Final Report

2004-2005 Evaluation

India dot-EDU Project

DATE: June 17, 2005

Prepared by:

James M. Royer, PhD
University of Massachusetts, Amherst

Executive Summary

Goals

The goals of the student learning evaluation were to:

- Determine if Karnataka students enrolled in schools receiving an IRI instructional intervention performed better on math, science and social studies tests than did the students enrolled in control schools
- Determine if Chhattisgarh students enrolled in schools receiving Interactive Radio Instruction (IRI) performed better on tests that measured English language comprehension and speaking than did student enrolled in control schools.

Brief Project Description

The India dot-EDU intervention evaluated in this study consists of distance education in three Indian states; Karnataka, Chhattisgarh and Jharkland. In Karnataka the intervention consists of instruction in math, science and social studies that is delivered to grade 4 and 5 students via radio instruction (IRI). Pretest, posttest and pretest to posttest gain of students enrolled in schools receiving the IRI treatments was compared to performance of students enrolled in comparable schools not receiving the distance education treatments (controls).

In Chhattisgarh, the distance education treatment consists of English language instruction that is delivered to grade 1 and 2 students via radio (Interactive Radio Instruction). Pretest, posttest and pretest to posttest gain of students listening to the IRI instruction was compared to pretest, posttest and pretest to posttest gain of control students enrolled in schools that did not receive the IRI instruction.

Test development. Student learning was measured by tests that were developed via a process that began with a series of meetings in July 2003 and that involved the local convening of subject matter experts in workshops that were conducted in January of 2004. These workshops produced draft tests in math, science, and social studies to be used in Karnataka and English language tests to be used in Chhattisgarh. These draft tests were then finalized in interactions between local experts and project staff members. The Karnataka tests were subsequently translated into Kannada. The Chhattisgarh tests were tape recorded in several versions using local speakers with accents that were familiar to students in different geographical regions.

Another important part of the testing process was the training of test administrators and the piloting of the tests used to assess student learning. This phase of the project occurred in June 2004 in workshops conducted in Bangalore and Raipur.
These workshops had the dual purpose of developing skilled test administrators and developing final versions of the tests, and final versions of the instructions that accompanied the tests.

School sampling. The selection of schools to participate as treatment and control schools was roughly a stratified random sampling plan. The process initially involved the identification of variables that were likely to influence student learning. This process resulted in the identification of the following attributes of schools that were likely to influence student learning performance:

- Whether the school was a lower primary school (LPS) or a higher primary school (HPS). The distinction involves the number of grades contained in a school but also is an indicator of school size with HPS schools enrolling more students and having more teachers.
- School location. Whether a school was located in an urban area, a rural area, or a tribal area was thought to be a factor in student achievement.
- Geographical region. The location of a school in each of the states was thought to be a factor would influence student performance.

Lists of schools were generated in Karnataka, and Chhattisgarh and the factors described above were then used to select treatment and control schools that were thought to be comparable to one another and matched with respect to the above stratification variables.

In addition to coding the above attributes in the data files, a number of characteristics of individual students were also coded in the files. These characteristics included:

- Sex of Student
- Father's Education Level
- Mother's Education Level
- Grade
- Caste

Test Administration

The pretests were administered in July 2004 and an interim test was administered in December 2004. The posttests were administered in April and May of 2005.

In Karnataka:

- Pretests were administered to 2,214 students
- Midtests were administered to a sample of 1,110 students who took the prestest.
- Posttests were administered to a total of 3,844 students. 2083 of these students had also taken the pretest.

In Chhattisgarh:

- Pretests were administered to 831 students.
- Midtests were administered to a sample of 335 students who had taken the pretest.
- Posttests were administered to 764 students, 567 of whom had also taken the pretest.

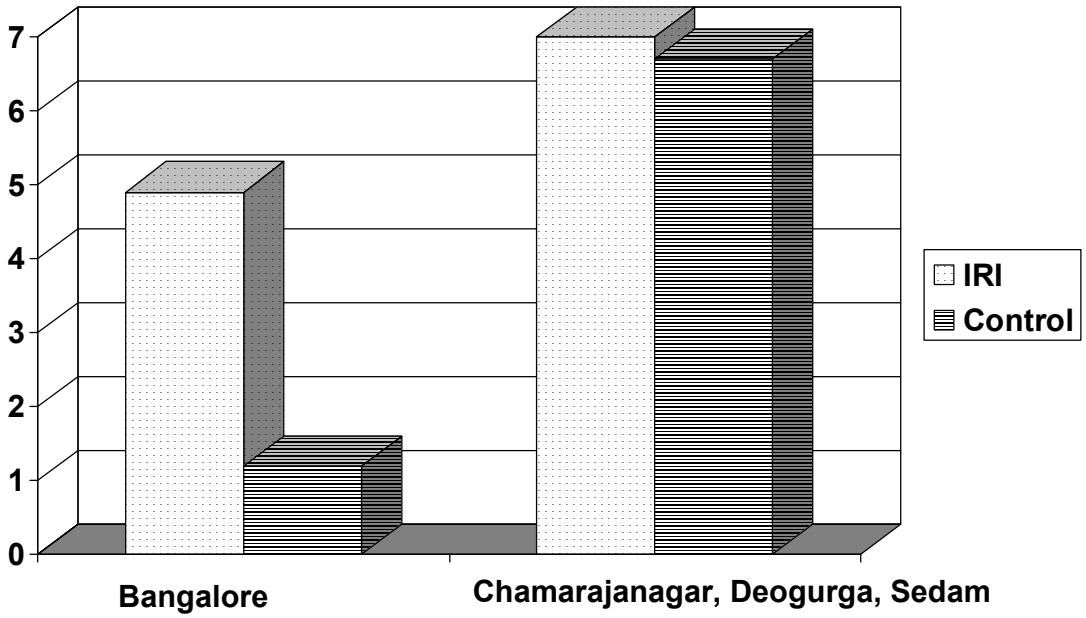
Results

The results reported in this document focuses on treatment situations that were well implemented during the project year. Results will not be reported for the Video and the IRI + Video interventions in Karnataka because those interventions were not well implemented. In addition, results for the Jharkland part of the project will not be reported. Data for poorly implemented treatments is coded in data files and is available upon request to project management. However, it will not be described further in this report.

The complete report to follow this executive summary presents detailed results broken down by treatment condition and by many of the other variables that were coded into the data files. The executive summary will focus on outcomes associated with treatment condition and outcomes associated with student sex, a variable of considerable interest in this and other educational intervention studies.

Karnataka results. The graph below shows the amount of gain occurring from pretest to posttest for students in Karnataka. The project staff indicated that there were differences in the quality of treatment delivery between Bangalore and the remaining three regions (Chamarajangar, Deogurga and Sedam). Specifically, the the later three regions there were some indications of poor signal quality and late or no delivery of teacher guides. This resulted in the decision to perform separate data analyses for Bangalore and the remaining three regions of Karnataka. The Pretest to posttest gains for the Karnataka regions are reported in the graph below.

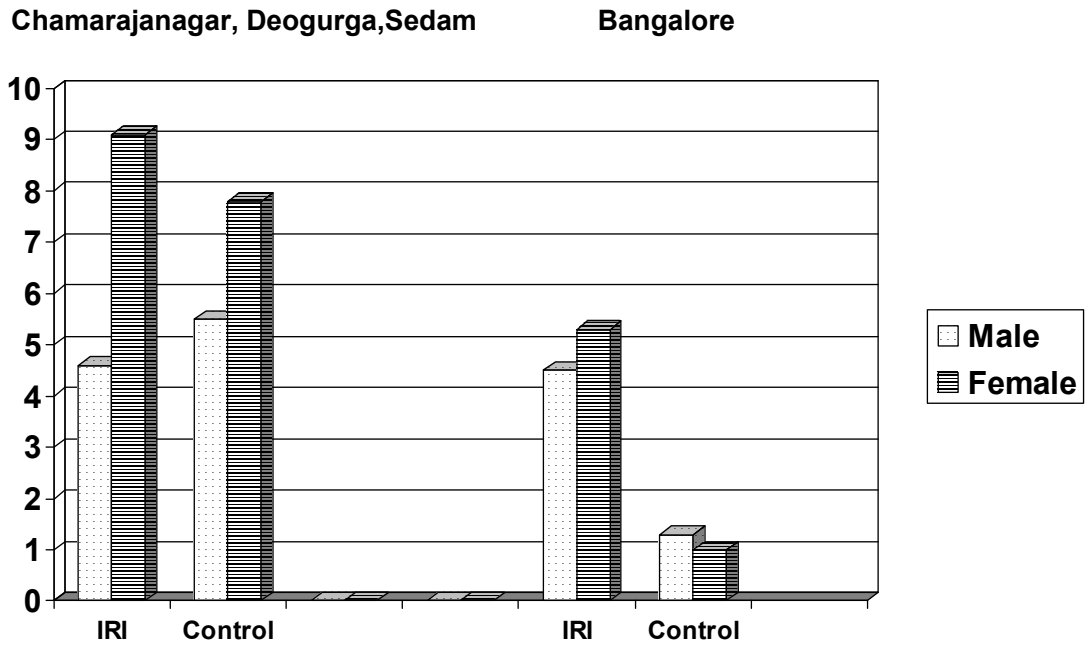
Pretest to Posttest Percent Correct Gain in Karnataka for IRI and Control Groups



The data depicted in the above graph was subjected to a statistical analysis that asked the question of whether the gains made by the IRI group were statistically superior to the gains made by the control group. The analysis indicated that the IRI group gained significantly more than the control group in Bangalore. However, there were no statistical differences in pretest to posttest gain for students from Chamarajangar, Deogurga and Sedam.

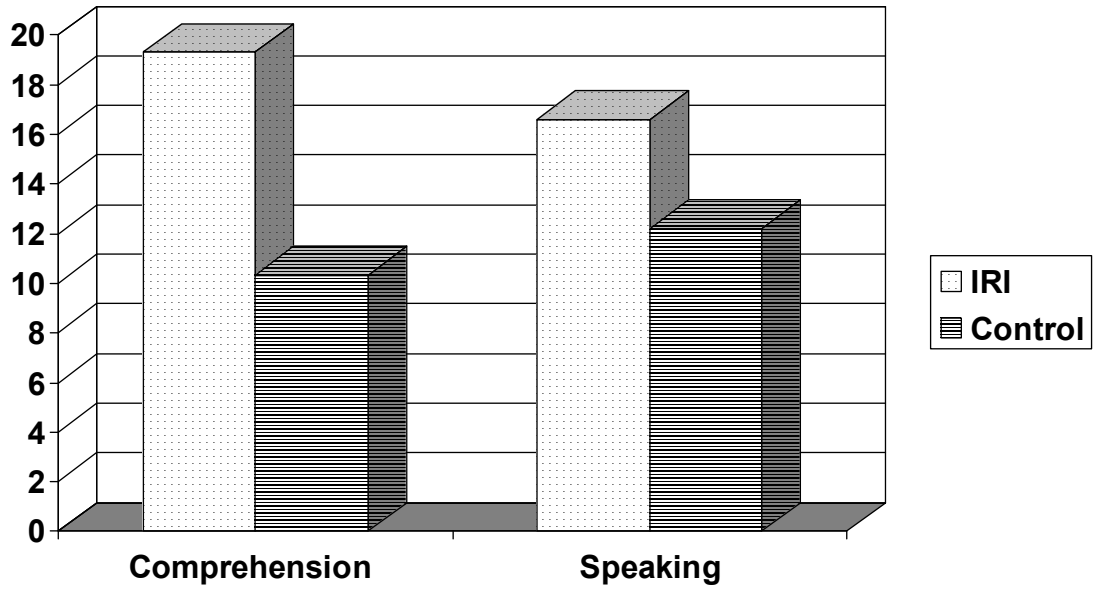
The next graph depicts the percent correct gain from pretest to posttest for girls and boys in the two treatment conditions. The graph shows that girls make more gain than boys in both the IRI group in both Bangalore and the other three regions in Karnataka. Girls also make greater gains than boys in the control condition in Chamarajangar, Deogurga and Sedam. However, there was little difference between the performance of boys and girls in the control condition in Bangalore. A statistical analysis of this data indicated that the girls made significantly greater gain than boys overall, and there was evidence that student sex impacted on performance differently in Bangalore than it did in the three other regions of Karnataka.

Pretest to Posttest Percent Correct Gain for Karnataka Males and Females as a Function of Intervention Condition

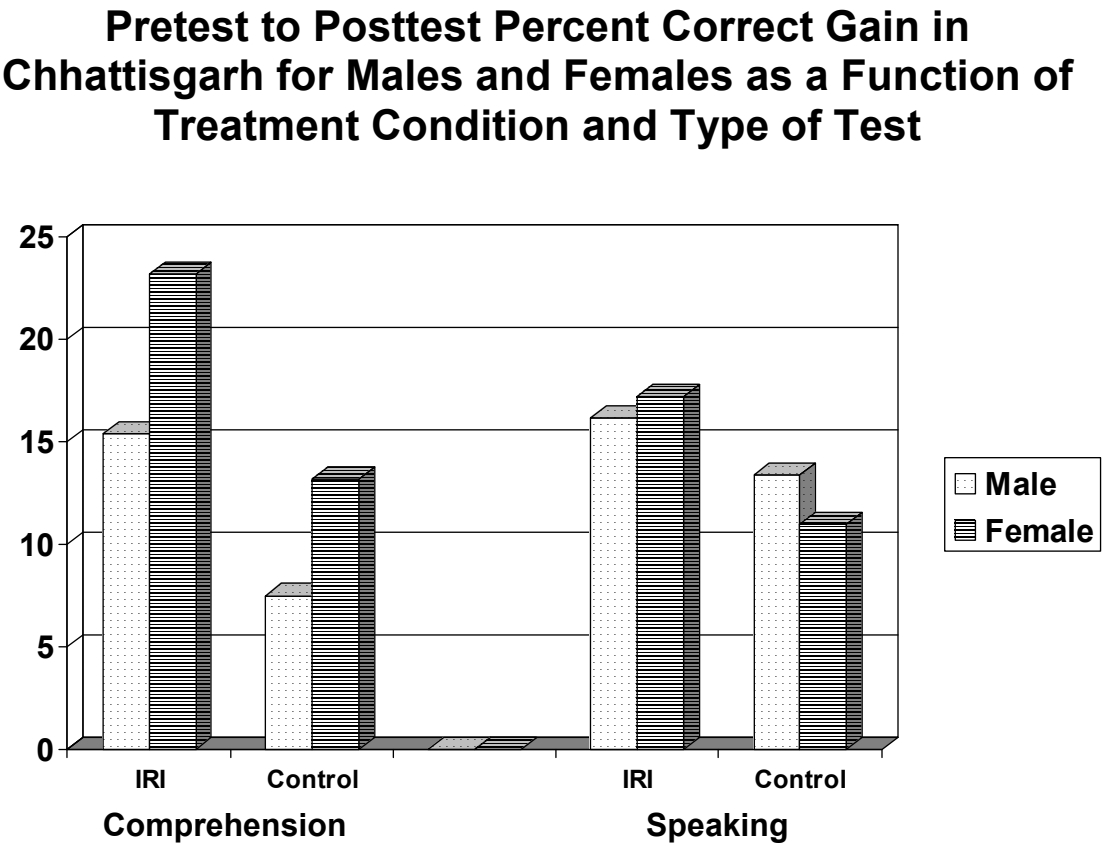


Chhattisgarh results. The tests administered in Chhattisgarh assessed student ability to listen to and understand English (comprehension), and the ability to speak in English (speaking) when asked a question and directed to respond in English. The graph below depicts the pretest to posttest gain in both comprehension performance and speaking performance as a function of treatment group. A statistical analysis indicated that the IRI group made significantly greater gains than the control group on both the comprehension test and the speaking test.

Pretest to Posttest Percent Correct Gain in Chhattisgarh for IRI and Control Groups



The gains on the English language tests were also examined for boys and girls. The graph below depicts the performance of boys and girls as a function of type of test (comprehension, speaking) and treatment condition. The graph shows that girls perform somewhat better than boys on the comprehension tests, but there is little difference between the two sexes on the speaking test. Statistical analyses of this data indicated that there were no sex differences on either the comprehension or the speaking test, and there were no interactions between student sex and treatment condition.



Summary of Major Results

The analyses of the Karnataka data indicated that there were no intervention treatment differences in the Chamarajangar, Deogurga and Sedam districts. There was, however, a significant advantage for IRI students in the Bangalore region. Specifically, the IRI students in Bangalore made significantly greater pretest to posttest percent correct gains than did control students. In addition, analyses reported in the main section of this report show that Bangalore IRI students made significantly greater gains than their control counterparts on each of the subject matter areas represented in the test; math, science and social studies.

The analyses examining the relative performance of boys and girls in Karnataka showed that girls made larger overall pretest to posttest percent correct gains than did boys. However, formal analyses of this data indicated that the differences were not statistically significant.

The analyses of the Chhattisgarh data show that the students listening to the IRI broadcasts made significantly greater pretest to posttest percent correct gains than their control counterparts on both the comprehension and speaking tests. An examination of the relative performance of boys and girls on these tests showed that there were mean

differences favoring girls on the comprehension tests, but these differences were not statistically significant. The differences between boys and girls on the speaking tests were quite small.

Interpretive Comments

The student learning evidence strongly supports the interpretation that Chhattisgarh students listening to English Language IRI programs perform better on comprehension and speaking testes than do control students not listening to the program. The trend showing that the impact is somewhat stronger in comprehension than it is in speaking makes sense in terms of what we know about language acquisition. Receptive competence in a language always precedes productive competence.

The Karnataka the evidence is mixed, though there appears to be reasonable explanations for the mixed results. First a review of the findings. Students from the Chamarajangar, Deogurga and Sedam regions of Karnataka who listened to IRI broadcasts made slightly more pretest to posttest gain in science and social studies than did control students. These differences were not statistically significant and in math control students made greater gains than did IRI students. The situation was different in Bangalore where students listening to the IRI broadcasts made significantly greater overall pretest to posttest gain than did control students, and the gains they made were consistent across subject matter areas. That is, IRI students significantly out-gained control student in math, science and social studies.

Now for some background with respect to the way the programs were implemented and with respect to the way testing was conducted. In some areas of Chamarajangar, Deogurga and Sedam reception was poor and some of the IRI schools were not receiving broadcasts for several months of the intervention period. In addition, the IRI broadcasts were designed to be broadcast successively within an instructional module. For instance, a particular science module might consist of 4 programs that were supposed to be broadcast in succession, and this science sequence might be followed by a 3 program sequence in math. In some areas of Chamarajangar, Deogurga and Sedam the modules were broken up so that in 3 successive broadcasts there would be a math program, a science program and a social studies program. This breaking up of the modular properties of the programs greatly reduced the effectiveness of the programs as an instructional procedure. Finally, the IRI treatment was supposed to have a teacher handbook that advised the teacher about how to take best advantage of the IRI programs. This handbook was not ready for delivery until November, well after the start of program broadcasts in July.

All of the negative factors mentioned above were not present in Bangalore. Teachers had handbooks, programs were broadcast in sequence, and signal strength was strong and consistent. What did happen in Bangalore, however, is that broadcasts and pretesting started in November rather than in July. This meant that there was a shorter interval between pretesting and posttesting in Bangalore than in the other 3 Karnataka regions. This in turn probably resulted in a lesser amount of overall gain in Bangalore than in the

other three regions. However, the important point is that in the location where program implementation occurred as it was supposed to occur (in Bangalore) students listening to IRI programs made greater gains than did students enrolled in control schools. In contrast, in situations where program implementation was not optimal (Chamarajangar, Deogurga and Sedam) students listening to IRI broadcasts displayed no learning advantages relative to control students.

2004-2005 Evaluation of Student Learning

India dot-EDU Project

Goals

The overall goal for the India dot-EDU evaluation is to determine if students enrolled in schools receiving a distance education treatment make greater academic gains than do students enrolled in schools not receiving the treatments (controls). The evaluation effort involved an examination of performance on three tests. A pretest administered in July of 2004, an interim test (midtest) administered in December of 2004, and a posttest administered in April and May of 2005. This report describes the activities associated with developing and administering the tests and it describes the learning outcomes on the tests with a particular emphasis on the answering the question of whether there are learning differences between students receiving the distance education treatments and students serving as controls.

Project Description

The dot-EDU intervention consists of a distance education intervention targeted at differing academic skills in three states. In Karnataka, radio (IRI), video, and a combination of radio and video instructional treatments provided instruction in mathematics, social studies and health-related content (science) to Standard 4 and 5 students. The learning performance of students receiving this instruction was compared to the learning performance of comparable students who did not listen to or watch the distance learning programs.

As noted above, some students in Karnataka received a treatment consisting of watching video instruction and some received a treatment consisting of IRI plus the video instruction. Project staff indicated that the delivery of video instruction during the first year of the project was problematic and did not really constitute a valid intervention treatment. Hence the video and IRI + video treatments will not be evaluated in this report. The data for the video and IRI + video is, however, available upon request to the EDC project staff.

In Chhattisgarh, grade 1 and grade 2 students listened to IRI programs that were designed to improve English language literacy. Specifically, the programs were designed to enhance the correct speaking of English and to enhance the ability of students to understand English that they are listening to. Again, success of the programs in achieving project goals is determined by comparing the English language performance of students who listen to the radio programs with the English language performance of comparable students who do not listen to the programs.

A Jharkland project was also part of the original dot-EDU program. Pretests and posttests were administered in Jharkland, but the implementation of the IRI intervention did not occur optimally, and the data from Jharkland will not be examined in this evaluation report.

Whereas the primary purpose of the evaluation effort was to determine whether IRI students made greater learning gains than control students, there were also secondary goals in the evaluation effort. These secondary goals involved determining whether the impact of the distance education interventions varied as a function of characteristics of children and schools. The evaluation design allowed an examination of the learning performance of participating children as a function of the following factors:

- Treatment Condition (IRI or control)
- Sex of Student
- Type of School (LPS or HPS)
- School
- School Setting (Rural, Urban, Tribal)
- Father's Education Level
- Mother's Education Level
- Grade
- Caste

In addition to examining the impact of each of the above factors, where possible, the evaluation effort also examined the interaction between geographical factors, demographic factors, and instructional treatment conditions.

Test Development

The development of the learning tests occurred in phased activities that began with a series of meetings in Bangalore, Mysore and the Sedam region in July 2003. At that time the author of this report (Royer) met with Professor Phalachandra and a number of teachers with content area expertise in the areas to be targeted by the Karnataka programs. The Karnataka meetings in particular focused on discussions about the "hard spots" in the curriculum. The hard spots that had been identified in earlier work played a central role in the subsequent development of both the instructional programs and the learning tests. This series of meetings resulted in a rough description of the content to be tested and the development of a tentative time line for completing the student learning evaluation component of the project.

The next phase of the test development effort occurred in January 2004 when Royer and Kimberly Parekh from EDC traveled to India to conduct two test development workshops, one in Bangalore, and one in Raipur. Both workshops were conducted by Royer, Parekh, and Phalachandra, and the participants were content area teachers or specialists from the Bangalore and Raipur areas. The goal of these workshops was to develop potential test items that could be included on learning tests. The materials available to workshop participants included textbooks in use in local schools and

curriculum descriptions that described instructional goals to be attained in the targeted grades. The participants also had available the master plan of the distance education programs that were under development.

Following the January 2004 workshops Parekh and Royer returned to the U.S. and proceeded to develop rough drafts of tests to be used in the evaluation effort. The development of the rough drafts involved a considerable amount of interaction between project participants in the U.S. and India and the tests went through a number of drafts during the process. The Karnataka pretest that was produced from this process consisted on 13 test items in math,10 in science, and 11 in social studies for a total of 34 items. The Chhattisgarh English language tests ultimately consisted on 8 items that measured comprehension competence and 6 items that measured the ability to speak in English.

The posttests were developed to mirror the content of the pretests but with different test items. The posttests finalized after a review of pretest performance and an examination of performance of the previously mentioned midtest.

Training Test Administrators and Pilot Testing

The next phase of the test development process occurred in June 2004. Royer again traveled to India and he, Lindsay Crinklaw of EDC, and Professor Phalachandra trained test administrators for both Karnataka and Chhattisgarh in workshops held in Bangalore and Raipur. In addition to training test administrators, the workshops had the goal of collecting pilot data on the learning tests. These workshops took place over a week in both Bangalore and Raipur and consisted of three days of training in test development, one day of actually administering the tests in local schools, and a final day of debriefing and examination of the pilot test results.

Additional changes in the tests were made after examining the pilot test results and Professor Phalachandra then proceeded to develop final versions of the tests. This involved getting the tests used in Karnataka translated into Kannada, and it entailed making tape recorded versions (by local speakers) of the comprehension and speaking tests to be used in Chhattisgarh.

Sampling of Schools to Participate in the Evaluation Effort

Sampling Considerations

Several factors were considered when selecting the schools that participated in the evaluation effort:

- The primary goal of the evaluation effort which is to determine if students receiving the dot-EDU treatment made greater educational gains than students enrolled in control schools.

- The inclusion of other factors in the analysis plan that were thought to be related to student performance. These factors include type of school (Lower Primary Schools or Higher Primary Schools) and school location (rural, urban or tribal)
- The evaluation budget. The number of schools sampled during the evaluation effort was constrained by the funds that were available to conduct the evaluation.
- Distribution of treatment and control schools. Where possible, an attempt was made to balance out treatment conditions and the "additional factors" such as, school type and school location.

Sampling Process

The sampling process was carried out with the goal in mind to balance out the demographic and treatment factors in the evaluation design to the extent it was possible to balance them.

Karnataka

Table 1 below shows the sampling distribution for Karnataka. A total of 2,214 students from 49 schools were pretested in Karnataka and the breakdown of the location, type of school, and treatment condition for pretested students in presented in Table 1.

Table1
Number of Schools and Number of Students in Karnataka Broken Down By the Demographic and Intervention Conditions in the Evaluation Study

	School Location					
	Rural		Urban		Tribal	
School Type	LPS	HPS	LPS	HPS	LPS	HPS
Region						
Chamarajanagar						
IRI	Schools =1 Students=24	Schools=1 Students=79	Schools = 1 Students=8	Schools =1 Students=30		
Control	Schools=1 Students=14	Schools=1 Students=28	Schools=2 Students=52	Schools=2 Students=106	Schools=1 Students=11	
	Totals for Chanarajangar: 11 Schools, 352 Students					
Deogurga						
IRI	Schools=1 Students=28	Schools=1 Students=58		Schools=1 Students=98	Schools=1 Students=37	
Control	Schools=1 Students=45	Schools=1 Students=31	Schools=2 Students=53	Schools=1 Students=75	Schools=1 Students=33	Schools=1 Students=55
	Totals for Deogurga: 11 Schools, 513 Students					
Sedam						
IRI	Schools=1 Students=26	Schools=1 Students=78	Schools=1 Students=57	Schools=1 Students=64		
Control	Schools=3 Students=72	Schools=2 Students=130	Schools=2 Students=48			
	Totals for Sedam: 11 Schools, 475 Students					
Bangalore						
IRI		Schools=4 Students=185	Schools=1 Students=4	Schools=5 Students=279		
Control		Schools=2 Students=144		Schools=4 Students=395		
	Totals for Bangalore: 16 Schools, 1007 Students					

Chhattisgarh

The sampling plan for Chhattisgarh was based on similar concerns as used to create the sample for Karnataka, though there were some differences. For example, the distinction between LPS and HPS schools was not maintained and it was not possible to sample tribal schools in Chhattisgarh. The sample for the 786 children participating in the pretest in Chhattisgarh is presented in Table 2 below. It should be noted that the number of students tested in Chhattisgarh is considerably less than in Karnataka because testing in Chhattisgarh involved administering tests to groups of five children. In Karnataka the tests could be administered to entire classrooms of children.

Table2
Number of Schools and Number of Students in Chattisgarh Broken Down By the Demographic and Intervention Conditions in the Evaluation Study

	Rural	Urban
Region		
Abhanpur		
IRI	Schools=7 Students=129	Schools=1 Students=15
Control	Schools=6 Students=105	Schools=1 Students=19
Totals for Abhanpur: Schools = 15, Students =268		
Kanker		
IRI	Schools=8 Students=160	Schools =2 Students=40
Control	Schools=4 Students=80	Schools=1 Students=20
Totals for Kanker: Schools = 15, Students = 300		
Kondegaon		
IRI	Schools=7 Students=111	School=1 Students=20
Control	Schools=5 Students=87	
Totals for Kondegaon: Schools = 13, Students = 218		

Test Scoring

The Karnataka pretest test contained 34 test items. Items were scored as either a 1 (correct) or 0 (incorrect) and then transformed into percent correct scores, which will be reported in the result section to follow.

The Karnataka posttest contained 25 items and was scored in the same way as the pretest.

The comprehension and speaking test items used in Chhattisgarh contained 8 comprehension items and 6 speaking items. The test administration procedure involved working with five children at a time. In the comprehension test, comprehension question number 1 would be asked to child number 1, comprehension item number 2 would be asked to child number 2, and so on until comprehension item number 6 would

then come back to child number 1 to answer.. A similar procedure was followed with the speaking items where child number 1 would be asked to answer speaking item number 1, and speaking item number 6 would come back to child 1 to answer. Test scoring generally consisted of giving a 1 for a correct response or a 0 for an incorrect response, though in some cases partial credit or additional credit was given for answers. As was the case with the Karnataka tests, scores are reported in terms of percent correct.

Test were scored in India and item level data (1=correct, 0=incorrect) was entered into Excel files that were then transmitted to Royer who transformed the Excel files into SPSS files and performed statistical analyses of the data.

Pretest, Posttest, and Pretest to Posttest Gains for Karnataka

Test Reliability

The reliability of the pretests and posttests used in Karnataka was examined using the Chronbach alpha procedure. The reliabilities of the tests are reported in Table 4 below. The reliabilities are less than desirable but high reliabilities are unlikely in situations where many of the students are performing at chance level. As will be seen in the sections to follow, many of the Karnataka students performed at chance or below on both of the tests.

Low test reliabilities decrease the overall ability of the evaluation effort to detect statistically significant differences between treatment groups. However, lower than desirable reliabilities typically do not effect estimates of mean performance if the tests are unbiased and administered to large numbers of students, as they are in the present evaluation effort. The overall low performance on the tests does, however, create a floor effect that may disguise indicators of treatment impact.

Table 4
Reliabilities (coefficient alpha) for Pretest and Posttest

Test	Number of Items	Number of Examinees	Test Reliability Based on Total Sample
Pretest	34	2,21410	.571
Posttest	34	3,844	.593

General Observations on Test Performance

Prior to presenting the student learning results for Karnataka a number of cautions regarding the interpretation of the data should be stated. One general observation regarding test performance was that students generally performed very poorly on the tests. Average percent correct performance on the pretest hovered around the 20% correct mark, which is below chance if a student guessed the answer to every question on the test. That is, each of the test items offered 4 alternative answers and if a student guessed on every question, by chance they would get 25% of the items correct. The fact that average performance is below 25% correct reflects the fact that many students did not even hazard a guess on all of the test questions. The pattern of many students not even guessing on items occurred on the posttest as well. Items not responded to were scored as incorrect on all of the tests. It is worthwhile mentioning that there are several reasons why students might not respond to an item. First, and most straightforward, they might not respond if they have no idea what the correct answer is. In this case scoring the item as wrong is the correct scoring option. Other reasons why students might leave an item unmarked is that they do not understand the directions for responding to items, they might not understand the wording of the item, or they might believe there was more than one correct answer in the alternatives and decide to leave it blank rather than making a choice between the believed correct alternatives. These all represent cases where scoring an item as wrong when it is blank may not reflect the true state of a student's knowledge.

Another way to look at pretest performance is to look at the frequency with which scores occurred. Ninety-nine students (4.5% of the total number of students taking the test) attained a score of zero, meaning that they either did not answer any of the questions or they answered so few that they could not get one correct by chance. Sixty-five percent of the students taking the test scored 25% correct, or worse. This means that over half of the students completing the pretest scored lower than chance if they had answered all of the questions. Another interesting comparison is to look at the score at which 90% of the students taking the test fell below that score. The score point that separated the lower 90% of the examinees and the top 10% was 33% correct. Finally, 15 students out of the

2,064 who took the pretest scored 50% or better, with the top score on the test being 61% correct.

An examination of the pretest data also indicated that the performance of intervention groups and the control group were relatively comparable before the interventions began. This provides a situation where intervention effects can be shown, if they are present.

It is apparent that the vast majority of the students completing the pretest had a great deal of difficulty with it, and there was ample room for scores to improve during the academic year.

The percentage of students performing below chance on the posttest was the same as the percentage performing below chance on the posttest (53%). There was, however, evidence of improved performance from the posttest in that the number of students achieving a score of zero declined from 15 to 7, and the highest score on the test improved from 68% correct (1 student) to 79% correct (3 students).

Karnataka Posttest Performance and the Comparison of Pretest and Posttest Gains

The analyses in the sections to follow will focus on comparisons between pretest and posttest performance. There is an earlier report comparing pretest and midtest performance that is available from EDC staff. The analyses reported in the section to follow have several purposes. First, descriptive statistics will be presented (group means) that show how students in the differing treatment groups perform on the pretests and the posttests. Tables will also be presented that show how males and females performed on the tests, how students from differing geographical regions performed on the tests, and how students with varying demographic characteristics performed. Additional analyses will also be performed that address the question of whether there are statistically significant differences between groups.

Comparisons Between Treatment Groups:

The first question addressed in the analysis of pretest and posttest data was whether the treatment groups differed from one another, and in particular, whether the groups differed in the amount of gain made from pretest to posttest. Table 5 shows the mean percent correct performance on the pretest and the posttest, and it shows the average amount of gain from pretest to posttest for students in the IRI and Control groups. The data in these tables was collected from students who took *both* the pretest and posttest. Students who completed one test but not the other will not be considered in the analyses.

The data for Karnataka students is presented separately for two groups of students. The data in the section of the table labeled “Karnataka Students” is for students in the Chamarajanagar, Deodurga and Sedam regions of the state. The second set of data is for students in the Bangalore region of Karnataka. There were distinct differences between the treatments presented to students in the different regions and these differences

warranted a separate comparison of performance. Specifically, in the Chamarajanagar, Deodurga and Sedam regions of the state the IRI programs were inadvertently broadcast out of sequence whereas in Bangalore they were broadcast in the correct order. Two other differences suggested that the regions be treated differently. In Bangalore students listened to the broadcasts on a daily basis whereas in the other regions the broadcasts were listened to three times per week. Finally, in Bangalore teachers were provided study guides to accompany the broadcasts whereas teachers in the other regions did not receive study guides.

Table 5
Mean Percent Correct Performance on the Pretest and Posttest,
And Pretest to Posttest Gain as a Function of Treatment Condition

Intervention		Percent correct on Pretest	Percent Correct on Posttest	Pretest to Posttest Gain
Chamarajangar, Deogurga and Sedam Students				
IRI	Mean	**23.1**	**30.1**	**7.0**
	N	427	427	
	Std. Deviation	9.32	12.5	
Control	Mean	**22.9**	**29.6**	**6.7**
	N	554	554	
	Std. Deviation	9.6	11.2	
Bangalore Students				
IRI	Mean	**27.1**	**32.0**	**4.9**
	N	411	411	
	Std. Deviation	9.7	11.8	
Control	Mean	**27.0**	**28.2**	**1.2**
	N	490	490	
	Std. Deviation	9.9	10.2	

The data summarized in the table above were subjected to statistical analysis using an analysis of covariance procedure (ANCOVA) where pretest performance was used as a covariate in the analysis of treatment differences in posttest performance. In essence, the ANCOVA procedure captures that part of the variance in posttest performance that is attributable to differences between students on the pretest. This increases the power (in the sense of being able to detect significant differences between groups) of the statistical analyses that look for differences in posttest performance between the IRI group and the Control group.

The ANCOVA analysis of the Bangalore data indicated that the IRI group scored significantly higher on the posttest (corrected for pretest performance) than did the Control group, $F(1,898) = 26.6$, $p < .01$. A similar analysis of the Chamarajangar,

Deogurga and Sedam regions of Karnataka indicated that there were no posttest differences (corrected for pretest differences) between the IRI and control group, $F(1,988) = .5$, NS.

Pre test to posttest gain by subject matter area . Another way of examining gain from pretest to posttest is to break performance down by subject matter area. Table 6 below shows test performance gain from pretest to posttest for each of the IRI and Control groups broken down by math, science and social studies test items. The table also breaks down the performance of Bangalore students separately from the performance of students in Chamarajangar, Deogurga and Sedam. As can be seen in the table, in most cases IRI students make more gains than Control students, though it is the case that the Control students make greater gains than the IRI students on the math tests for Karnataka students from Chamarajangar, Deogurga and Sedam. It is also noteworthy that the gap between IRI and Control groups (favoring the IRI group) is generally larger in Bangalore than it is for the other regions of Karnataka. Finally, the negative values in the table indicate that students performed better on the pretests than they did on the posttest.

Analyses of covariance (ANCOVA) were performed on each of the three subject matter area posttests for both the Karnataka region students and the Bangalore students. The form of this analysis was that posttest performance was used as the dependent variable and pretest performance was the covariate. The analysis of math posttest performance for the Chamarajangar, Deogurga and Sedam students indicated that the IRI group and the Control group did not differ from one another, $F(1,1000) = .792$, NS. An analysis of the same students' performance on the science posttest also indicated that the IRI and Control groups were not statistically different from one another, $F(1,1000) = 1.5$, NS, and a similar outcome occurred in the analysis of the social studies items where again there were no differences between the IRI and Control group, $F(1,1000) = 1.3$, NS.

The same analyses were performed on the data from the Bangalore students. The ANCOVA analysis of math posttest performance indicated that the IRI students performed significantly better on the posttest (adjusted by the ANCOVA) than did the Control students, $F(1,904) = 11.0$, $p < .01$. The analysis of the science posttest performance also indicated that the IRI group significantly outperformed the Control group, $F(1,905) = 13.1$, $p < .01$. Finally, the analysis of the social studies items again indicated a significant advantage for IRI students over the Control students, $F(1,905) = 19.5$, $p < .01$.

Table 6
Mean Pretest to Posttest Gain in Subject Matter Areas as a Function of Treatment Condition

Intervention		Math Pre to Post Gain	Science Pre to Post Gain	Social Studies Pre to Post Gain
Chamarajangar, Deogurga and Sedam Students				
IRI	Mean	**3.0**	**11.6**	**7.8**
	N	439	439	439
	Std. Deviation	18.8	20.9	19.2
Control	Mean	**4.6**	**9.8**	**6.5**
	N	564	564	564
	Std. Deviation	19.9	19.8	17.6
Bangalore Students				
IRI	Mean	**2.09**	**11.2**	**2.55**
	N	412	412	412
	Std. Deviation	19.4	21.5	21.8
Control	Mean	**-.76**	**7.6**	**-1.5**
	N	495	496	496
	Std. Deviation	19.9	20.4	20.6

Comparisons Between Boys and Girls

The next set of analyses address the questions of whether boys and girls differ in test performance. Table 7 below shows the average performance of boys and girls on the pretest and posttest and it shows the average pretest to posttest for boys and girls. The table shows that the differences between boys and girls were generally small on both the pretest and the posttest, though in the Chamarajangar, Deogurga and Sedam regions girls made somewhat larger pretest to posttest gain than did the boys. A statistical analysis (ANCOVA) indicated, however, that the difference in gain between boys and girls was not significant, $F(1,2201) = 2.4$, NS.

The analysis of overall performance of girls and boys for the Bangalore data also showed that there were no differences between the sexes on adjusted posttest performance, $F(1,904) = .795$, NS.

Table 7
Mean Percent Correct Performance on the Pretest and Posttest As a Function of Student Sex

Student Sex		Percent correct on Pretest	Percent Correct on Posttest	Pretest to Postest Gain
Chamarajangar, Deogurga and Sedam Students				
male	Mean	**22.9**	**27.2**	
	N	1060	1060	**4.2**
	Std. Deviation	9.22	12.7	
female	Mean	**21.9**	**27.6**	
	N	1077	1077	**5.6**
	Std. Deviation	9.2	11.3	
Bangalore Students				
male	Mean	26.3	29.6	
	N	540	478	**3.1**
	Std. Deviation	10.8	10.9	
female	Mean	27.5	30.3	
	N	466	430	**2.9**
	Std. Deviation	10.2	11.4	

Male and female performance by treatment group. Another way to look at the performance of boys and girls is to examine the results broken down by treatment groups. Table 8 does this for both Bangalore students and for students from the other Karnataka regions. The ANCOVA analysis of the Chamarajangar, Deogurga and Sedam posttest data indicated that sex was a significant factor in the analysis (girls outperforming boys), $F(1,998) = 9.4$, $p < .01$, but there was no main effect for intervention and the sex X interaction was non-significant.

An analysis (ANCOVA) of posttest performance for Bangalore students that included both treatment and student sex as factors in the analysis indicated that treatment was a significant source of variance, but student sex and the interaction between student sex and intervention were not statistically significant.

Table 8
Mean Performance on the Pretest, Posttest and Pretest to Posttest Gain as a Function of Intervention Treatment and Student Sex

Intervention	Student Sex		Percent correct on Pretest	Percent Correct on Posttest	Pretest to Posttest Gain
Chamarajangar, Deogurga and Sedam Students					
IRI	male	Mean	23.8	28.4	
		N	203	203	**4.6**
		Std. Deviation	9.4	10.2	
	female	Mean	22.5	31.6	
		N	224	224	**9.1**
		Std. Deviation	9.2	14.2	
Control	male	Mean	23.5	29.0	
		N	295	295	**5.5**
		Std. Deviation	9.5	11.1	
	female	Mean	22.3	30.1	
		N	259	259	**7.8**
		Std. Deviation	9.6	11.4	
Bangalore Students					
IRI	male	Mean	27.0	31.5	
		N	236	236	**4.5**
		Std. Deviation	9.9	11.9	
	female	Mean	27.3	32.6	
		N	175	175	**5.3**
		Std. Deviation	9.4	11.7	
Control	male	Mean	26.4	27.8	
		N	237	237	**1.3**
		Std. Deviation	9.6	9.4	
	female	Mean	27.6	28.6	
		N	253	253	**1.0**
		Std. Deviation	10.5	10.8	

Male and female performance by subject matter and treatment group. Performance of boys and girls was also examined broken down by treatment group and subject matter area. Table 9 below shows that there is some variation between treatments in terms of the amount of gain that boys and girls made on the different subject matter tests. In

general, girls make more gains than boys in all three subject matter areas, though Bangalore boys make greater gains than Bangalore girls in math.

Table 9
Mean Performance Gain on Math, Science and Social Studies Items as a Function of Intervention Treatment and Student Sex

Intervention	Student Sex		Math Pre to Posttest Gain	Science Pre to Posttest Gain	Social Studies Pre to Posttest Gain
Chamarajangar, Deogurga and Sedam Students					
IRI	male	Mean	**.43**	**9.0**	**5.5**
		N	205	205	205
		Std. Deviation	17/9	19.7	17.1
	female	Mean	**5.3**	**13.8**	**9.9**
		N	234	234	234
		Std. Deviation	17.9	21.6	20.6
Control	male	Mean	**3.8**	**8.7**	**6.0**
		N	304	304	304
		Std. Deviation	20.5	19.7	17.6
	female	Mean	**5.5**	**11.1**	**7.0**
		N	260	260	260
		Std. Deviation	19.1	19.8	17.5
Bangalore Students					
IRI	male	Mean	**2.2**	**10.5**	**1.8**
		N	236	236	236
		Std. Deviation	19.9	20.9	21.6
	female	Mean	**1.87**	**12.3**	**3.5**
		N	176	176	176
		Std. Deviation	18.7	22.3	22.0
Control	male	Mean	**0.3**	**8.6**	**-2.66**
		N	242	242	242
		Std. Deviation	19.6	19.4	20.1
	female	Mean	**-1.51**	**6.6**	**-.53**
		N	254	254	254
		Std. Deviation	20.2	21.2	21.1

Pretest and Posttest Performance by Geographical Region:

The next set of analyses examined student performance on the pretest, the posttest, and pretest to posttest gain as a function of geographical region. Table 10 below presents this data. As can be seen in the Table, students in Deogurga made larger gains than the other regions and Bangalore tended to make smaller gains than the other regions. The smaller gains for Bangalore probably were due to different testing periods in Bangalore than in

the other regions. In Chamarajangar, Deogurga and Sedam pretest were administered in July 2004 and posttests in April/May 2005. In Bangalore the pretests were administered in November 2004 and the posttests in April/May 2005. Thus, there was smaller time period for the Bangalore students to show pretest to posttests gains.

Table 10
Mean Percent Correct on the Pretest, the Posttest and Pretest to Posttest Gain as a Function of Geographical Region

Region		Percent correct on Pretest	Percent Correct on Posttest	Pretest to Posttest Gain
Chamarajanagar	Mean	**22.8**	**30.3**	
	N	323	302	**7.5**
	Std. Deviation	9.91	10.7	
Deogurga	Mean	**22.5**	**32.2**	
	N	438	337	**9.7**
	Std. Deviation	10.6	14.2	
Sedam	Mean	**21.9**	**26.9**	
	N	419	352	**5.0**
	Std. Deviation	8.7	9.3	
Bangalore	Mean	**27.0**	**29.9**	
	N	901	901	**2.9**
	Std. Deviation	9.9	11.1	

The performance of students in different regions was also broken down by treatment group and this data is presented in Table 11 below. This data shows that the largest gap between IRI and control groups occurs in the Bangalore region.

Table 11
Mean Performance on the Pretest, Posttest, and Pretest to Posttest Gain as a Function of Intervention and Geographical Region

Intervention	Region		Percent correct on Pretest	Percent Correct on Posttest	Pretest to Posttest Gain
IRI	Chamarajanagar	Mean	**23.0**	**30.7**	**7.6**
		N	112	112	
	Deogurga	Mean	**22.9**	**32.1**	**9.2**
		N	146	146	
	Sedam	Mean	**23.3**	**28.0**	**4.7**
		N	169	169	
	Bangalore	Mean	**26.8**	**32.0**	**5.2**
		N	468	412	
Control	Chamarajanagar	Mean	**22.7**	**30.0**	**7.3**
		N	211	190	
	Deogurga	Mean	**22.2**	**32.2**	**10.0**
		N	292	191	
	Sedam	Mean	**21.0**	**25.9**	**4.9**
		N	250	183	
	Bangalore	Mean	**26.7**	**29.9**	**3.2**
		N	1006	908	

Pretest and Posttest Performance as a Function of School Location:

The analyses in this section show student performance for students broken down by whether they are enrolled in rural, urban or tribal schools. This data comes from the Chamarajanagar, Deodurga and Sedam regions of Karnataka and does not include data from Bangalore. Table 15 shows average student performance on the two tests and the amount of pretest to posttest gain for the three school locations, and Table 16 shows student performance for schools in the three locations broken down by treatment condition. As can be seen in Table 15, students in tribal schools make twice as much gain from pretest to posttest as did students in urban and rural schools.

Table 16 shows that there is considerable variation between the three school locations as a function of treatment group. This data should be interpreted cautiously though because of the small number of schools sampled for the posttest. And in particular the number of students in the IRI tribal group is very small (only 17 students).

Table 15
Mean Performance for Chamarajangar, Deogurga and Sedam Students on the Pretest, Posttest, and Pretest to Posttest Gain as a Function of School Location

Location of School		Percent correct on Pretest	Percent Correct on Posttest	Pretest to Posttest Gain
Rural	Mean	22.5	29.1	
	N	525	440	**6.6**
	Std. Deviation	9.6	11.5	
Urban	Mean	23.1	30.1	
	N	539	483	**7.0**
	Std. Deviation	9.7	11.7	
Tribal	Mean	18.4	30.9	
	N	116	68	**12.5**
	Std. Deviation	9.6	14.2	

Table 16
Mean Performance on the Pretest, Posttest, and Pretest to Posttest Gain as a Function of Treatment Intervention and School Location

Intervention	Location of School		Percent correct on Pretest	Percent Correct on Posttest	Pretest to Posttest Gain
IRI	Rural	Mean	20.6	28.4	
		N	293	210	**7.8**
		Std. Deviation	9.5	10.8	
	Urban	Mean	22.4	31.9	
		N	257	212	**9.5**
		Std. Deviation	10.4	14.0	
	Tribal	Mean	19.7	20.9	
		N	37	17	**1.2**
		Std. Deviation	8.2	8.0	
Control	Rural	Mean	22.6	29.5	
		N	320	235	**6.9**
		Std. Deviation	9.7	12.1	
	Urban	Mean	22.5	28.4	
		N	334	278	**5.9**
		Std. Deviation	10.1	9.5	
	Tribal	Mean	17.9	34.2	
		N	99	51	**16.3**
		Std. Deviation	9.7	14.3	

Pretest and Posttest Performance as a Function of Caste:

The next set of analyses examined student performance as a function of caste membership. Caste membership is coded in the files in the following way:

SC: Scheduled Caste. This classification refers to caste groups that have received formal legal protection in the constitution

ST: Scheduled Tribe. This classification refers to tribal groups that have received formal legal protection in the constitution

OBC: Other Backward Classes. This classification contains members of lower castes that have not received formal protection in the constitution.

Minority: These are members of religious groups other than Hindu.

Forward: These are members of caste groups that typically are more advantaged than lower castes groups

Any Other: These are members of groups that do not fit into the other categories. One example is members of nomadic tribes.

As can be seen in Table 17 below, there is relatively little variation between the caste groups in terms of the amount of gain they make from pretest to posttest.

Table 17
Mean Performance on the Pretest, Posttest, and Pretest to Posttest Gain as a Function of Caste

Caste		Percent correct on Pretest	Percent Correct on Posttest	Pretest to Posttest Gain
SC	Mean	21.7	28.9	
	N	386	311	**7.2**
	Std. Deviation	10.1	12.4	
ST	Mean	22.7	30.3	
	N	232	202	**7.6**
	Std. Deviation	9.2	11.1	
OBC	Mean	22.4	30.5	
	N	289	232	**8.1**
	Std. Deviation	10.0	11.8	
Minority	Mean	22.6	28.6	
	N	152	130	**6.0**
	Std. Deviation	8.8	9.9	
Forward	Mean	24.1	30.5	
	N	85	83	**6.4**
	Std. Deviation	10.8	14.1	
any other	Mean	21.5	29.9	
	N	35	32	**8.4**
	Std. Deviation	8.6	10.4	

Table 18 below presents the pretest and posttest performance for Chamarajangar, Deogurga and Sedam students and Bangalore students broken down by Caste and intervention condition. There does not appear to be any consistent noteworthy pattern in the data.

Table 18
Mean Performance on the Pretest, Posttest, and Pretest to Posttest Gain as a Function of Treatment Intervention and Caste

Intervention	Caste		Percent correct on Pretest	Percent Correct on Posttest	Pretest to Posttest Gain
Chamarajangar, Deogurga and Sedam Students					
IRI	SC	Mean	23.86	27.7	**3.9**
		N	113	113	
	ST	Mean	21.8	31.2	**9.3**
		N	94	94	
	OBC	Mean	24.2	32.0	**7.7**
		N	88	88	
	Minority	Mean	22.6	29.2	**6.6**
		N	73	73	
	Forward	Mean	22.3	31.3	**8.9**
		N	54	5	
	any other	Mean	25.7	31.6	**5.9**
		N	4	4	
Control	SC	Mean	22.3	29.9	**7.5**
		N	191	191	
	ST	Mean	23.9	29.6	**5.6**
		N	108	108	
	OBC	Mean	22.1	29.6	**7.6**
		N	143	143	
	Minority	Mean	23.0	28.0	**5.0**
		N	57	57	
	Forward	Mean	28.4	29.0	**.63**
		N	28	28	
	any other	Mean	22.1	30.4	**8.3**
		N	27	27	
Bangalore Students					
IRI	SC	Mean	26.8	30.3	**3.5**
		N	158	158	
	ST	Mean	21.9	25.4	**3.4**
		N	17	17	
	OBC	Mean	28.1	35.3	**7.2**
		N	138	138	
	Minority	Mean	27.7	29.4	**6.0**
		N	42	42	
	Forward	Mean	27.4	33.4	**6.0**
		N	39	39	
	any other	Mean	24.7	30.8	**6.1**
		N	14	14	
Control	SC	Mean	26.7	28.2	

		N	189	189	**1.4**
	ST	Mean	29.9	27.4	**-2.5**
		N	16	16	
	OBC	Mean	28.2	29.1	**.92**
		N	217	217	
	Minority	Mean	24.4	24.7	**.36**
		N	40	40	
	Forward	Mean	23.5	26.3	**2.7**
		N	19	19	
	any other	Mean	18.3	26.8	**8.5**
		N	9	9	

Summary of Chamarajangar, Deogurga, Sedam and Bangalore Student Learning Analyses

The most important analyses in the above section involved comparisons between treatment groups. The analyses involving the Chamarajanagar, Deodurga and Sedam regions of Karnataka indicated that there were no statistically significant differences between students enrolled in schools that listened to the IRI programs and those students enrolled in control schools. The same outcome occurred when analyses were conducted on the pretest to posttest gain on math, science, and social studies test items. That is, there were no difference between IRI and Control groups in these analyses.

There were, however, significant advantages for the IRI student in the Bangalore analyses. The overall analysis indicated that Bangalore students listening to the IRI programs made significantly greater gains than their Control counterparts, and follow-up analyses of math, science and social studies items indicated that IRI students made greater gains that did Control students in all three subject matter areas.

Additional analyses in the section examined the performance of boys and girls, the performance of students from urban, rural, and tribal locations, the performance of students from differing geographical regions, and the performance of students belonging to different castes. These results may be worthy of further study but there did not appear to be any striking differences in student performance as a function of these factors.

Pretest, Posttest, and Pretest to Posttest Gains for Chhattisgarh

Test Administration and Test Scoring

The English language comprehension and speaking tests were administered to groups of five children. In most occasions one group of 5 children in a school was administered the comprehension test and the speaking test was administered to another group of 5 children, though there were occasions where both the comprehension and speaking test was administered to the same 5 children.

Both the comprehension and the speaking test was administered via tape recordings though instructions for taking the test were supplemented by having the examiner assist in situations where it was clear that children were not understanding what they were supposed to do. The comprehension test involved presenting a child with a direction and then asking them to either select a picture or complete an action that corresponded to the direction in English. So, for example, a child might be directed to select the picture of a flower from a group of pictures or to touch his or her nose. The first child would respond to question 1 and the second child in the group of 5 would respond to question 2, with the cycle continuing until all of the questions had been answered. Questions responded to correctly were scored as 1 and incorrect questions as zero.

The speaking test was administered in a similar way with each child in a group of 5 being presented with a direction that they were to respond in English to. So, for example, the examiner might point to a picture of a dog and the child was expected to provide the English name for the thing being pointed to. Speaking items were scored as zero if the item was answered incorrectly, and as either .5 or 1 if it was answered correctly. For example, if the examiner pointed to a horse and asked what it was, the child that said "horse" was given a score of .5. If the child used a complete sentence, "That is a horse" the child would be given a score of 1.

Test Reliability

The reliabilities for the comprehension and speaking tests are listed in the table below. Prior to discussing the reliabilities it is important to understand what the comprehension and speaking tests are measuring. They should be conceptualized as providing an estimate of the English comprehension and speaking ability of a classroom of students. The tests were administered by randomly sampling five students from a classroom, and then having students answer items in succession. So item 1 would be answered by student 1, item 2 by student 2, and so on. This process introduces considerable variability within a set of items associated with the individual competence of each student. Thus, it would not be expected that a given assessment for a class would have high inter-item consistency. However, since the tests were administered pre and post to the same five students, any gains in performance should be associated with the collective gain in competence of the five students, and, since the five students were selected randomly from classrooms, improvements in test performance could be interpreted as improvements in the competence of the class as a whole.

Given this explanation, as can be seen in the table below, the reliabilities are relatively low. In addition to the reason mentioned above, another factor contributing to the low reliabilities is the small number of items on both the comprehension and the speaking test. Most tests based on 8 or fewer items would have low reliabilities.

Tests with reliabilities as low as the ones used in this study would be unacceptable for making decisions about individual students. However, the reliabilities are not too low to be used to detect differences between large groups of students, as is the goal for the tests

in this evaluation study. On a final note about test reliability, one encouraging factor is that the reliability for both the comprehension and the speaking tests improved from pretest to posttest.

Table 19
Reliabilities (coefficient alpha) for Pretest and Posttest

Test	Number of Items	Number of Examinees	Test Reliability Based on Total Sample
Comprehension Pretest	8	417	.358
Speaking Pretest	6	398	.238
Comprehension Posttest	8	304	.488
Speaking Posttest	6	301	.418

Test Results

Comparisons Between Treatment Group

The first set of analyses examined whether there were differences between the comprehension and speaking performance of children listening to the IRI broadcasts and those enrolled in control schools. Table 20 below reports the performance on the pretests, the posttests and the pretest to posttest gain. The comprehension test contained 8 items so if one item were answered correctly it would be 12.5% correct, 2 correct would provide a percent correct score of 25%, and so on. The speaking test contained 6 items so 1 item correct would be scored as 16% correct, 2 items would be 32% correct and so on.

The data was examined using and analysis of covariance procedure (ANCOVA) where the dependent variable was performance on the posttest, the independent variable was intervention treatment, and pretest performance was the covariate. The analysis of the comprehension data indicated that the IRI group performed significantly higher on the posttest than did the Control group, $F(1,298) = 6.54$, $P < .01$. A similar analysis of the speaking data indicated that again the IRI group performed significantly higher on the posttest than did the Control group, $F(1,281) = 15.6$, $p < .01$.

Table 20
Percent Correct on the Comprehension and Speaking Pretests, Posttests, and Pretest to Posttest Gain as Function of Intervention Treatment

Intervention		Percent Correct on Pre Test	Percent Correct on Posttest	Pretest to Posttest Gain
Comprehension Test				
IRI	Mean	**14.2**	**33.5**	
	N	257	191	**19.3**
	Std. Deviation	15.2	21.3	
Control	Mean	**18.3**	**28.6**	
	N	160	113	**10.3**
	Std. Deviation	15.6	19.6	
Speaking Test				
IRI	Mean	**.14**	**16.8**	
	N	246	169	**16.6**
	Std. Deviation	.35	13.3	
Control	Mean	**.17**	**12.4**	
	N	152	132	**12.2**
	Std. Deviation	.35	12.5	

Comparisons Between Boys and Girls

Table 21 below shows pretest scores, posttest scores and pretest to posttest gain made by boys and girls on the comprehension and speaking tests. The table shows that girls make more gain than boys on the comprehension test but they are very close to the same on the speaking test. An ANCOVA using speaking posttest as the dependent variable, sex as the independent variable, and speaking pretest performance as the covariate indicated that boys and girls did not differ on the speaking posttest. As noted in Table 21 there was a substantial difference in the gain made by boys and girls on the comprehension test. However, an analysis of the same form as used in the speaking test indicated that there was not a statistically significant difference between the sexes on the comprehension test, $F(1,298) = 1.44$, NS.

Table 21
Pretest, Posttest, and Pretest to Posttest Gain on the Comprehension and Speaking Tests as a Function of Student Sex

Intervention		Percent Correct on Pre Test	Percent Correct on Posttest	Pretest to Posttest Gain
Comprehension Test				
Males	Mean	**18.5**	**30.9**	
	N	209	151	**12.4**
	Std. Deviation	16.7	21.6	
Females	Mean	**13.0**	**32.4**	
	N	208	153	**19.4**
	Std. Deviation	13.6	19.9	
Speaking Test				
Males	Mean	**.19**	**15.1**	
	N	205	154	**14.9**
	Std. Deviation	.38	13.8	
Females	Mean	**.11**	**14.6**	
	N	193	147	**14.5**
	Std. Deviation	.32	12.5	

The next set of analyses addressed the question of whether intervention condition played a role in the gain made by boys and girls on the two tests. This data is displayed in Table 22 below. Looking first at performance on the comprehension test, the table shows that the girls make more gains than boys in both the IRI group and the Control group. An ANCOVA analysis of this data using comprehension posttest performance as the dependent variable, sex and intervention treatment as independent variables, and comprehension pretest performance as the covariate, showed that sex was not a significant source of variance, and the interaction between sex and treatment was also non-significant. Consistent with the previous analysis, intervention was a significant source of variance with the IRI group performing better than the Control group.

The next analysis used the same form of analysis in that speaking posttest performance was the dependent variable, sex and intervention the independent variables, and speaking pretest performance was the covariate. The pattern of outcomes was the same in that the effects for sex and the sex by treatment interaction were non-significant, but consistent with the previous analysis, there was a significant effect for intervention with the IRI group scoring higher on the posttest than the Control group.

Table 22
Pretest, Posttest and Pretest to Posttest Gain on the Comprehension and Speaking Tests as a Function of Student Sex and Intervention Treatment

Intervention	Student Sex		Pretest Percent Correct	Posttest Percent Correct	Pretest to Posttest Gain
Comprehension Test					
IRI	Male	Mean	**16.9**	**32.3**	
		N	128	95	**15.4**
		Std. Deviation	16.7	22.2	
	Female	Mean	**11.4**	**34.6**	
		N	129	96	**23.2**
		Std. Deviation	13.0	20.4	
Control	Male	Mean	**21.0**	**28.5**	
		N	81	56	**7.5**
		Std. Deviation	16.6	20.7	
	Female	Mean	**15.5**	**28.7**	
		N	79	57	**13.2**
		Std. Deviation	14.1	18.7	
Speaking Test					
IRI	Male	Mean	**.18**	**16.4**	
		N	126	86	**16.22**
		Std. Deviation	.36	13.7	
	Female	Mean	**.10**	**17.3**	
		N	120	83	**17.2**
		Std. Deviation	.35	13.1	
Control	Male	Mean	**.22**	**13.6**	
		N	79	68	**13.4**
		Std. Deviation	.41	14.0	
	Female	Mean	**.13**	**11.2**	
		N	73	64	**11.07**
		Std. Deviation	.27	10.8	

Performance on the Comprehension and Speaking Tests as a Function of Urban or Rural School Location.

The results reported in Table 23 below show student performance broken down by whether the students are enrolled in an urban or rural school. As can be seen in the table, rural students make considerably more gain than urban students. However, this data should be interpreted with caution given that the relatively small number of students in the urban groups would suggest that they are coming from a single school.

Table 23
Pretest to Posttest Gain on the Comprehension and Speaking Tests as a Function of School Location (Urban or Rural)

Location		Comprehension Pretest to Posttest Gain	Speaking Pretest to Posttest Gain
Rural	Mean	**16.57**	**13.58**
	N	270	260
	Std. Deviation	23.4	13.2
Urban	Mean	**4.4**	**7.6**
	N	31	24
	Std. Deviation	27.49	13.8

Performance on the Comprehension and Speaking Tests as a Function of Geographical Block

The data files were coded so as to be able to separate the results by geographical blocks in Chhattisgarh. Specifically, schools were selected from the Abhanpur, Kanker and Kondegaon regions. Table 24 reports the pretest to posttest gain for each of the regions on both the comprehension and the speaking tests. An ANCOVA was performed on the data summarized in Table 24. For the comprehension data the form of this analysis was that comprehension posttest performance served as the dependent variable, block served as the independent variable and comprehension pretest performance was the covariate. This analysis indicated that block was a significant source of variance, $F(2,297) = 109.8$, $p < .01$. Follow up statistical contrasts of the three blocks indicated that Kondegaon students scored significantly higher than Kanker and Abhanpur students, and that Kanker students scored significantly higher than Abhanpur students. One aspect of the data is the surprisingly good performance of the Kanker students on the speaking test. It would be worthwhile to conduct a more detailed follow-up study to identify what schools in Kanker are doing to bring them up to a level equivalent to the Kondegaon schools on the speaking tests.

Table 24
Pretest to Posttest Gain on the Comprehension and Speaking Tests as a Function of School Block

Block		Comprehension Pretest to Posttest Gain	Speaking Pretest to Posttest Gain
Abhanpur	Mean	**6.5141**	**5.0373**
	N	142	134
	Std. Deviation	19.50849	9.02625
Kanker	Mean	**10.4911**	**20.9091**
	N	56	55
	Std. Deviation	22.45915	17.48550
Kondegaon	Mean	**30.0971**	**19.9123**
	N	103	95
	Std. Deviation	23.89961	8.62047

The next set of analyses examined the performance of the students in different blocks in terms of whether they were in the IRI intervention condition or the Control intervention condition. Table 25 shows this table. As can be seen in the table, the IRI groups in each of the 3 geographical blocks makes substantially greater gains than do their Control counterparts in the same block. These differences are particularly striking in the Kanker block.

Table 25
Pretest to Posttest Gain on the Comprehension and Speaking Tests as a Function of School Block and Intervention Condition

Intervention	Block		Comprehension Pretest to Posttest Gain	Speaking Pretest to Posttest Gain
IRI	Abhanpur	Mean	**8.5938**	**6.4010**
		N	80	69
		Std. Deviation	20.14522	10.51618
	Kanker	Mean	**14.4444**	**26.6204**
		N	45	36
		Std. Deviation	21.64334	14.74915
	Kondegaon	Mean	**35.1190**	**21.1310**
		N	63	56
		Std. Deviation	23.53116	7.10034
Control	Abhanpur	Mean	**3.8306**	**3.5897**
		N	62	65
		Std. Deviation	18.46972	6.90480
	Kanker	Mean	**-5.6818**	**10.0877**
		N	11	19
		Std. Deviation	18.84446	17.47550
	Kondegaon	Mean	**22.1875**	**18.1624**
		N	40	39
		Std. Deviation	22.55469	10.27256

Performance on the Comprehension and Speaking Tests as a Function of Standard

The data in Table 26 shows the amount of pretest to posttest gain for students in Standard 1 and 2. Students in Chhattisgarh are in a multigrade school setting so students in both standard 1 and standard 2 are receiving very similar instruction. Nonetheless, there should be some evidence that students in standard 2 perform better on the tests than students in standard 1 given that they have had experience with English instruction provided the previous year. Given this expectation the data in Table 26 below is somewhat surprising because it shows that there is very little difference between students in standard 1 and standard 2.

The next set of analyses divided the data in terms of both intervention condition and standard. Table 27 shows that students in the IRI condition show gains in pretest to posttest performance that are similar to what would be expected in that standard 2 students perform better than standard 1 students. In contrast, in Control schools standard 1 students actually perform better (in terms of gain) that standard 2 students, and there is

very little difference between the two standards on the speaking test. These results may suggest that there is either little actual instruction occurring in control schools, or that instruction is sporadic and sometimes of poor quality.

Table 26
Pretest to Posttest Gain on the Comprehension and Speaking Tests as a Function of Standard

Standard		Comprehension Pretest to Posttest Gain	Speaking Pretest to Posttest Gain
1	Mean	**15.4561**	**12.8289**
	N	148	152
	Std. Deviation	22.83860	12.20236
2	Mean	**15.1961**	**13.3838**
	N	153	132
	Std. Deviation	25.36498	14.69789

Table 27
Pretest to Posttest Gain on the Comprehension and Speaking Tests as a Function of Grade and Intervention Condition

Intervention	Standard		Comprehension Pretest to Posttest Gain	Speaking Pretest to Posttest Gain
IRI	1	Mean	**17.0699**	**15.3846**
		N	93	91
		Std. Deviation	24.22613	12.66393
	2	Mean	**20.6579**	**16.9048**
		N	95	70
		Std. Deviation	24.91798	14.87816
Control	1	Mean	**12.7273**	**9.0164**
		N	55	61
		Std. Deviation	20.19736	10.46324
	2	Mean	**6.2500**	**9.4086**
		N	58	62
		Std. Deviation	23.67659	13.53703

Performance on the Comprehension and Speaking Tests as a Function of Caste

The final set of analyses for Chhattisgarh examines the performance of students belonging to different caste groups. As can be seen in Table 27, there is large variation between the caste groups with respect to pretest to posttest gain. The largest gains occur in the OBC group and the smallest in the SC. These results should be interpreted with considerable caution given that the number of students contributing data is very small in the SC and the "others" caste groups. The next table reports the performance of students belonging to different castes as a function of treatment condition. Table 28 shows a pattern of results that are very similar to other patterns reported in this section of the evaluation report. Specifically, students in a particular caste show greater learning gains if they are enrolled in schools listening to IRI programs.

Table 27
Pretest to Posttest Gain on the Comprehension and Speaking Tests as a Function of Caste

Caste		Comprehension Pretest to Posttest Gain	Speaking Pretest to Posttest Gain
SC	Mean	**-.8065**	**8.3333**
	N	31	16
	Std. Deviation	24.77723	9.62250
ST	Mean	**15.5963**	**7.6211**
	N	109	117
	Std. Deviation	22.51260	11.08206
OBC	Mean	**19.5364**	**17.8657**
	N	151	139
	Std. Deviation	23.43325	13.55496
Others	Mean	**-1.2500**	**17.3611**
	N	10	12
	Std. Deviation	24.61509	15.26630

Table 28
Pretest to Posttest Gain on the Comprehension and Speaking Tests as a Function of Intervention Condition and Caste

Intervention	Caste		Comprehension Pretest to Posttest Gain	Speaking Pretest to Posttest Gain
IRI	SC	Mean	**9.3750**	**13.8889**
		N	12	9
		Std. Deviation	28.26669	8.33333
	ST	Mean	**18.4211**	**8.1633**
		N	57	49
		Std. Deviation	24.21828	11.59824
	OBC	Mean	**21.9318**	**19.8029**
		N	110	93
		Std. Deviation	23.41209	13.45472
	Others	Mean	**-2.7778**	**21.6667**
		N	9	10
		Std. Deviation	25.60043	12.54621
Control	SC	Mean	**-7.2368**	**1.1905**
		N	19	7
		Std. Deviation	20.54627	5.75055
	ST	Mean	**12.5000**	**7.2304**
		N	52	68
		Std. Deviation	20.26177	10.76513
	OBC	Mean	**13.1098**	**13.9493**
		N	41	46
		Std. Deviation	22.52624	13.03537
	Others	Mean	**12.5000**	**-4.1667**
		N	1	2
		Std. Deviation	.	5.89256

Summary of Chhattisgarh Student Learning Analyses

The major result of interest in the Chhattisgarh data was the comparison of the learning gains of students enrolled in schools receiving the IRI broadcasts versus the gains made by students enrolled in control schools. The analyses showed that IRI students made

significantly greater gains on both the comprehension and the speaking tests than did the control students.

The examination of the gains made by boys and girls tended to show that girls made greater gains than boys, though these differences were not statistically significant.

The general conclusion that IRI students made greater gains that control students was not moderated in examinations of other variables coded in the data files. That is, IRI students were superior to control students in both urban and rural populations, in the three geographical blocks contributing data to the evaluation effort, and in each of the caste groups examined in the analyses. There was also an interesting comparison of IRI impact that involved standard 1 and standard 2 students. In the control schools the expected pattern of standard 2 students performing better than standard 1 students was not present. Instead, standard 1 students were actually better than standard 2 students on the comprehension test. However, in the IRI schools the expected pattern did emerge in the standard 2 students performed better than standard 1 students on both the comprehension and the speaking tests.

Overall Summary and Interpretive Comments

The evidence for positive impact of the India dot-EDU program is mixed. There is clear evidence of positive program impact in Chhattisgarh where students enrolled in schools receiving the IRI broadcasts made significantly greater gains on both the comprehension and speaking tests than did the students enrolled in control schools.

The evidence in Karnataka shows program impact in the Bangalore region of Karnataka, but not in the Chamarajangar, Deogurga and Sedam regions of the state. In Bangalore the IRI students made significantly greater gains than control students in overall pretest to posttest gain, and this significant advantage was consistent across the three subject matter areas of math, science and social studies.

In Chamarajangar, Deogurga and Sedam there was no statistical differences between the overall pretest to posttest gain made by the IRI students, and there was no evidence that IRI students' performance differed from control students' performance on the math, science and social studies sections of the tests.

The difference in the pattern of results for Bangalore and the other regions of Karnataka is likely to be attributable to differences in the way the IRI programs were implemented in the different regions. Three factors, in particular, were probably important. IRI broadcasts in Chamarajangar, Deogurga and Sedam began in July of 2004. In some regions of Chamarajangar, Deogurga and Sedam early reception of the IRI broadcasts was quite poor and it is certain that many IRI students were not exposed to some broadcasts early in the school year. Given that the IRI instructional modules are hierarchical in structure, meaning that early broadcasts contain content important for the learning of material presented in later broadcasts, the missing of early broadcasts probably had a deleterious impact on the overall learning of students receiving the IRI

broadcasts. The problem of signal strength was discovered and corrected later in the school year, but by that time the negative instructional impact had probably already occurred.

A second factor that probably muted the impact of the IRI experience was the way some schools in Chamarajangar, Deogurga and Sedam utilized the broadcasts. The IRI programs were constructed in modular form with each module containing multiple broadcasts that were designed to be heard in sequence. For instance, a particular math module might consist of 3 programs that were supposed to be broadcast one after another, and this module of math programs might be followed by a module of science broadcasts consisting of several programs that were again designed to be broadcast in succession. In some areas of Chamarajangar, Deogurga and Sedam the modules were broken up so that in 3 successive broadcasts there would be a math program, a science program and a social studies program. This breaking up of the modular properties of the programs greatly reduced the effectiveness of the programs as an instructional procedure.

A final factor that may have lessened the impact of the IRI programs in Chamarajangar, Deogurga and Sedam was the late delivery of teacher handbooks. These handbooks instructed teachers about how to use the IRI broadcasts most effectively, and how to implement the interactive part of the IRI instruction. The handbook was finished and delivered in November, but this late delivery may have lessened the effectiveness of the broadcasts that began in July.

The negative factors mentioned above did not occur in Bangalore. The Bangalore broadcasts started late, in November, but the programs were broadcast with good signal strength, the modules were broadcast in the correct sequence, and teachers were given handbooks that allowed them to optimally use the broadcasts in their classrooms. The late starting date for the Bangalore broadcasts undoubtedly reduced the amount of possible gain that IRI students might have made. That is, the shorter interval between pretesting and posttesting in Bangalore probably resulted in a lesser amount of overall gain in Bangalore than in the other three regions. However, the important point is that in the location where program implementation occurred as it was supposed to occur (in Bangalore), students listening to IRI programs made greater gains than did students enrolled in control schools. In contrast, in situations where program implementation was not optimal (Chamarajangar, Deogurga and Sedam) students listening to IRI broadcasts displayed no learning advantages relative to control students.

Another concern regarding the evaluation is associated with the learning tests. Both the Chhattisgarh and the Karnataka tests had reliabilities that were lower than desirable. Given the testing method it is unlikely that the Chhattisgarh tests will move much beyond a .5 reliability level. Using a procedure where five different children respond to test items in succession is an atypical procedure than is unlikely to produce high reliabilities. However, as argued in the body of the report, this procedure does provide an index of overall classroom performance that is adequate for the purpose of documenting changes in the English language skill of the class as a whole.

The Karnataka tests are more problematic. There were many children in Karnataka who did not even guess at a large number of the test items, and there were some who left the entire test blank, even on the posttest after a year of instruction on the content being tested. There are two likely possibilities for this situation, and perhaps the outcome was a combination of both. The first possibility was that the content of the tests exceeded the academic capability of the majority of the students being tested. It should be mentioned that the test development process involved procedures for trying to prevent this from happening, but nonetheless it is possible that it did.

The second possibility was that the multiple choice testing procedure that was used was unfamiliar to the students and they were unwilling to respond in an unfamiliar situation. Again, the test development process should have prevented this occurrence in that there were ample opportunities during test development for local educators to inform the development process. It is possible, however, that the problem occurred.

In an effort to improve the quality of the Karnataka tests a workshop will be held in late June, 2005, in Bangalore that will thoroughly review the Karnataka tests at both the content and procedural level. Hopefully this will result in the production of tests that are accurate reflections of student knowledge.