<u>WORKING PAPERS</u>:
EXPANDED BASIC EDUCATION PROGRAM

The Child Friendly Schools Movement and Impacts on Children's Learning:

Practical Applications in Cambodia

Kampuchean Action for Primary Education

Kurt Bredenberg
You Heeyit

July, 2004

Phnom Penh, Cambodia

Funded with Support from UNICEF and the Swedish International

Development Agency (Sida)

Abstract

This research paper is the first of two studies that are being funded through a national program to support Child Friendly Schools (CFS) in Cambodia. The study's primary expectation was that program interventions would lead to significant and positive changes in critical and creative thinking skills among target groups relative to the comparison group in non-program sites. Conclusions in this regard were based on analysis of mean differences between comparison groups and effect sizes on locally designed testing instruments. Interventions consisted primarily of changes in the physical learning environment and teacher practice following a child rights approach to school development. The study employed a cluster sampling technique in which class was identified as the primary sampling unit. Sampling was not random but was based on equivalency matching of sampling units. Research results tentatively suggest that expectations relating to positive impact on higher order thinking skills among children in CFS classrooms are being fulfilled. In all cases, total mean scores among experimental group children were higher than those for control groups of comparable composition and that these differences were statistically significant with a very low probability of error ($p<.01$). As importantly, effect sizes indicated that the differences in total mean scores between comparison groups were moderate to large. Nevertheless, researchers stopped short of widely generalizing these conclusions due to the pre-experimental nature of the research design.

1. Introduction

1.1. Background

Since the adoption of the Convention on the Rights of the Child by most of the world's countries, educators have been searching for practical ways to translate many of the Convention's ideals into concrete applications, particularly in the developing world. One of the movements that best capture a rights-based approach to educational development has come to be known as the Child Friendly Schools Movement. Because of its emphasis on *participatory learning and action* approaches (PLA), this movement has become highly eclectic in character, manifesting different directions in different places depending on input from the grassroots (Hopkins and Chaimuangdee, 2000). Nevertheless, there are a number of defining themes that do seem common to programs in most countries including, improvements in psychosocial learning environments, gender sensitivity, inclusiveness, health and nutrition, and parental engagement among others.

Child Friendly School (CFS) approaches have variously been adopted in many different places as a counterweight to the "effective schools" movement, which has been a powerful philosophy to improve school management and classroom practices during the last three decades. Although effective schools applications began originally in North America and the United Kingdom, its potential to improve education in developing countries was soon realized. But while applications of the effective schools movement in the developing world originally began as a means to improve participation among the most marginalized children of society (e.g., Lockheed and Levin, 1988), its co-option by economic models of educational development in many national programs has led to permutations that focus more on efficiency and numbers than on quality and children's welfare. As a result, many educators have turned to CFS approaches as a means to put the emphasis back on quality, local participation, and child welfare.

1.2. The Development of Child Friendly Schools in Cambodia

A national Child Friendly School program began in Cambodia in 2001 with funding support from UNICEF and the Swedish International Development Agency (Sida). The program was established as a tripartite partnership between the national Ministry of Education, UNICEF/Sida, and a local NGO known as Kampuchean Action for Primary Education (KAPE). KAPE has provided technical direction for the program and has formulated an implementation strategy that is appropriate to the Cambodian context (see below). The program is now in its 4th year and is operating in x schools in 5 provinces. As is true of many other national CFS initiatives, this program is designed to promote educational innovation in Cambodia's state school system. The program is characterized by a *child rights approach* and includes activities in 4 major components:

- Gender Sensitive and Inclusive Education
- Psychosocial Learning Environments
- Health and Nutrition
- Parental Engagement

CFS activities generally begin each year with local planning exercises that include a child sensitization workshop in which children indicate through pictures, skits, and other devices what they would like to see happen in their school. Parallel exercises also occur for parents with respect to their expectations of the local school. Schools then develop improvement plans based on a menu of activities to help meet the needs stated by children and parents. Selection of activities is totally voluntary and stakeholder driven. Sample activities include action research by teachers relating to gender issues, experimental classrooms, student associations, breakfast programs, scholarships for poor children, life skills,

creative writing, debate clubs, and others.
One of the key interventions in the national Child Friendly Schools Program in Cambodia, however, relates to the emergence what have become known as *model* or *experimental classrooms.* These are classrooms where the resident teacher *volunteers* to radically change both the physical environment and teaching practice in ways that mark a radical departure from traditional methodologies. Those teachers who volunteer are then provided with material assistance to change the classroom environment and technical support to learn about new child friendly teaching methodologies. Physical learning conditions are changed to make the classroom environment less formal and more comfortable. Some teachers, for example, have used program funds to refurbish classrooms to allow children to sit on the floor around low tables as they do at home. Teachers also use portfolios of children's work not only as a means to make student evaluation less threatening but to actually involve children in their own assessment. But the main change in classroom learning has come from modifications in teaching practice. This refers especially to heavy reliance on cooperative learning principles in which children work together and help each other as teams.

The use of teacher volunteers reflects a strategic decision early on in program design to operate on a principle of *stakeholder engagement.* The Cambodian CFS program has tried to internalize lessons of the past in which millions of dollars of donor money allocated to mandatory, school-wide in-servicing for teachers has led to little actual change in classroom practice (e.g., Wheeler, 1998). These well-intentioned but unsuccessful past initiatives to improve classroom practice reflected decision-making at the upper levels only, which was then imposed on schools from without. Given the poor and irregular nature of teacher payment as well as low levels of professionalism and accountability in much of the school system, such decisions were based on very vulnerable assumptions that change through mass in-servicing was possible. The CFS program design in Cambodia, however, recognizes that such fundamental change in educational practice cannot be imposed from above but must be built up consensually from the bottom. Although this approach requires more time than top-down initiatives, program resource personnel has so far found the results to be both inspiring and more sustained. More change in classroom practice has been detected through peer pressure, internalized values of professionalism, and an emerging local culture of quality in target schools than was true of previous projects that used externally imposed mass in-servicing as a means to change teaching methodology.

1.3. Focus on "Maximal" Learning

In order to assess changes in student learning as a result of technical inputs in experimental classrooms, the program has been conducting periodic testing of children with a focus on cross-cutting skills in *critical* and *creative* thinking. Historically, the Child Friendly School program has been designed to avoid a "minimalist" approach to student learning and has instead focused on building learning competencies in the higher order thinking skills. While there have been calls from some quarters for the program to refocus on basic minimum learning competencies, program planners have so far resisted the temptation to do so. This is not only due to the fact that other projects have already undertaken complementary activities to monitor minimum learning standards but also because of Ministry expectations to develop a program whose approach to quality is "maximal" in nature, not minimalist. To be sure, there are questions as to whether conventional testing can measure changes in "creativity" or even whether creative thinking can be defined at all (see below). Nevertheless, program designers have tried to use the latest research to map out a set of domains that approximate what many com-

monly understand to mean creative and critical thinking.
As part of their training, volunteer teachers who manage experimental classrooms study a module that focuses exclusively on critical and creative thinking. The module begins with an attempt to engage teachers in a discussion that defines and distinguishes between critical and creative thinking. This leads to the establishment of some defining characteristics between the two modes of thinking possesses many of the elements listed below:

Critical Thinking
- Identifying common relationships
- Identifying common patterns
- Ordering things into logical sequences
- Classifying elements by like attributes
- Problem solving
- Application of principles
- Etc.

Creative Thinking
- Deriving original ways of doing things
- Combining existing ideas to produce new ones
- Flexible expression of ideas
- Fluency in the use of concepts
- Etc.

These discussions help teachers to realize that critical and creative thinking are not mutually exclusive areas but have areas of overlap, particularly in the area of problem solving.

Teachers are assisted in the above activity by a previous module that introduces Bloom's taxonomy of educational objectives. This taxonomy outlines a hierarchical organization of thinking skills that starts with Knowledge (or Memory) and moves up through 6 stages to what Blooms defines as Evaluation. In this respect, critical thinking skills are most analogous to skills in the taxonomy that are known as Application and Analysis while creativity links with what Bloom calls Synthesis.

Following the above exercise, teachers review selected lessons from their textbooks and try to develop questions or tasks that reflect the characteristics of critical and creative thinking that were discussed previously. Teacher groups then present tasks/questions that they have developed and describe the characteristics that they exemplify (e.g., classification, application, etc.). An important theme in the Critical/Creative Thinking Module relates to the non-curriculum specific focus of the discussions, that is teachers try to think of thinking skills that are not linked to any particular subject but rather which are cross-cutting in nature. These capacity building activities with teachers, coupled with periodic follow-up by district-based staff, are among the primary interventions orchestrated by the project to change what children learn and how they learn it.

2. Research Design and Methodology

2.1. Overview

The national CFS program in Cambodia has attempted to build into its design educational research activities that would allow program managers to make some tentative conclusions about impacts of pedagogical interventions, such as those described above, on children's learning. As mentioned several times earlier, these activities have focused very heavily on impacts with respect to the higher order thinking skills, i.e., critical and creative thought. *The study's primary expectation, therefore, is that program interventions will lead to significant changes in thinking skills among target groups relative to a comparison group.* Accordingly, learning outcomes, operationalized by test scores, were the study's primary *dependent variable* while the receipt of program interventions or treatments was the primary *independent variable*. Manipulation of the independent variable was defined simply as participation or nonparticipation in a CFS classroom. The research design also allowed for analyses of measured variables such as *sex* and *age* to determine whether these played any role in observed learning outcomes.

Those responsible for research activities adopted a pre-experimental *static group comparison* design that generated comparative data from a terminal test administration in two independent samples of children in the closing months of an academic year. These comparisons derive from test scores of children studying in experimental conditions (i.e., those learning in experimental classrooms) and those in a control condition in non-CFS program sites. The tests that were developed to measure selected learning outcomes are standardized, non-curriculum specific, and focus on higher order thinking skills. For purposes of interpretive clarity, test scores have been expressed in a percentage scale to enable easy comparisons for readers.

Test administrations span several years and reflect efforts to concentrate technical interventions in a limited number of grades initially followed by incremental expansion to other grade levels as target cohorts move through the educational system. Accordingly, experimental classrooms were first established in the program at Grades 1, 5, and 6 with subsequent expansion to Grades 2, 3, and 4 as children moved to higher grades from Grade 1. In order to ensure that the same children initially targeted continued to benefit from the new pedagogy, schools received assistance in keeping cohorts together as they were promoted to higher grades. The first round of testing activity was, therefore, administered to children in Grades 1, 5, and 6 in 2003/4 with a second round planned for Grades 2, 3, and 4 in 2005.

2.2. Sample Construction

Test administrations comprised 2,492 children in all conditions. Following the project design of the CFS program, researchers chose *class* as the primary sampling unit. Children in experimental conditions were selected from approximately 60 different classes, which were among the first to participate in the CFS program in two provinces. The final sample in the experimental condition numbered 34 classes comprising 1,324 children. Within the sample, 16 classes were chosen from Grade 1 and 9 each from Grades 5 and 6. Although subject assignment to groups was not random, experimental condition classes were chosen in a way to ensure a balance of schools in urban, semi-urban, and rural areas. In this respect, about 30% of target schools are located in urban or semi-urban areas while 70% tend to be rural.

A control group that also comprised 34 classes (and 1,168 children) was matched with counterparts in the experimental condition in a way to promote valid comparisons between classes. Classes in the control group were matched with those in the experimental condition on the basis of rough equivalence with respect to the following characteristics:

- Commune Poverty Rates
- Pupil Class Ratio (PCR)
- Urban/Rural Setting
- Sex of Teacher (not possible in all cases)

With the exception of teacher sex, overall pairings of classes did approximate rough equivalence in most cases (see Table 1). In cases where class characteristics by poverty or PCR did not fall within the arbitrarily set demarcation points between categories, study design still allowed for equivalent matching with a control class if the difference did not exceed 25%.

Table 1: Equivalency Matching of Experimental and Control Groups

Criteria	Experimental	Control
Commune Poverty Rate	32.9%	30.3%
PCR	39.1 : 1	38.6 : 1
Urban/Rural	10 Urban classes 24 Rural Classes	10 Urban classes 24 Rural Classes
Sex of Teacher	21 Male Teachers 13 Female Teachers	16 Male Teachers 18 Female Teachers

It should be noted that sub-samples for each grade are not always comparable in composition suggesting a need to avoid comparisons between sub-sample groups. In this respect, children in Grade 1 reflected an even split between urban and rural populations while classes in Grades 5 and 6 manifested a rural bias in composition. As a result of greater representation from among urban schools, more classes from Grade 1 tended to be from communes that had poverty rates of 30% or less than was true of sub-samples for Grades 5 and 6, which showed a more even balance with respect to commune poverty rate. The distribution of classes by each classification criteria within the experimental group is summarized in Table 2.

Table 2: Characteristics of the Sample by Classification Criteria (Experimental Group)

Classification Criteria	**Number of Classes**			
	Grade 1	***Grade 5***	***Grade 6***	***Total***
Commune Poverty Rate				
Less than 30% Poor	13	2	2	17
30% - 50% Poor	3	5	5	13
More than 51% Poor	0	2	2	4
Pupil-Class Ratio				
Less than 35 : 1	5	0	4	9
35 : 1 to 45 : 1	7	8	5	20
More than 45 : 1	4	1	0	5
Demographic Setting				
Urban/Semi-urban	8	1	1	10
Rural	8	8	8	24
Sex of Teacher				
Male	6	7	8	21
Female	10	2	1	13

Classifications of schools (and the classes that reside there) according to urban/rural differences and commune poverty rates have been based on national census data. Information on PCR and teacher sex on the other hand has been provided by Provincial Offices of Education in each respective province.

2.3. Test Development and Correction

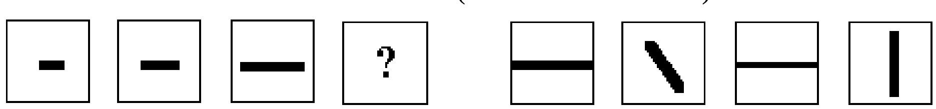
Table 3: Critical and Creative Thinking Domains

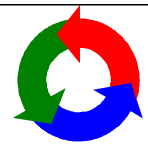
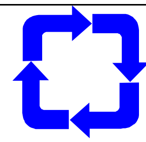
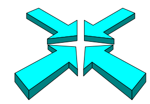
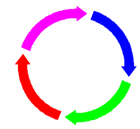
Critical Thinking Domains	Grade 1	Grade 5	Grade 6
1. Alternating Series	✓	✓	✓
2. Logical Sequences	✓	✓	✓
3. Classification	✓	✓	✓
4. Mazes and Puzzles	✓	✓	✓
5. Task Analysis	✓	--	--
6. Figural Series	--	✓	✓
7. Analogies	--	✓	✓
8. Concept Analysis (Fact/Opinion)	--	✓	✓
9. Textual Analysis	--	✓	✓
10. Application of Principles	--	✓	✓
Creative Thinking Domains			
11. Determining Consequences	--	✓	✓
12. Unusual Uses	--	✓	✓
13. Word Fluency	--	✓	✓
14. Expressional Fluency	--	✓	✓
15. Associational Fluency	--	✓	✓
16. Figural Elaboration	--	✓	✓
17. Simile Interpretation	--	✓	✓
18. Remote Association	--	✓	✓

Actual measurement of learning impacts in critical and creative thinking has relied on a battery of tests especially designed for the purpose. Using psychometric research studies that were also interested in similar assessments of higher order thinking skills, the program has designed its test batteries with 18 distinct domains to represent critical and creative thinking processes. These are shown in Table 3. Tests were field-tested with a small sample of about 100 children. Items that were objectively scored underwent item analysis followed by appropriate revisions. Test items of medium difficulty and high discrimination were retained for the final tests. Similarly, subjectively scored items were reviewed for clarity of directions, time required for responding, and the appropriateness of student answers. Given the difficulty of administering tests to Grade 1 children for whom written tests are not appropriate, testing activities with this group were limited to critical thinking

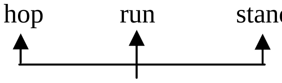
While there is general consensus among many psychometricians about what sorts of tasks constitute critical thought, there is less agreement about what constitutes creative thinking or whether such thinking skills can be measured at all. This fact notwithstanding, several theorists have shown intense interest in efforts to measure creativity (e.g., Guilford, 1974). In the main, such efforts have focused on what is known as *divergent thinking,* that is, reasoning processes (as opposed to products) that emphasize movement in multiple directions and which allow for the possibility of multiple solutions to any given problem. This is to be contrasted with *convergent thinking* (more generally used in critical thought) where test takers are required to use established principles to solve problems that have but one possible solution. Thus, whereas theorists use such terms as analysis, application, and unitary problem solving to describe critical thinking, other terms such as flexibility, originality, synthesis, and problem solving that may lead to multiple solutions are used to describe creativity.

Table 4: Question Exemplars Reflecting Selected Domains

Domain	Question Exemplar
(1) Alternating Series	What comes next? (Circle the best answer) 

<table>
<tr><td>(2) Classification</td><td>Circle the letter of the picture that does not belong.
<table>
<tr><td>a</td><td>b</td><td>c</td><td>d</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table></td></tr>
<tr><td>(3) Analogies</td><td>
<table>
<tr><td><u>Pair 1</u>
planet : sun</td><td><u>Pair 2</u>
moon : ____</td><td>a. earth</td><td>b. star</td><td>c. meteor</td><td>d. Venus</td></tr>
</table></td></tr>
<tr><td>(4) Determining Consequences</td><td>What would happen if it rained in Cambodia for one entire year without stopping? List as many consequences as you can think of in the space provided.</td></tr>
<tr><td>(5) Word Fluency</td><td>List as many words as you can think of that mean the same as the underlined word. Write your answers in the space provided.

Example: <u>big</u>
1. ________; 2. ________; 3. ________; 4. ________; 5. ________; 6. ________</td></tr>
<tr><td>(6) Remote Association</td><td>Look at the 3 words given below and try to find a fourth word that is related in the same way to the other 3. Be sure to state its relationship to the other three words.

hop run stand

Answ: _______ Relationship: _________________________</td></tr>
</table>

Efforts to measure creativity in the present context have borrowed considerably from existing test batteries that pertain to measure creative thinking processes. Mednick (1962), for example, has developed a test battery that uses *remote associations* in which a test taker is presented with 3 words and asked to find a fourth that is associated with the other three (an example is provided in Table 4). Along a different track, Guilford has developed tests that ask children to identify multiple consequences of an event or multiple uses for a given object such as an elastic band (1974). The evaluation of student answers focuses on the number, diversity, and originality of student responses. Several of these techniques have been incorporated into the test instruments developed for the present undertaking.

Tests in the present study are characterized by a wide range of question types ranging from task work, objective test items, and open-ended questions. Test administration modes also varied between interviews, written tests, and task work requiring mediation by proctors. Particularly in the case of Grade 1 children where written tests are not appropriate, proctors administered all questions in a one-on-one interview format that lasts about 10 to 15 minutes per child. A sampling of questions that represent each domain in the tests used is presented in Table 4 above. In general, critical thinking question items tend to be close-ended while those relating to creative thinking tend to be more open-ended. This has made correction of test items relating to creativity somewhat problematic. Correctors generally establish rather low thresh holds for a "correct" response and are required to alternate correction of such items in order to maximize inter-scorer reliability.

2.4. Data Analysis Methods

Most of the data compiled in this study is quantitative in nature comprising primarily student test scores. All scores have been expressed as percentages to facilitate comparisons. The data analysis techniques employed relied heavily on the use of *t-tests* to determine whether differences in mean scores between experimental and control groups were statistically significant at a probability level of p<.01. Mean differences for all domain scores as well as total test scores were analyzed in this way.

Researchers also looked at correlational interaction between student scores and independent variables relating to sex and age to determine whether these factors had any influence on student performance in the experimental group of students. These analyses used the Pearson's *r* to compute simple correlation coefficients to determine the nature of relationships, if any, between selected variables.

Finally, the study has also included a brief analysis of *effect size* of total mean scores of experimental and control groups. Because of the relatively large sample sizes used in the study, it is likely that even small variations between mean scores will yield significant differences. Some researchers have, therefore, recommended that studies such as this also include an analysis of the *magnitude* of mean differences, that is, the effect size (e.g., Cohen, 1988). Such analyses give some indication of the practical importance of a study's findings and whether the impacts are major or minor. Effect size is calculated by establishing the size of a mean difference relative to the standard deviation of the comparison groups. This may involve taking an average of the standard deviations of two groups or the larger of the two. This study has opted for the latter approach. In general, at probability levels of p<.05, effect sizes which constitute 80% or more of the of the standard deviation of the control group are considered major; those that constitute 50% of the control group standard deviation are considered moderate, and those that constitute 20% are considered minor.

3. Results Section

3.1. Overview

This section presents mean scores by grade for test batteries in critical and creative thinking. Mean scores were calculated and compiled for each of the domains listed in Table 3 as well as for the total overall score for all sections. The presented results also indicate mean differences between experimental and control groups. Starred differences (*) indicate that experimental-control group variations are *not* significant at the probability level of p<.01. Mean score tables also present average age of students to validate assumptions that comparison groups were roughly equivalent with respect to this variable. Other tables in this section summarize effect sizes and correlation analyses of mean scores for all domains with *age* and *sex.* Pearson *r* values that are *not* significant at a probability level of p<.05 are also indicated with a star (*).

3.2. Results in Critical Thinking Domains

<u>Grade 1 Results</u>: A comparison of total scores for experimental and control groups shows a statistically significant difference in favor of the experimental group at a probability level of p<.01 (see Table 5). In this respect, the average score among children in the experimental group was 67% compared with an average score of 53% for the control group. This tendency was true of all 5 domains that were included in the test for Grade 1. These results appeared to occur in spite of the observation that the average age of children in the control group was slightly higher than that of children in the experimental group. The thinking domains that showed the greatest difference between experimental and control groups included *logical sequences, mazes/puzzles,* and *alternating series.* The mean difference be-

tween scores in this respect ranged from 20% to 31% in favor of the experimental group. Thinking domains relating to *classification* and *task analysis* showed less extreme differences though these were still significant at p<.01 and in favor of the experimental group. With respect to best performance in absolute terms, children in the experimental group scored highest in domains relating to *alternating series, classification,* and *mazes/puzzles,* which with the exception of classification tended also to be domains where mean differences were greatest.

Table 5: Comparative Mean Scores and Correlation Coefficients for Grade 1 in Critical Thinking Domains

Test Section	Domain	Correlation Coefficients				Mean Scores		
		Experimental		*Control*		*Experimental*	*Control*	*Mean Difference*†
		Sex	*Age*	*Sex*	*Age*			
1	Alternating Series	-0.01*	0.11	0.02*	0.16	85%	65%	20% (3)
2	Logical Sequences	0.01*	0.11	-0.11	0.12	68%	37%	31% (1)
3	Classification	0.01*	0.16	-0.01*	0.18	81%	73%	8%
4	Mazes & Puzzles	-0.05*	0.16	-0.09*	0.19	74%	51%	23% (2)
5	Task Analysis	0.03*	0.13	-0.05*	0.19	36%	27%	9%
Total Score		**0.01***	**0.19**	**-0.06***	**0.25**	**67%**	**53%**	**14%**
Standard Deviation		**--**	**--**	**--**	**--**	**19.76%**	**19.57**	**--**
Average Age		**--**	**--**	**--**	**--**	**6.9 yrs**	**7.6 yrs**	**0.7 yrs**

Experimental Group: N=642 children; Control Group: N = 490

†Significant at p<.01 unless otherwise indicated; *Correlation coefficients *not* significant at p<.05; (-) indicates rank of mean difference

An analysis of the average age in comparison groups using a *t-test* revealed a significant difference between groups at p<.01. The mean difference in average age between the two groups was 0.7 years, thereby dashing an expectation for relative age equivalence. Nevertheless, though the control group manifested an older average age, this did not appear to work to the advantage of the group with regard to overall performance.

With few exceptions, correlation analyses of performance patterns among children in comparison groups showed no relation with *sex.* With respect to total mean scores for the entire test, the calculation of the Pearson product moment coefficient of correlation yielded r = .01 for the experimental group and r = -0.06 for the control group neither of which was significant at p<.05. This outcome was true for all sex-mean score comparisons for each individual cognitive domain with the exception of *logical sequences* which did show a significant albeit weak correlation with sex for the control group.

In contrast to sex, *age* showed significant relationships with mean performance scores in all cases. In this respect, the experimental group yielded a correlation of r = 0.19 for total mean score and age while the control group yielded a correlation of r = 0.25; both were significant at the p<.05 level. With respect to individual cognitive domains, correlations tended to be strongest for tasks involving *classification* and *mazes/puzzles* for the experimental group (r = 0.16) and *classification, mazes/puzzles,* and *task analysis* for the control group (r = 0.18, r = 0.19, and r = 0.19, respectively).

<u>Grade 5 and 6 Results</u>: Written tests at the Grade 5 and 6 level for critical thinking items yielded total mean scores that showed significant differences at p<.01 for both Grades 5 and 6 in favor of the experimental group (see Table 6). Mean differences stood at 6% for Grade 5 and 7% for Grade 6. To be sure, however, this outcome was not true across all domains. Total mean scores for Grade 5 children who studied in CFS classrooms came out at 39% against 33% for the control group. Among Grade 6

children, the experimental group mean score stood at 47% against 40% for the control.

Domains that showed the greatest differences between comparative groups that were also in favor of the experimental group included *textual analysis* and *classification* for Grade 5 and *analogies, figural series,* and *classification* for Grade 6 children. Average mean differences favoring the experimental grouup for both grades were greatest for *textual analysis* and *classification.* In this regard, mean differences ranged from 12% to 13.5%. On the other hand, the difference in mean scores for *mazes/puzzles* was not significant for Grade 5 children nor was there any significant difference in mean scores for domains in *concept analysis* or *application of principles* among Grade 6 children. Indeed, the mean difference for the latter domain was significantly in favor of the control group by a large margin (22%) for Grade 5. This outcome was clearly in opposition to what was expected.

In terms of absolute scores, children in experimental conditions appeared to do best in domains such as *classification* and *concept analysis* at Grade 5 and *figural series, classification, analogies, mazes/puzzles* and *concept analysis* at Grade 6. Test scores in this respect all tended to be 50% or more. With the exception of *concept analysis* and *mazes/puzzles* at Grade 6, these were also the same domains where experimental group children exhibited the largest mean differences in test scores with control group children as well. Domains where children in experimental conditions did most poorly included *logical sequences* and *application of principles* for Grade 5 children (scores of 30% or less) and *logical sequences, textual analysis,* and *application of principles* among Grade 6 children (scores of 40% or less).

T-test analyses for mean age differences between comparative groups indicated that there was no significant difference for Grade 5 at the probability level of p<.01. The mean difference between comparative groups in this regard was 0.02 years. In contrast, the average age difference among Grade 6 children was 0.5 years, which was found to be a significant difference at p<.01, thereby indicating group composition was not equivalent by age among the Grade 6 sample.

Table 6: Comparative Mean Scores for Grades 5 and 6 in Critical Thinking Domains

Test Section	**Domain**	**Mean Scores**						**Average Mean Difference**
		Grade 5			**Grade 6**			
		Experi-mental	*Control*	*Mean Difference*†	*Experi-mental*	*Control*	*Mean Difference*†	
1	Figural Series	44% (3)	36%	8%	52%	41%	11% (1)	9.5%
2	Logical Sequences	28%	20%	8%	36%	32%	4%	6%
3	Classification	55% (1)	41%	14% (3)	59% (1)	49%	10% (2)	12% (2)
4	Analogies	43%	38%	5%	55% (2)	44%	11% (1)	8%
5	Concept Analysis (Fact/Opinion)	51% (2)	41%	10%	52%	52%	0%*	5%
6	Mazes & Puzzles	37%	36%	1%*	53% (3)	47%	6%	3.5%
7	Textual Analysis	34%	16%	18% (2)	32%	23%	9%	13.5% (1)
8	Application of Principles	25%	47%	22% (1)	36%	36%	0%*	11% (3)
Total Mean Score		**39%**	**33%**	**6%**	**47%**	**40%**	**7%**	**6.5%**
Standard Deviation		**11.93%**	**11.94%**	**--**	**13.34%**	**13.19%**	**--**	**--**
Average Age		**13.17 yrs**	**13.15 yrs**	**0.02 yrs***	**14.1 yrs**	**14.6 yrs**	**0.5 yrs**	**0.26 yrs**

Grade 5 Experimental Group: N = 363; Grade 5 Control Group: N =359
Grade 6 Experimental Group: N =319; Grade 6 Control Group: N =319
†Significant at p<.01 unless otherwise indicated; *Difference is *not* significant at p<.01; (-) indicates rank of mean difference

Correlation analyses found no significant relationship between *sex* and overall test performance for either comparison group at Grade 5 (see Table 7). In both cases, the calculation of the Pearson product moment coefficient of correlation yielded *r* = -0.03, which was not significant at the probability level of p<.05. Among Grade 6 sub-samples, however, this outcome held true only for the control group with *r* = -0.03. For the experimental group, the resulting correlation coefficient yielded a value of *r* = -0.15, which was significant at p<.05. This outcome, though weak, indicated that boys in the Grade 6 experimental group tended to outperform girls in most tasks with the exception of *analogies*, *textual analysis,* and *application of principles.*

Table 7: Correlation Coefficients for Mean Scores and Age/Sex for Critical Thinking Domains

Test Section	Domain	Grade 5				Grade 6			
		Experimental		*Control*		*Experimental*		*Control*	
		Sex	*Age*	*Sex*	*Age*	*Sex*	*Age*	*Sex*	*Age*
1	Figural Series	-0.03*	-0.06*	-0.09*	-0.16	-0.18	-0.17	-0.05*	-0.11
2	Logical Sequences	0.01*	-0.13	0.05*	-0.18	-0.15	-0.03*	0.00*	-0.19
3	Classification	-0.01*	-0.17	-0.08*	-0.20	-0.15	-0.24	-0.22	-0.01*
4	Analogies	-0.02*	-0.16	0.02*	-0.24	-0.02*	-0.26	0.04*	-0.15
5	Concept Analysis (Fact/Opinion)	0.00*	0.01*	0.06*	0.03*	-0.14	-0.07*	0.09*	-0.10*
6	Mazes and Puzzles	-0.09*	-0.03*	-0.17	-0.01*	-0.16	-0.09*	-0.07*	-0.06*
7	Textual Analysis	0.09*	-0.25	-0.03*	0.06*	-0.05*	-0.35	0.06*	-0.20
8	Application of Principles	-0.13	-0.12	0.00*	-0.18	-0.05*	-0.16	0.08*	-0.19
Correlation w/Total Mean Score		**-0.03***	**-0.24**	**-0.03***	**-0.25**	**-0.15**	**-0.31**	**-0.03***	**-0.24**

Grade 5 Experimental Group: N = 363; Grade 5 Control Group: N =359
Grade 6 Experimental Group: N =319; Grade 6 Control Group: N =319
*Difference is *not* significant at p<.05

As was true of Grade 1 children, *age* tended to show a more robust relationship with test performance mean scores than was true of sex. In this respect, correlation analyses for Grade 5 yielded *r* = -0.24 and *r* = -0.25 for experimental and control groups, respectively. For Grade 6, the result was *r* = -0.31 and r = -0.24 for the experimental and control group, respectively. These *r* values were significant at the p<.05 level. In contrast to Grade 1, however, the negative *r* values indicated an inverse relationship between age and test performance; that is, younger children tended to perform better than older children. This inverse relationship was true across all comparative groups. Nevertheless, there were several domains in which age did not manifest a significant relationship with test performance including *concept analysis* and *mazes/puzzles.* This too was consistently true across all groups.

3.3. Results in Creative Thinking Domains

<u>Grade 5 and 6 Results</u>: Analysis of mean differences between comparative groups yielded the result that mean scores among the experimental group were significantly higher than those of the control group at p<.01 in both Grades 5 and 6. In addition, these statistically significant differences were consistent at the same probability level across all domains (see Table 8). In this respect, the total mean score for Grade 5 was 38% for the experimental group as against 22% for the control group and 40% for the experimental group at Grade 6 as against 23% for the control group. Total mean score differences were16% at Grade 5 and 17% at Grade 6. Among the various thinking domains defining creative thought, mean differences were greatest for *determining consequences* and *figural elaboration* at Grade 5 and for *determining consequences, figural elaboration,* and *remote associations* for Grade 6.

Among the various domains, the highest absolute scores for the experimental group also tended to be the same domains with highest mean differences with the control group. Domain scores that were highest among experimental group children included *figural elaboration, determining consequences,* and *simile interpretation* at Grade 5 and *determining consequences, unusual uses,* and *figural elaboration* at Grade 6.

Table 8: Comparative Mean Scores for Grades 5 and 6 in Creative Thinking Domains

Test Section	Domain	Mean Scores					
		Grade 5			Grade 6		
		Experi-mental	*Control*	*Mean Difference*†	*Experi-mental*	*Control*	*Mean Difference*†
1	Determining Consequences	55% (2)	20%	35%	65% (1)	30%	35%
2	Unusual Uses	49%	32%	17%	54% (2)	37%	17%
3	Word Fluency	20%	16%	4%	29%	14%	15%
4	Expressional Fluency	40%	46%	6%	40%	46%	6%
5	Associational Fluency	43%	24%	19%	41%	21%	20%
6	Figural Elaboration	64% (1)	24%	40%	54% (2)	28%	26%
7	Simile Interpretation	52% (3)	33%	19%	43%	29%	14%
8	Remote Association	22%	3%	19%	24%	2%	22%
Total Mean Score		**38%**	**22%**	**16%**	**40%**	**23%**	**17%**
Standard Deviation		**17.03%**	**13.21%**	**--**	**15.43%**	**10.22%**	**--**
Average Age		**13.17 yrs**	**13.15 yrs**	**0.02 yrs***	**14.1 yrs**	**14.6 yrs**	**0.5 yrs**

Grade 5 Experimental Group: N = 363; Grade 5 Control Group: N =359
Grade 6 Experimental Group: N =319; Grade 6 Control Group: N =319
†Significant at p<.01 unless otherwise indicated; *Difference is *not* significant at p<.01; (-) indicates rank of highest scores

Analyses using the Pearson product moment coefficient of correlation again generally found no significant relationship between overall mean score and *sex.* (see Table 9). On the whole, this was also true of most domain mean scores as well. The Grade 5 experimental group was an exception to this observation, however, with r = 0.11 for total mean score and r = 0.15 for *determining consequences* and r = 0.14 for *simile interpretation.* These relationships were significant at the probability level of p<.05. Though these relationships are weak, they suggest a performance advantage among girls. Other correlation coefficients calculated for total mean score/individual domain scores and sex found no significant relationships.

Table 9: Correlation Coefficients for Mean Scores and Age/Sex for Creative Thinking Domains

Test Section	Domain	Grade 5				Grade 6			
		Experimental		*Control*		*Experimental*		*Control*	
		Sex	*Age*	*Sex*	*Age*	*Sex*	*Age*	*Sex*	*Age*
1	Determining Consequences	0.15	-0.08*	0.02*	-0.22	-0.02*	-0.20	0.00*	-0.22
2	Unusual Uses	0.04*	-0.11	-0.04*	-0.20	-0.01*	-0.19	0.02*	-0.12
3	Word Fluency	0.09*	-0.20	0.07*	-0.32	0.03*	-0.14	0.05*	-0.17
4	Expressional Fluency	0.03*	-0.26	-0.03*	-0.21	0.04*	-0.15	0.06*	-0.15
5	Associational Fluency	0.07*	-0.17	-0.04*	-0.20	-0.04*	-0.16	-0.09*	-0.15
6	Figural Elaboration	0.02*	-0.11	-0.07*	-0.01*	-0.06*	-0.19	0.04*	-0.15
7	Simile Interpretation	0.14	-0.09*	-0.03*	-0.21	0.00*	-0.14	0.10*	-0.09*
8	Remote Association	-0.02*	-0.17	0.07*	-0.02*	-0.08*	-0.05*	-0.01*	0.03*
Total Mean Score		**0.11**	**-0.24**	**-0.01***	**-0.31**	**-0.03***	**-0.23**	**0.04***	**-0.23**

Grade 5 Experimental Group: N = 363; Grade 5 Control Group: N =359
Grade 6 Experimental Group: N =319; Grade 6 Control Group: N =319
*Difference is *not* significant at p<.05

In contrast to sex, inverse relationships of some strength were again in evidence for analyses of relationships between *age* and mean score performance. The inverse nature of the relationships indicates that younger children exhibited a performance advantage. At Grade 5, correlation analyses of age and total mean score yielded r = -0.24 for the experimental group and r = -0.31 for the control group; at Grade 6, these analyses yielded r = -0.23 for both the experimental and control groups. These relationships were significant at $p<.05$. With few exceptions, age evidenced a highly uniform, significant relationship with individual mean scores for most thinking domains as well. Notable exceptions in this regard included non-significant correlation coefficients for *determining consequences, simile interpretation,* and *remote association* among experimental groups and *figural elaboration, remote association*, and *simile interpretation* among control groups.

3.4. Analysis of Effect Sizes with respect to Total Mean Differences

The computation of effect sizes relating to mean differences for total mean scores relative to the largest standard deviation for each group is summarized in Table 10. According to this analysis, the effect size for findings on the Grade 1 critical thinking test was 0.71 thereby indicating a moderate to major impact. Effect sizes for critical thinking tests in Grades 5 and 6 were smaller at 0.50 and 0.52, respectively. These results nevertheless suggest moderate impacts on children's ability to complete tasks relating to critical thinking.

Table 10: Summary of Effect Sizes for Total Mean Scores, All Grades

Grade	Critical Thinking			Creative Thinking		
	Mean Difference	***Standard Deviation Used***	***Effect Size***	***Mean Difference***	***Standard Deviation Used***	***Effect Size***
1	14%	19.76%	0.71	--	--	--
5	6%	11.94%	0.50	16%	17.03%	0.94
6	7%	13.34%	0.52	17%	15.43%	1.10

Effect sizes appeared to be greatest on tests that pertained to measure creative thinking. In this respect, the effect size for Grade 5 was 0.94 while for Grade 6 the effect size was 1.10. These effect size magnitudes imply impacts on children's learning that are large although in absolute terms, percentage scores among both groups were relatively low. Because the mean difference for all total mean scores were in favor of the experimental group, these results suggest that impacts on those children who received program interventions were not only significant but also moderate to major in magnitude.

4. Discussion and Interpretation of Results

4.1. Overview of Findings

Analyses of test performance tentatively suggest that expectations for program interventions to have a positive impact on higher order thinking skills among children in CFS classrooms are being fulfilled. In all cases, total mean scores among experimental group children were higher than those for control groups of comparable composition and that these differences were statistically significant with a very low probability of error. As importantly, effect sizes indicated that the differences in total mean scores between comparison groups were moderate to major. Both in absolute terms as well as with respect to effect sizes, mean differences were greatest among children at Grade 1with respect to critical thinking and among children at Grades 5 and 6 with respect to creative thinking. With few exceptions, experimental group subjects outperformed those in control groups across most of the 18 cognitive domains selected for inclusion in test batteries.

Among the critical thinking domains that were tested, *classification* skills tended to emerge as an area where children in experimental groups performed best. Indeed, among Grade 1 children, this domain had a rank score of 2 and among Grade 5 and 6 children, a rank score of 1. In addition, Other cognitive domains where children in experimental groups performed well included *concept analysis, analogies,* and *mazes and puzzles.*

This study did not venture any predictions as to how *age* and *sex* might be related to test performance, only that there might be some relationship between them and the dependent variable, which was defined as test performance. In the actual event, correlation analyses did not generally suggest any significant relationship between test performance and sex though there were several exceptions to this observation. This was true of both experimental and control groups. In the few cases where significant relationships were observed, they tended to be weak with magnitudes in a range of 0.11 to 0.15. Interestingly, when these exceptions occurred in critical thinking domains, they tended to indicate a slight performance advantage for boys; where they occurred in creative thinking domains, they tended to indicate a slight performance advantage for girls.

In contrast, age tended to show a consistent relationship with test performance though this was not always in line with what one might expect. Among children at Grade 1 level, correlation coefficients exhibited significant albeit weak relationships with age in the range of 0.19 to 0.25. These results suggested that older children tended to perform slightly better than younger children. Similarly significant relationships also emerged in the higher grades but these were in a direction reverse to what was observed among Grade 1 children. Without exception, correlation coefficients were negative indicating that younger children had a performance advantage. These correlation coefficients tended to be slightly greater than those observed in the lower grades and ranged from –0.23 to –0.31 among both experimental and control groups.

4.2. Implications, General Constraints, and Conclusions

This study has been a first step for Cambodian educators to assess the impact of program interventions involving Child Friendly Schools on learning outcomes for target groups. Preliminary findings have been both highly positive and very encouraging. Nevertheless, researchers would stop short of making definitive conclusions regarding impact on learning due to the pre-experimental nature of the research design. Although experimental and control groups were roughly matched for equivalence on criteria of relevance including local poverty rates, PCR, and demographic factors, the assignment of members to comparison groups was not statistically random.

References

(1) Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, Hillsdale, N.J.: Lawrence Erlbaum Associates.
(2) Guilford, J.P., et al. (1974) *Structure of Intellect Abilities.* Orange, CA: Sheridan Psychological Series.
(3) Hopkins, J. and Chaimuangdee, K. (2000) *Making Schools More Child Friendly: Lessons from Thailand.* Bangkok: UNICEF.
(4) Lockheed, M.E., and Levin, H.M. (1988) *Effective Schools in Developing Countries*, Levin, H.M. and Lockheed, London: Falmer Press.
(5) Mednick, S.A. (1962) *The associative basis of the creative process.* Psychological Review, 69, 220-232.
(6) Wheeler, C. (1998) *Rebuilding Technical Capacity in UNICEF/Sida Supported School Clusters*, Phnom Penh, Cambodia: UNICEF/Sida.