or transfers in from other schools) was 55.5 percent in the Standard program and 34.3 in the Kom program. (We have no current explanation for this finding since both would have had English instruction in their previous class.)

Even with the softness in the numbers and the problem of name confusion, gains in efficiency due to the KEP program seem to be substantial. However, we will not consider these data or claims to be reliable or valid until we have been able to resolve the names issue.

The Issue of Age as a Variable

In the Kom area as in much of Africa, it is common to find a rather broad range of ages in the classroom. In the case of Kom, the range is large though we lack the data to say that the disparity is any greater in the Kom schools than in other parts of the country or the continent. Although there are official policies on age requirements to go to school, several factors work together to circumvent these policies. These include the fact that older children are charged with the care of younger siblings, the desire of parents to send their children to "pre-school", and the need of the schools to have as many children enrolled as possible to maximize their incomes from fees. As a result, we found children as young as three years of age in Class 1 and four years of age in Class 2. The presence of these under-age children clearly constitutes a management and instructional problem for the teachers. We also present evidence that their presence negatively impacts school performance.

For the purposes of this analysis, we will define 'under-aged' as having an age of less than six when entering Class 1 and less than 7 when entering Class 2. Table 15 reports performance by age in each of the programs for Class 1.

Table15. Comparison of performance of children in the two programs by age in Class 1.

	Standard Program		KEP Program		
Age	N	Mean Score	N	Mean Score	Differential
3	3	3.82	--	--	--
4	15	14.12	33	35.33	21.21
5	76	18.88	98	39.92	21.04
6	139	21.29	121	50.09	28.80
7	86	22.63	57	53.27	30.54
8	25	26.46	14	62.35	35.89
9	4	21.47[13]	7	64.78	43.31
10			3	66.40	--
11					
TOTAL	348	20.9	333	47.1	26.2

Several observations stand out from this data. First, it is evident that schools in both programs are admitting a substantial number of under-age children—27 percent in the case of schools in the Standard program and 39.3 percent in the experimental program. Second, it is

[13] Two of the four 9 year olds in this grouping scored extremely low—a mean score of 7 percent— suggesting they probably have significant to severe learning limitations. The mean score of the other two was 35.3 percent.

clear that under-age children perform well below the level of appropriately aged children in both educational models (26 percent lower in the KEP program and 20.3 percent lower in the Standard program). If the mean scores of both groups were adjusted to exclude the under-age children, the mean score in the Standard program would be 22.1 and 52.4 in the KEP program. Third, we note in both programs that mean performance increases with age. The 8 and 9 year olds in both groups scored well above their younger classmates (cf. the footnote regarding the 9 year olds in the Standard program).

The data in Table 16 provide clear support for the national policy that children are not really ready for the intellectual demands of Class 1 until they reach 6 or 7 years of age. While the 5 year olds are not dramatically behind the 6 year olds, there is clearly a developmental divide between ages 5 and 6 in terms of the learning demands of Class 1. While it may be convenient for parents to send under-age children to school with older siblings and for schools to accept them (along with the accompanying fees), the practice is not a positive contribution to the educational development of Cameroon.

The children in the core age group (6 and 7 year olds) in the KEP program realize an absolute advantage of approximately 29 percentage points over their age peers in the Standard program and a relative improvement gain of 135 percent—a very large gain.

Table 16. Comparison of the performance of children in Class 2 by program and age.

	Standard Program		KEP Program		
Age	N	Mean Score	N	Mean Score	Differential
4	2	25.65	3	31.74	6.09
5	9	23.91	12	43.55	19.64
6	61	33.10	54	54.48	21.38
7	122	31.78	74	53.86	22.08
8	73	33.14	67	53.93	20.79
9	34	39.62	49	54.55	14.93
10	20	40.70	29	55.46	14.76
11	4	54.35	8	69.95	15.60
12	--		1	75.22	--
	325	33.76	297	54.12	20.36

The age/performance profile of Class 2 is quite different from that of Class 1. The 6 year olds who might otherwise be classified as under-age appear to do just as well as (actually slightly better than) the 7 year olds. On the basis of performance, only the 4 and 5 year olds are truly under-age in terms of their ability to keep up with their older peers in both the Standard and KEP programs. Based on this classification, the Standard program has only 3.4 percent under-age children with the KEP program having 5.1 percent. These figures represent a distinct improvement over the presence of under-age children in Class 1. It would appear that either teachers or school administrations managed to convince parents that most of the under-age children from Class 1 should either repeat or wait under they are old enough to attend rather than moving on to Class 2 and doing poorly there as well.

As we saw in Class 1, the data indicate that performance goes up with age. Class 2 children who were 11 or 12 (no indication why they were in Class 2) score well above the age core of Class 2 (those 6, 7, and 8).

In the age core of Class 2, we note a consistent absolute advantage of about 21 percentage points and a relative advantage or gain of 65 percent. Put differently, the KEP innovation produced a 65 percent improvement in educational outcomes. This is substantially less than observed in Class 1 but still very large and larger than the relative gain seen in similar multilingual education projects in other countries.

Primarily for the sake of completeness we have included a listing of all of the schools in the study—both control and experimental—and the percentage of their Class 1 students who were under-age. The proportion ranges from 0 to 90 percent. The KEP schools seemed to be slightly more likely to have under-age children. We can also observe that schools generally performed more poorly when they had a higher percentage of under-age children though the correlation is not tight. We suspect that those schools with higher proportions of under-age children probably admitted them for financial reasons.

Table 17. Percentage of children in each school who would be characterized as under-aged.

	Program	Enrolled	Percent Under-aged	Mean Score
CBC Abuh	English	5	80	26.64
CBC Belo	KEP	14	0	72.63
CBC Fujua	KEP	15	46.7	31.65
CBC Fundeng	English	21	42.9	20.37
CBC Kikfuini	KEP	7	0	54.63
CS Kindo	English	13	76.9	14.06
CS Wombong	English	12	66.7	37.61
GS Ameng	KEP	18	38.9	50.53
GS Atondum	English	58	5.2	24.62
GS Baichu	English	19	5.3	23.24
GS Bolem	KEP	34	5.9	46.95
GS Fundong Village	English	33	63.6	24.79
GS Ilung	KEP	29	75.9	65.62
GS Kitchu	KEP	38	89.5	42.10
GS Laikom	KEP	20	35	33.93
GS Mboh	KEP	39	25.6	50.52
GS Meli	English	64	0	15.02
GS Mentang	English	43	16.3	10.31
GS Muteff	KEP	62	35.5	47.88
GS Ngwah	KEP	45	40	31.82
GS Njinikijem	English	41	26.8	28.26
GS Wombong	KEP	12	16.7	64.73
GS Yuwi	English	27	74.1	14.49
PS Ngwah	English	12	0	36.29

Analysis shows a weak correlation between the percentage of under-age children in a school and the mean performance of all children in that school. This correlation shows about a 1 point decline in school mean scores for each 12 percent increase in the proportion of under-age children in the school. If a school has 50 or 60 under-age children, school mean performance will be significantly affected. The strength of the correlation is not strong, however, (r-sq = .027) so that other factors such as teacher quality and classroom size play bigger roles in determining the extent of the impact of having a large number of under-age children in the classroom. What is clear is that some schools are admitting large numbers of under-age children and such schools experience a performance penalty for having done so.

Gender as a variable

Program enrollment. Analyses of education data in Africa and most other developing regions have typically shown a marked gender effect in educational outcomes to the advantage of the male population. Therefore, it is useful to examine the data from the Kom area and the KEP project to observe the behavior of these variables in this project. Walter and Davis (2005) found evidence in Eritrea that mother-tongue education strategies tended to have a leveling effect of gender bias in education, at least in terms of performance. Table 18 presents

the relevant initial data in terms of the distribution of children by gender in control versus experimental schools.

Table 18. Comparison of enrollment by gender, program, year, and class.

	Standard Schools		KEP Schools		
	Male	Female	Male	Female	
2008 – Class 1	187	149	164	161	n.s.[14]
2009 – Class 1	183	166	148	188	p = 0.0281
2009 – Class 2	175	152	146	151	n.s.

From Table 18, it is evident that there were numerical differences in the enrollments in the schools of the two programs by gender. Only one of these differences—Class 1 in 2009—was statistically significant with there being more males in the Standard schools and more females in the KEP schools. Since the difference was not great and there is no discernible pattern (yet) in the enrollment, we have judged this one case to be primarily accidental.

Performance. Table 19 categorizes, by program, year, and class, the relative performance by gender in the schools which are included in this research project. Across years and classes, females scored higher than males on 9 of the 15 metrics of comparison in the Standard or English-medium schools. In the KEP schools, the comparison more strongly favors females 12 to 3. However, none of the differences was statistically significant so we are not able to draw very solid conclusions from this comparison. The mother-tongue schools appear to have a slight tendency to favor the female students. However, in this research population, the female students are more than holding their own in both program modes.

Table 19. Comparison of performance by program, gender, class, and year.

	2008 – Class 1		**2009 – Class 1**		**2009 – Class 2**	
	Males Higher	Females Higher	Males Higher	Females Higher	Males Higher	Females Higher
Standard Schools	Reading Oral English	Math Written Overall		Reading Math Written Oral English Overall	Reading Math Written Overall	Oral English
KEP Schools		Reading Math Written Oral English Overall	Reading Written Overall	Math Oral English		Reading Math Written Oral English Overall

The 'School' Variable

A number of researchers (Walter and Davis, 2005; Lynd, 2008) have noted that variation between schools in developing countries is a major factor in explaining variation in children's academic performance. In this section, we take a look at the school variable first looking at year-by-year comparisons and then citing statistical evidence in supporting the impact of this factor.

[14] n.s. is an abbreviation indicating that a comparison was not statistically significant.

Table 20. A general overview of the performance of the KEP schools for the period 2008-2009.

	Class 1		Class 2
School	2008	2009	2009
CBC Belo	60.0	72.6	62.7
CBC Fujua	69.2	31.7	46.0
CBC Kikfuini	63.0	54.6	50.0
GS Ameng	38.9	50.5	55.8
GS Bolem	62.2	46.9	63.1
GS Ilung	51.7	65.6	63.9
GS Kitchu	54.0	42.1	46.8
GS Laikom	35.3	33.9	54.7
GS Mboh	63.3	50.5	65.4
GS Muteff	46.5	47.9	40.3
GS Ngwah	54.3	31.8	51.6
GS Wombong	30.1	64.7	51.8

Scanning the table we observe dramatic differences both between schools within a given year and within a given school for two different years. In 2008, the lowest performing school trailed the highest performing school by 39 percentage points. In 2009 the difference between high and low schools was almost 40 points. Across years, we see similar changes—as high as 37 points in the case of the school CBC Fujua (this school was negatively impacted by a philosophical rift within the sponsoring organization with the result that the KEP trained teacher was removed and could not be replaced with a similarly trained teacher). Conversely, the school, GS Wombong made a huge improvement between 2008 and 2009. These massive changes up or down are usually the result of teacher issues. In fact, project staff reported that half of the KEP schools experienced prolonged teacher absences due to maternity leave, sickness, and transfers. In most cases, this meant that classes were closed for the period of time involved as suitable substitutes were not available. Such events also happen in the Standard schools but substitutes can usually be assigned since no special training is needed to teach as a substitute in the Standard schools.

Note from Kristine:

At CBC Fujua the Class 2 teacher didn't go on maternity leave, he spent the year at the teacher-training school. The class 1 teacher was transferred for political reasons.

The standard practice in the when teachers go on maternity leave is to combine classes so that one teacher teaches both classes for three months. They do not hire substitute teachers. In the KEPP project we didn't want to hinder the KEPP teacher (not on leave) by making him or her teach both classes. Therefore we paid for substitutes for the KEPP classes for the teachers on leave. We gave the substitute teachers some training.

With the exception of one school, GS Wombong, scores between Class 1 in 2008 and Class 2 in 2009 were relatively stable among the schools in the KEP program.

Table 21 compares the English-medium schools between years and classes in a similar manner. The differences in performance of Class 1 students between the two years are quite similar—within 10 percentage points—with just three exceptions. The school CBC Abuh had to be dropped from the study because it appeared to be on its way out of existence with just a few students and no teacher. Two schools—GS Atondum and PS Ngwah—apparently brought

in new Class 1 teachers so upgraded their performance substantially. All schools showed improvement of performance in Class 2 in some cases more than doubling mean student performance.

Table 21. A general overview of the performance of the Standard schools for the period 2008-2009.

	Class 1		Class 2
School	2008	2009	2009
CS Kindoh	13.2	14.1	40.5
CS Wombong	28.6	37.6	49.6
CBC Abuh/Wainchia	8.6	26.6	28.8
CBC Fundeng	23.7	20.4	26.4
GS Atondum	12.2	24.6	37.1
GS Baichu	14.0	23.2	33.1
GS Fundong Village	16.3	24.8	46.9
GS Mentang	11.2	10.3	21.1
GS Njinikijem	23.2	28.3	28.0
GS Yuwi	14.6	14.5	38.4
GS Meli	15.6	15.0	30.7
PS Ngwah	14.3	36.3	41.7

Additional, but more technical, analysis on the variable of 'Schools' can be found in Appendix E on Model Fit.

Observations about the 'absolute' level of performance of schools in the study

Most of the analysis to this point has focused on a comparison of students and schools in two different educational programs. At this point it is worth adding a little additional commentary about performance in general compared to what is happening in other developing countries. The following observations are based on our experience in several other countries having similar multilingual education programs.

The differential between control and experimental schools

There is not yet an adequate body of research to establish what is "normal" in terms of an expected differential or advantage for children and schools using the first language as a medium of instruction. Thomas and Collier (1997) report minimal differences in performance in the first three years of school in the US. In the Philippines, Walter and Dekker (2008) found a consistent advantage of 40 – 45 percent for children in first language classrooms beginning in Class 1 and extending through Class 3. Several researchers have reported somewhat vaguely that children in first language classrooms outperformed their peers in second language classrooms but did not quantify the differences. Lynd, for example, evaluating test scores from Uganda reported that children in first language schools outperformed their peers in second language schools but did not quantify the differential.

In the testing of Class 1 children done in 2008, KEP children showed a 225 percent advantage (53 percent versus 16 percent). This differential seemed too large to be sustained and, some would suggest, to be believed. Yet it was a consistent difference across subject matter (apart from oral English) as well as between schools so cannot be dismissed as meaningless. Our view at this point is that the quality of instruction in the schools of Boyo—and probably in rural Cameroon as a whole—has slowly and almost imperceptibly declined over the last twenty years to such a low level that even a modest innovation like the KEP program shows a huge differential advantage over the existing schools.

The differential at the Class 1 level in 2009 has declined to 125 percent—still large but a sharp decrease in difference compared to 2008. Many of the Standard schools seemed to have achieved sizable gains compared to 2008—maybe because 2008's results were embarrassingly low. At the same time there was a noticeable though not huge decline in performance in the KEP schools—partly for technical reasons and partly because of political, health and administrative issues involving teachers at several KEP schools having nothing to do with the KEP innovation per se.

'Absolute' performance of both programs

The test administered was a criterion-referenced test meaning that it only covers content that the curriculum defines as "to be mastered" during the specified year. On such tests one normally expects mean performance to be somewhere between 60 and 80 percent depending on the many variables that make up the instructional process. Performance of students in other similar projects where the authors have worked or are doing research have typically been in the 65-75 percent range with students in the Standard model having mean scores of 45 to 55 percent.

This year (2009) mean performance of children in the KEP program was 47.1 percent—approximately 20 points below the expectation of 60-75 percent. The performance of those in the Standard program was 22.9 percent—approximately 30 points below the observed or expected mean in similar programs elsewhere. We note in both cases that the programs are scoring 20-25 points below what we have seen in similar programs elsewhere. What accounts for this 20-25 point deficit or handicap? The answer to this question may well be the key to the future of basic education in Cameroon.

Conclusion

The second year of the KEP innovation in education is now "in the books." The results from the testing for Class 1 generally duplicate the findings for 2008 when the children in the KEP program outscored those in the Standard program by more than 3 to 1. The results for Class 2 continue to demonstrate that the innovation is capable of bringing dramatic improvement to basic education in Cameroon. In the coming 2009-2010 school year, the innovation will extend to Class 3. We look forward with keen anticipation to seeing the results from year three and Class 3.

As the report has made clear, despite the apparent effectiveness of the KEP innovation, basic education in Cameroon, especially in rural areas, continues to face very significant challenges in the struggle to deliver quality education to the children of the country. While we are convinced that the KEP innovation can bring about dramatic improvement in educational outcomes in appropriate contexts (linguistically homogeneous areas), it is not, by itself, a complete answer to the educational needs of the country. Other innovations are needed as well.

References

Abedi, Jamal and Leon, S. (1999*). Impact of students' language background on center-based performance: Analysis of extant data*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Abedi, Jamal. (2002). Standardized Achievement Tests and English Language Learners: Psychometrics Issues. *Educational Assessment,* 8(3), 231-257.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Cummins, Jim. (1998). Beyond adversarial discourse: Searching for common ground in the education of bilingual students. Presentation to the California State Board of Education, Feb 9, 1998. Also in Ovando and McLaren (eds.) *The politics of multiculturalism and bilingual education: students and teachers caught in the crossfire*. New York: McGraw-Hill.

Dutcher, Nadine. 1995. *The Use of First and Second Languages in Education. A Review of International Experience*. Washington, D.C. The World Bank.

Lindholm-Leary, Kathryn J. (2001). *Dual language education*. Clevedon, UK: Multilingual Matters.

Lynd, Mark R. (2008). Measuring learning achievement 2008: Follow-up for P2 and baseline for P3. Core Report for the Unity Project, Uganda. (USAID contract). School-to-School International.

National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs, (2002). *Survey of the states' limited English proficient students and available educational programs and services*. Washington, DC: Author.

Thomas, Wayne C. and Virginia Collier. (1997). School effectiveness for language minority students. National Clearinghouse for Bilingual Education. www.ncbe.gwu.edu/ncbepubs/resource/effectivenss/index.htm.

Walter, Stephen L. and Patricia M. Davis. (2005). *Eritrea National Reading Survey – 2002*. Dallas, TX: SIL.

Walter, Stephen L. and Kristine Roth Trammell. (2008). *Results of the First Year of the Kom Experimental Education Project.* (Unpublished research report).

Walter, Stephen L. and Diane Dekker. (2010). *Testing the multilingual education model in the Philippines*. International Education Review (to appear).

Appendix A

IntraProgram Comparisons Across Years

Of particular interest to the program administrators is a comparison across years but within programs. The following two tables provide the data describing these comparisons.

KEP 2008 and 2009

Table A.1. Comparison between years within the KEP program.

Comparison of Results in the KEP Program between 2008 and 2009				
	2008	**2009**	**Statistics**	**Comment**
Language Arts				
Word Recognition	68.5	47.6	T = 6.29; p < 0.000	Significantly lower in 2009
Grammar	57.8	46.0	T = 5.45; p < 0.000	Significantly lower in 2009
Reading Comprehension	64.3	47.3	T = 6.38; p < 0.000	Significantly lower in 2009
Mathematics				
Counting	51.3	40.6	T = 3.20; p = 0.001	Significantly lower in 2009
Place Value	28.2	33.3	T = 1.66; p = .096	No difference
Addition	48.5	46.7	T = .585; p = .559	No difference
Subtraction	45.3	35.7	T = 3.18; p = .001	Significantly lower in 2009
Oral English				
Social Interaction	87.3	82.8	T = 2.52; p = 012	Significantly lower in 2009
Factual Response Items	73.8	79.8	T = 2.41; p = .015	Significantly higher in 2009
Free Response Items	47.7	24.5	T = 6.84; p < 0.000	Significantly lower in 2009

In the body of the report, it has been noted several times that performance within the KEP program was slightly lower in 2009 compared to 2008. Table A.1 compares all tested skill areas between the two years in order to make such a comparison more explicit. From looking at the table, we note that performance in 2009 was (statistically) significantly lower in 7 out of 10 skill areas, significantly higher in just one, and equal in two—place value and addition.

This kind of comparison obviously leads one to search for explanations of which there are many. Both of us expected performance to be higher in the experimental program in 2009 because the training was better done, materials were out and available before the training began, several technical adjustments were made, newly trained teacher had the support of their predecessors from the year before, there was better supervision of the new teachers, and the program had developed some momentum within the educational infrastructure of the area. Instead, performance was down somewhat.

On the flip side, there more problems with teachers in 2009. At least half reported significant time lost because of teacher absences. Two teachers went on maternity leave with no trained teacher to replace them. Three teachers had prolonged illnesses during which their classes did not meet. Two of the private schools were seriously affected by an ideological rift within the organization that was sponsoring them with accompanying loss of students, morale, and momentum. The biggest factor, however, seemed to be teacher motivation. For their involvement in a 'new' program, the teachers wanted to be paid extra. It did not matter to them that they benefited from free extra training, free books and teaching aids in their classrooms, additional supervisory support and encouragement, and the same salaries being paid to their peers to do the same job. They wanted additional payment for doing something different from what their peers were doing. The non-monetary professional benefits were apparently not meaningful to them.

It is also possible, of course, that the group of teachers named to teach the new Class 1 in 2009 was just not of the same level of capability as those named in 2008. In our opinion, the data does not provide much support for this interpretation, but we are not able to rule it out.

The Standard Program 2008 and 2009

Table A.1. Comparison between years within the KEP program.

Comparison of Results in the Standard Program between 2008 and 2009				
	2008	**2009**	**Statistics**	**Comment**
Language Arts				
Word Recognition	15.0	14.2	T = .437; p = .662	No difference
Grammar	5.62	15.9	T = 7.37; p < 0.000	Significantly higher
Reading Comprehension	4.17	6.9	T = 2.52; p = 0.011	Significantly higher
Mathematics				
Counting	14.1	15.2	T = .548; p = .584	No difference
Place Value	.45	6.2	T = 4.98; p < 0.000	Significantly higher
Addition	12.75	10.9	T = .947; p = 3.44	No difference
Subtraction	8.48	9.5	T = .585; p = .558	No difference
Oral English				
Social Interaction	72.5	79.8	T = 3.73; p < 0.000	Significantly higher
Factual Response Items	59.3	51.0	T = 2.87; p = 0.004	Significantly lower
Free Response Items	18.7	18.7	T = 0; p = 1	No difference

Appendix B

Oral English and Learning in the Schools of Kom

In the North West Province of Cameroon, English is the official language and the language of business, government, and education. It is natural, then, that parents see mastery of English as key to future opportunities for their children. In their logic, the best way to ensure that their children are going to learn English is to be sure that they are taught in English in the classroom. The same perspective and logic is widespread in developing countries, not only among parents but also within educational institutions and governments. The existence of generations of students who went through such schooling without effectively learning English has not dispelled this belief. It may be that the occasional students who did succeed in this system is sufficient evidence to the contrary to satisfy most observers.

At the same time, there is little doubt that children are not going to learn English without some instruction. The key issue thus becomes, "What kind of instructional model is going to work best to maximize mastery of English (and other content) in the schools of Boyo?" The Thomas and Collier (1997) in their research on multilingual education in the US found that children were more successful in learning English (and other content) when they had a solid and positive foundation in education laid through their first language. Dutcher (1995) also cites evidence that having a solid linguistic background in a familiar language enhances and accelerates one's ability to learn a second language.

Because of these contrasting perspectives and research findings, we have made a point of paying attention to the development of skill in English in the two programs in the Kom area. Table B.1 recounts the data from earlier in the paper on tested performance in the oral acquisition of English.

Table B.1. Comparison of performance on a test of oral English in both programs in Classes 1 and 2, 2009.

	Class 1 - 2009			**Class 2 – 2009**		
	Standard	**KEP**	**Statistics**	**Standard**	**KEP**	**Statistics**
Social Exchange	77.9	82.8	$p = 0.117$	80.1	86.0	$p = 0.002$
Factual Response	51.0	80.0	$p < 0.000$	56.5	69.8	$p < 0.000$
Free Response	18.5	24.5	$p = 0.006$	32.3	39.3	$p = 0.003$
Overall	42.8	55.0	$p < 0.000$	50.8	59.5	$p < 0.000$

The data in Table B.1 makes several points. First, contrary to the fears of many, children in the KEP program have not suffered in terms of their learning of oral English. In fact, quite to the contrary, they show a consistent advantage in this regard compared to their peers in the English-medium schools consistency outperforming them on all measures included in the assessment.

Second, the advantage seen in Class 1 seems to have narrowed slightly in Class 2 though the difference remains statistically significant. Looking in the Table, this seems mainly due to a decrease on the measure, factual response, in the KEP schools. We lack separate empirical evidence for the decline on this measure since it is not mirrored in the other measures.

Third, we note a large disparity between performance on the very common "social exchange" items and the more demanding items which require students to think of and generate responses. This is not surprising. However, what is more surprising is that the children in the KEP schools did better on this task than did those in the English-medium

schools. This finding is consistent with the research evidence supplied by Thomas and Collier and Dutcher (1995). It will be particularly interesting to see what happens in Class 3.

The research data coming from the KEP projects sheds some additional light on the "English question" which is worth making explicit. It has been widely assumed—whether implicitly or explicitly—that 'real learning' only takes place in English and that English is the best vehicle for that purpose. The research data provides a way of testing this assumption.

If the assumption is correct that the "best learning" takes place when English is used as a medium of instruction, then it would be reasonable to find that mastery of English would be the best predictor of learning for the students in the research project. We can test this assumption separately for both programs. The relevant data are in Table B.2.

Table B.2. Predictors of overall performance for Class 1 students.

	Standard		**KEP**		**Combined**	
	r-sq.	r	r-sq.	r	r-sq.	r
Word Recognition	.141	.375	.354	.595	.412	.641
Grammar	.155	.394	.293	.541	.403	.635
Reading Comprehension	.063	.252	.337	.581	.426	.653
Language Arts	.154	.392	.385	.620	.423	.650
Math	.244	.494	.431	.657	.495	.704
Oral English	.143	.378	.117	.342	.173	.416

Six sections of the achievement test administered were each compared to the overall score on the test to see which best "predicts" or correlates with the overall score.[15] The term, 'r-sq.' is a measure of the amount of variation in test scores explained by or accounted for by the sub-measure in question. So, for example, the correlation between Grammar and the Overall Score for students in the English-medium schools accounts for 15.5 percent of the variation in the Overall Score. Another way to say this is that performance on the Grammar test 'predicts' 15.5 percent of the variation in the Overall Score.

The value in the second column labeled 'r' is a measure of the strength of the correlation between each subsection and overall performance. All correlations in Table B.2 are statistically significant so specific values of the statistics, t or p, are not given.

With this little bit of technical background, note which subsection of the test best correlated with the Overall score? In both programs, Math correlates most highly with the Overall score. Of the six subsections analyzed, Oral English is the third-weakest predictor for students in the Standard program and the weakest for the entire sample tested.

[15] The statistically knowledgeable will ask about *autocorrelation*. To reduce or eliminate autocorrelation, achievement on each subsection of the test was removed from the overall score before the correlations were done.

Appendix C

The Language of Testing

Any program and any research which crosses linguistic boundaries is bound to raise questions about the role that language might have played in program outcomes. This is clearly an issue in any kind of educational innovation involving multiple languages. On the one hand, educators and researchers investigating the impact of language as a medium of instruction have found that children being instructed in the language they know best seem to learn much more effectively. On the other hand, how does one construct an assessment guaranteed to measure 'true' learning while factoring out language as an intervening variable since all normal educational assessments use language as the medium of assessment whether in written or oral form. In terms of the KEP project and associated research, the question is this, "Is it the case that Kom children being taught in English are learning less, or is it the case that being tested in English gives a false measure of what these children have learned?" And, of course, it is possible that both statements are true at the same time. Is there any research on this question? The answer is, "Yes," but this research fails to resolve all aspects of the problem.

Educators and psychologists in the US have had enough experience with language as a factor in instruction, assessment, and diagnosis that they are beginning to actively grapple with this question. The issue is large enough that the US *Standards for Educational and Psychological Testing* (AERA, APA, & NCME; 1999) state:

> For all test takers, any test that employs language is, in part, a measure of their language skills. This is of particular concern for test takers whose first language is not the language of the test. … In such instances, test results may not reflect accurately the qualities and competencies intended to be measured. (p. 91)

A series of studies done earlier in this decade by the US National Center for Research on Evaluation, Standards, and Student Testing found evidence that the performance of non-native speakers of English on assessments of math and science was affected by their reduced proficiency in English (Abedi & Leon, 1999). If this is true in the US, then it is probably true everywhere that we find language dissonance in the classroom. But is the problem sufficiently salient that we need to take it seriously in settings like that of Kom or Cameroon in general? Fortunately, we have some very specific research to help answer that question.

Abedi (2002) sought to investigate the *extent* of the impact of limited proficiency in English on student performance on standardized assessments. For this purpose Abedi gathered large-scale datasets from four different sites in the US having a fairly significant number of ELLs (English Language Learners). The datasets included 1.2 million children between Classes 2 and 11 of which 22.3 percent were ELLs. In an effort to control for and investigate other variables known to affect educational outcomes, the datasets included information on the socioeconomic status of the children and level of education of parents. The measurement variable was performance on either the Stanford 9 or the Iowa Test of Basic Skills[16], both widely used and nationally normed standardized tests.

The data in Table 1 below are all taken from Site 2 which was the largest site (an entire US state). The Grade 2 population was just over 414,000 students with 30.2 percent of these being ELLs. The Grade 7 population was just short of 350,000 children with 21.2 percent

[16] For readers of this paper not from the US, the Stanford 9 and the Iowa Test of Basic Skills are nationally standardized tests that schools can use to measure the educational progress of their children relative to all other children in the US at the same age and class level.

being ELLs. Since the dataset is very large and includes all children (in the respective grades) for an entire US state having a large ELL population, any observed differences in performance have to be taken seriously.

Table C.1. Disparity Indices from Site 2 for Grades 2 and 7 for mean performance on the Stanford 9.

Difference Index (DI)	*Reading*	*Math*	*Language*	*Spelling*	*Average (all language related)*	*Difference*
Grade 2						
ELL/Non-ELL	55.8	33.5	60.2	42.8	52.9	19.4
Free/reduced lunch	32.7	25.1	35.2	25.3	31.3	6.2
Parent education	106.3	84.9	118.5	87.5	104.1	19.2
Grade 7						
ELL/Non-ELL	96.9	50.4	70.7	81.1	82.9	32.5
Free/reduced lunch	47.2	29.5	32.9	31.1	37.1	7.6
Parent Education	98.4	76.2	79.0	80.5	86.0	9.8

The numbers in Table 1 are expressed in terms of a *Disparity Index*. In this case, the Disparity Index (DI) is equal to the extent to which the non-ELLs performed above the ELL group (expressed as a percentage gain or advantage)[17]. For example the DI for Grade 2 Reading between ELLs and non-ELLs is 55.8. This number does not mean that the scores of one group averaged a total of 55.8 percentage points above the second group. Rather, it means that the mean score of the non-ELLs (native speakers of English) was 55.8 percent higher than ELLs on the Reading subtest of the Stanford 9. Similarly, on the math subtest of the Stanford 9, non-ELLs scored 33.5 percent higher than did the ELLs.

In his analysis of the data, Abedi looked closely at three different variables—the first language of the test takers, the socioeconomic status of the test takers, and the educational level of the parents of the test takers. By now, it has been well established that parental education and socioeconomic status of the family of the test taker play a significant role in educational outcomes. In fact many educators had previously ascribed the observed mean low performance of non-native speakers of English on such tests to precisely these two variables. Abedi's objective was to attempt to determine, from the data, whether language was also a factor or maybe even a greater factor than parental education and socioeconomic status on students' scores. Abedi reasoned that the best way to look for evidence of a possible language effect in test results was to compare the performance of test takers on math versus all language-related test tasks. Math, he reasoned, was less dependent on language proficiency and thus would be less impacted by the fact that some takers did not speak English (the language of the test) as their first language.

He then made three two-way divisions of the test takers to observe the impact of each of these divisions on performance (on math versus on language-related tasks). These three divisions were as reflected in the table—ELLs versus nonELLs, children from high socioeconomic backgrounds versus those from low socioeconomic backgrounds, and children who parents were highly educated versus those whose parents were not highly educated. The data in the table capture the results of these comparisons. So, for example, the Class 2 children having highly educated parents scored 106 percent higher on the reading subtask of the Stanford 9 than did the children having poorly educated parents. Children from high socioeconomic backgrounds scored 32.7 percent higher on the reading subtask than did children from low socioeconomic backgrounds and so on.

[17] The actual formula for computing the DI is as follows: Mean Score for non-ELLs minus Mean Score for ELLs divided by the Mean Score for ELLs. The DI is not the simple difference between the two, but the relative or percentage advantage of one score over the other.

The column ‘Average (all language-related)’ is the arithmetic average of the three language-related columns, Reading, Language, and Spelling. The column ‘Difference’ is the simple differential between performance on the math subtest and mean performance on the language-related subtests. It is this 'Difference' column that Abedi was most interested in. What he found was that, for Class 2 children, the two variables, parental education and first language had the most impact on educational outcomes with socioeconomic status having the least impact. However, at the level of Class 7, there was a very large differential between ELLs and non-ELLs on math versus language-related subtasks of 32.5 percentage points. Furthermore, ELLs had fallen significantly behind native speakers on BOTH of these subtasks.

On the basis of these results Abedi concluded that the language of testing does constitute a source of bias as an assessment of what students have learned. At the same time, it should also be noted that language of instruction appears to have had a major impact on learner performance as the ELLs appeared to be falling further and further behind in school the longer they stayed in school (a disparity index of 82.9 percent in Class 7 on language tasks versus 52.9 percent in Class 2 and a disparity index of 50.2 percent in Class 7 on the math task versus 33.5 percent in Class 2).

In the Standard education model in the Kom area, Kom-speaking children are being educated in English and tested in English both by the KEP project and by the Ministry of Education. At the same time, we must also point out that there is a major conceptual difference in the Kom situation compared to that analyzed by Abedi. In the Kom case, ALL of the children speak the same language-Kom. Some of these are being instructed in English and tested in English while others are being instructed in Kom and tested in Kom (their first language). Given the findings reported above, we were curious to see whether a similar effect might be observed in the Kom data despite the obvious differences in educational design between Abedi's research data and that of the KEP project. Table C.2 presents the relevant data.

Table C.2. Disparity indices comparing performance in language-related versus mathematics subjects in the KEP project.

Disparity Index	Grammar	Reading Comprehension	All Language Related	Mathematics	Difference
Grade 1					
English/Kom	190.2	582.3	277.2	279.8	-2.6
Grade 2					
English/Kom	71.2	134.3	96.6	116.4	-19.8

Keeping in mind that the KEP project does not have the same educational design as that reported on by Abedi, what do we learn from the data? Our first major impression has to be that, in the case of the KEP program, the main effect to be observed is not an issue of language of testing, but language of instruction. The evidence is of two sorts. First, we note the very large disparity indices which exist especially in Class 1 but carry over to Class 2. Second, we note that the disparity index for mathematics is actually greater, not smaller in both Classes as would be expected by Abedi's logic. Both of these observations tend to suggest that the pattern we are observing is caused by instruction and not by language of testing.

The fact that the difference between the disparity indices for math versus all language related measures gets larger between Class 1 and Class 2 is very suggestive in terms of the theory underlying mother tongue education[18]. However, we need to be careful of over-

[18] The theory holds that one of the major benefits of mother tongue education lies precisely in the superior ability of the teacher to convey more difficult or more abstract concepts in the language of the child. Many of the

interpreting this trend since only two years' worth of data are available for analysis. If the size of this difference continues to grow consistently from Class to Class, we will, indeed have found yet another source of support for the theory of mother tongue education.

procedures and concepts of mathematics—even at the level of basic education—are not encountered or directly observable in daily live so must be explained verbally. If neither the teacher nor the student controls well the language of education, these explanations are too often ineffective.

Appendix D

The Schools Participating in the KEP Innovation

KEP Schools (Kom-medium instruction)

CBC Belo
CBC Fujua
CBC Kikfuini *
GS Ameng
GS Bolem
GS Ilung
GS Kitchu
GS Laikom
GS Mboh
GS Muteff
GS Ngwah
GS Wombong

* To be replaced in 2009-2010 because it is failing.

Standard Schools (English-medium instruction)

CS Kindoh
CS Wombong
CBC Abuh/Wainchia **
CBC Fundeng
GS Atondum
GS Baichu
GS Fundong Village
GS Mentang
GS Njinikijem
GS Yuwi
GS Meli
PS Ngwah

** CBC Abuh lost most of its students during the year and was unable to field a teacher so was replaced by another school as a comparison school.

GS – government school
CS – Catholic school
PS – Presbyterian school
CBC – Cameroon Baptist Convention

Appendix E

Model Fit (or "What combination of Variables best explains the results?")

Model Fit (accounting for variation in the data)

Researchers use a variety of strategies for presenting the results of their findings. One of these is known as 'model fit.' One can think of *model fit* as somewhat like putting together the pieces of a jigsaw puzzle. In the case of this kind of educational data, however, this means trying out different combinations of variables so see which ones best explain the outcome observed as a result of the testing. In doing this, it is necessary to point out ahead of time that we know that not all variables affecting student performance have been accounted for. For example, we know that some children are brighter than others. Our data does not include information on this variable so it will remain a source of unexplained variation. Another variable known to impact educational performance is that of education of the parents. We did not collect data on this variable so it will also be a source of unexplained variation in the data. Other variables commonly explored to explain outcomes include class size, instructional methodologies, funding support, socioeconomic status, etc. The only one of those on which we have data is class size.

The variables about which we do have information are the following: program (Standard or KEP), school attended by the children, age of the children, gender of the children, and class size. Which of these and in what combination best explains the outcomes we have observed?

Class 1 2009

Not surprisingly, the variable of 'Program' (whether or not a student is in the KEP) program is the largest single variable accounting for 33 percent of the total variation in the data. Since every child, in effect, gains or looses and average of 26 points on the year-end assessment, the variable of program is going to be large.

Because 'Program' is such a major source of variation AND because all of the schools were either in one program on in the other program, the remainder of the analysis of model fit for Class 1 is going to be done and presented by program.

The Standard Program. In the Standard program, the best analytical model is that of 'School' plus 'Age' plus 'Class size.' Taken independently, 'School' accounts for 27.5 percent of variation in performance, 'Age' accounts for 3.4 percent of variation in performance, and 'Class size' accounts for 3.1 percent of variation. Not surprisingly, then, the best overall model is one which includes these three in the order mentioned in the beginning of the paragraph. It is additionally instructive to look at the best-fit regression model:

```
Overall_Percent = - 17.2 + 0.962 School_Per + 3.65 Age - 0.102 Class_Size.
```

The coefficient for 'School_Per'(formance) is .962 indicating that a child's performance within the Standard program is very heavily dependent upon the school he or she attends. Mean scores for the schools participating in the Standard program ranged from a low of 10.3 percent (school-wide average) to 37.6 percent. With this amount of variation, a child's performance is heavily dependent on the school of attendance.

The coefficient of 3.65 for 'Age' tells us that, all other things being equal, a child's performance is going to increase (or decrease) 3.65 percentage points for each year of age. Since age 6 is the designated entry age for Class 1, a child's performance is going to decrease by 3.65 percentage points for each year the child is younger than 6. For example, if the

average performance of 6 year old children was 20.9, we can say that the estimated level of performance for a 5 year old will be 17.25 and for a 4 year old, 13.60. The model clearly supports the assertion that children are not developmentally ready for school before the age of six[19].

The coefficient of .102 for 'Class size' tells us that class performance is going to decrease 1 percentage point for each additional 10 children in the classroom. Since the average class size was 40 children, the implication is that a school with 60 children in a class can expect to see a two percentage point reduction in mean performance solely because of class size. Class size clearly has less impact than Age or School attended. Still, it appears, there is a measurable price to pay for having large class sizes.

The KEP Program. Not too surprisingly, the optimal model for the KEP program is similar to that of the Standard program. In the case of the KEP program, the three variables of School, Age, and Class size account for 25.8, 11.2, and 1.9 percent of data variance respectively. When put together in a single model, the three variables account for 37.1 percent of total variance with the following regression equation:

```
Overall_Percent = - 35.4 + 0.976 School_Per + 6.59 Age - 0.0585 Class_Size
```

As we saw earlier, the coefficient for School is close to 1 indicating that the school attended plays a very heavy role in determining or predicting a child's performance. From a low of 31.6 to a high of 72.6, mean performance by school was dramatic.

Of greater interest is the coefficient for Age of 6.59. This coefficient indicates that performance goes up or down by this amount for each year of age per child. A child two years underage would be predicted to score more than 13 points below the mean for a properly aged child, again underscoring the impact of having so many under-age children in classrooms in the Kom area. Since instruction in the KEP program is in Kom so that all children understand what the teacher is saying, it is likely that this coefficient (6.59) is the real or true coefficient (or penalty) for being under-aged.

The coefficient for Class Size is only .058—only about half of that in the Standard program. The implication is that large class size has a slightly less of a negative impact on performance when the local language is the language of instruction.

Class 2 2009

It is appropriate as well to examine possible model fits for the Class 2 data from this year. It turns out that the best model is not too different from that that found for Class 1 though the weighting of the variables changes somewhat.

The variable of 'Program' accounts for less variation of the data in Class 2 compared to Class 1 dropping to only 26.93 percent of variance. This is not surprising given that the mean scores of students in the two programs were closer together in Class 2 than in Class 1.

The Standard Program. Once we separate the programs and look at the variables of School, Age, and Class Size for the English program, we find that these account for 30.0, 4.9, and 3.6 percent respectively in the variation of the data. When placed in a single model, the three variables together account for 33 percent of the variation in the data. The resultant regression equation is as follows:

[19] As is often the cases with statistical analysis, it is almost certainly the case that the coefficient and the data for Age here under-represents the 'true impact' of Age because overall performance is so low on the part of those of normal age. The following section on the KEP program gives a more realistic picture of the consequences of the Age variable.

```
Overall_Percent = - 12.7 + 0.926 School_Per + 2.32 Age - 0.0567 Class_size
```

The coefficient for School_Per(formance) of .926 is still close to 1 but slightly smaller than the analogous coefficients seen for Class 1. The implication is that School of attendance is slightly less dominant in predicting student performance. The coefficient of 2.32 for age indicates that performance increases or decreases this amount for each year of age (older or younger) and the coefficient of .0567 for class size indicates that performance suffers approximately one percentage point for each additional 18 students added to class size—a measurable but not very large factor in student performance.

The KEP Program.

The best model for the KEP program is similar to that of the Standard program. The three variables of School_Per(formance), Age, and Class Size, when modeled separately, account for 23.2, 2.1, and 0.7 percent of variance respectively. When combined into a single model, the three variables together account for 25.9 percent of variation. The resultant regression equation is as follows:

```
Overall_Percent = - 13.5 + 0.995 School_Per + 2.06 Age - 0.0671 Class_size
```

In contrast to the Standard program, the variable of School_Per(formance) has a coefficient almost equal to 1 (.995) indicating a close relationship between school performance and student performance. The coefficient of 2.06 for Age is slightly lower than see in the Standard program but has the same meaning. The coefficient for Class_Size is very slightly larger probably being a simple artifact of the data set.

Conclusion

The dominant variable impacting children's performance in Class 1 (outside of Program) is that of School attended and this is true for both programs. The second largest variable is that of Age and the smallest is that of Class size. (Gender has not been discussed because it was found to have essentially no impact at all.) This analysis makes clear the steps that could be taken to improve student learning outcomes: (1) use the local language as a language of instruction, (2) improve delivery of instruction at the classroom level, (3) deal with the age problem, and lastly (4) worry about class size. Since this latter variable has minimal impact, it is the least important to consider at this point.

Unexplained variation in children's scores is due to such factors as intelligence, variation in how well children understood test items, attitude of children towards the test, etc. With sufficient resources and tools, one can measure these additional explanatory variables. In the case of this research project, however, no effort was made to do so since it was assumed these are probably equally distributed among the students participating in the two programs.