The AGEM Pilot
at
Maria Niketan School, Bangalore, India

Samar Singh, PhD
Email: samar@agem.in
www.agem.in

November, 2009 to March, 2010

Reading this report

1. On the next page is an executive summary which is extensively cross referenced to the main body of the report.
2. The cross references are indicated in red e.g. Section 7.1, which takes the reader to the last section on Conclusions.
3. The reader may return to this page by using the Previous View icon on the Adobe Reader which looks like this: 
4. The Table of Contents, List of Figures and List of Tables are similarly hyperlinked.

Part I

Executive Summary

1 Brief Statement of the Pilot

1.1 Overview

We, at AGEM, have been researching and implementing technologies, methods and processes to support the growth of meaningful education in the Indian context in India for more than a decade. Recently, we conducted a Pilot Study at the Maria Niketan School in Bangalore, with a view to measuring the impact of our technologies, processes and methods on an experimental group of 22 students in a Grade 6 Class comprising 76 students using Maths as the object of our research. Although we could only get about 30 hours of student contact for teaching Maths over a three month period between Nov., 2009 and March, 2010, the impact seems to have been significant.

The primary characteristics of our teaching approach lay in using pre-prepared multimedia content with embedded questions, which were answered by students in parallel using our wireless keyboards with a near immediate representation on a projected display of the class response in various forms including frequency charts, word frequencies and concordances. Student responses were archived and later analysed to acquire insights into what makes students perform, how our materials could be improved etc.

We sought to measure the impact of our teaching by providing the experimental group a pre-test comprising 113 test questions. At the end of the course a post-test was taken by the experimental group and also by an additional group of 41 students from the same class. The average score for the experimental group went from 24% to 75% while the group of 41 students had an average of 61%.

We also compared the School's Term 2 examination scores with the Term 1 scores as we had commenced our work and finished it within Term 2. This analysis indicated that in Maths the average increase in the experimental group was 7.5% while for the remaining 53 students who took the same examinations, the average increase was 2.7%. Any impact that we had on these scores was unlikely to be related to content, as the text book for Term 2 shows greatly advanced content compared to our fundamental lessons on issues such as the concept of number culminating with the subject of fractions.

1.2 Objectives of the Pilot at Maria Niketan School

Our objectives in conducting this program (Section 4.3) were to:

- establish the robustness of our hardware, software and processes in a school environment
- acquire performance data and determine the efficacy of our systems in a specific discipline i.e. maths
- determine on a prima facie basis, how the system impacts other aspects of the classroom e.g. classroom management.
- gather research data to answer other questions such as the differences in the nature of responses from students with different attributes, e.g. gender, age etc.

1.3 Stages of development of the AGEM system

1. 1998 - 2000: First generation hardware and systems comprising infra red clickers. (Section 3.4.1)
2. 2006: Second generation radio frequency clickers for HAAP Project (Section 3.4.2)
3. 2007-2008: USB Keyboards and radio frequency keyboards. (Section 3.4.3)
4. 2009-2010: The Maria Niketan Pilot with radio frequency keyboards. (Chapters 4, 5, 6)

1.4 The AGEM system

1.4.1 Characteristics

Section 2.1

- Feedback systems - one per student - to allow parallel input by individual students to pre-designed questions
- Topic Plans comprising video, audio, instructional content, and embedded questions representing small increments of learning.
- Immediate, short term and long term feedback of performance based on analysis of embedded questions
- Deficiencies in response of individual students are never made public. All representations displayed are of class response
- Supplementary mechanisms such as activities, projects and reflection sessions.

1.4.2 Principal design objectives

- Serving large classes typically of 20 to 100 students.
- Providing for a large student base e.g. dozens of classes in hundreds of schools
- Making differentiated instruction practical
- Serving the need for scientific feedback of performance for:
  - design of Topic Plans
  - school management
  - reporting to parents.

1.5 The Maria Niketan School Pilot

The experimental group consisted of 22 students (13 male and 9 female) who had been assigned to an extra-curricular remedial maths program which we took over. The control group constituted the rest of the class which was not in the remedial maths program. This control group comprised an additional 41 students.

The test questions for the pre-test were a set of 113 questions selected from the Number Worlds website. Regrettably this was not provided to the control group. The post test undertaken by both groups was a 65 question subset of the 113 questions comprising the pre-test. (See Section 5.2.4)

The instructional component comprised 4 Topic Plans called TP2, TP3, TP4 and TP5 which extended across several classroom sessions. Each Topic Plan contained between 30 and 51 embedded questions which students answered in parallel using the wireless feedback devices. Hence, it was possible to obtain scores for each Topic plan.

The program with the 22 students of the Class 6 remedial maths group serving as the experimental group and the rest of 41 students serving as a control group commenced on 23 Nov 2009. The time spent was (Section 5.2.2) as follows:

Time in hours	Topic
3.75	Pre-test with experimental group
18.75	Topic Plans 1 to 5
11.25	Other activities
2.1	Post test for full group
35.85	Total Program

Table 1.1: Nature of Program and Time spent

1.5.1 The outcomes of the pre and post testing

Statistical summary (Section 6.3.2)

	Test Score Statistics (%)			
	Mean	**Max**	**Min**	**Std. Dev.**
Experimental Group - Pre Test	24	50	5	9
Experimental Group - Post Test	75	89	46	12
Control Group - Test at end of session	61	82	35	9

Table 1.2: Comparison of group statistics for pre-test and post-test experimental group and the control group

The normalized gain for the experimental group (Section 6.3.3)

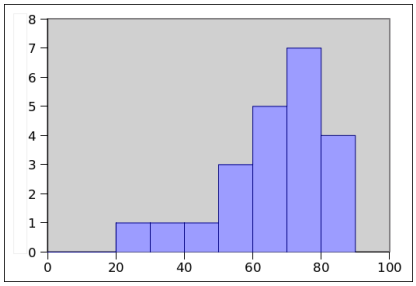


Figure 1.1: Frequencies of Normalized gain for Experimental Group - pre and post test comparison

1.5.2 Outcomes across topic plans

Analysis of scores based on questions embedded within the Topic Plan

These results show a declining gap between the best and worst scores over successive Topic Plans. At the same time the mean score also increased over the Topic Plans. (Figure 6.3)

Analysis of Topic Plan scores against ratings of inferential ability based on post-test questions

Four students showed significantly lower normalized average scores across Topic Plans 2 to 5 as compared to averages for these Topic Plans. (Section 6.3.5)

The post-test questions were categorized into Calculation, Inference, Observation and Reporting dominant categories. It was found that the same four students had scored the lowest four scores in the inference category in the post-test.

However, the sample is too small to draw conclusions other than to say that more research into this area may be justified.

Other outcomes

1. In terms of pre and post test comparisons for the experimental group, the least gains were made for questions that lay within an inference categorization. (Section 6.4.2)
2. There was no evidence of correlation of performance with age in days in post test scores. (Section 6.4.3)
3. In general, in the post test the experimental group scored higher in all four categories of questions than the control group. (Section 6.4.3)
4. There was little difference in the scores between male and female students for post test scores within the experimental group. (Section 6.4.5)
5. The behavior of experimental group students in the classroom was exemplary with very few instances of distraction in spite of very high background noise. (Section 5.3.1)

6. The ability for communicative expression seemed to correlate with performance on the Topic Plan embedded question scores. However, further work is needed before any conclusions can be drawn. (Section 5.3.3)

1.5.3 Maria Niketan's end of term examinations

The school conducted one set of examinations internally at the end of Term 1 around October 2009. These concluded before we commenced our engagement with the school. Term 2 examinations were conducted shortly after we completed our program in March/April 2010. We were of the opinion that our program may impact the results of the Term 2 examinations in Maths, English and Science. We asked for and received individual student results for both Term 1 and Term 2 exams. We have no idea of what was taught or examined or the nature of the examination process. We do know that the three subjects being considered had 2 components each - Written and Oral.

It appears there are 75 students who took the Term 1 and 2 examinations out of a class strength of 76. Only 65 students attended our post test examinations which form the basis for our pre and post test assessment described earlier. It is assumed for this exercise that the control group comprised the 53 students who were not part of the experimental group.

Change in Scores between Term 2 (After AGEM Program) and Term 1 (Before AGEM Program)

Characteristic	Group (No. of Students)	Change (Term 2-Term1) scores		
		English	Math	Science
Average Change	Experimental Group (22)	9.9	7.5	13.8
	Control Group (53)	9.4	2.7	10.3
No change instances	Experimental Group (22)	2	1	1
	Control Group (53)	6	10	4
Negative change	Experimental Group (22)	3	4	3
	Control Group (53)	13	16	12

Figure 1.2: Difference of Term 2 and Term 1 School examination scores - Experimental and Control groups

1.6 Primary conclusions

Section 7.1

- Hardware, software, methods and practices seem to work reliably after fine tuning as a result of the early stages of the pilot.
- Pre and Post test results show significant increments for the experimental group. Experimental group performance exceeds that for the remainder of the class.This has to be taken in the context that the experimental group was a remedial class for maths.
- Classroom behavior of the experimental group and of the control group during the testing process was exemplary. This could have been the result of the inclusiveness generated by the keyboards.
- There were very few instances within the experimental group where students did not respond to Topic Plan questions.
- Within some areas of the topic plans there could have been a slower progression to benefit the group as a whole.
- It is difficult to observe the performance of the class as a whole and as individual for assessing performance. The capture of performance data and its analysis has been invaluable in making possible the progress that the outcomes depict.

Part II

The Maria Niketan School Pilot

Contents

I Executive Summary 2

1 Brief Statement of the Pilot 3
- 1.1 Overview . . . 3
- 1.2 Objectives of the Pilot at Maria Niketan School . . . 3
- 1.3 Stages of development of the AGEM system . . . 3
- 1.4 The AGEM system . . . 4
- 1.5 The Maria Niketan School Pilot . . . 4
- 1.6 Primary conclusions . . . 6

II The Maria Niketan School Pilot 7

2 Introduction 11
- 2.1 The AGEM process . . . 11
- 2.2 Caveats . . . 11

3 Setting the stage 12
- 3.1 The need for reform of the educational system . . . 12
- 3.2 A sad anomaly . . . 12
- 3.3 Prelude to this work . . . 12
- 3.4 Personal milestones of development during the last twelve years . . . 12
- 3.5 Experiences and Learnings . . . 15

4 Designing the Pilot Program 19
- 4.1 The rationale for maths . . . 19
- 4.2 The overarching goals of the Maths Program at Maria Niketan . . . 19
- 4.3 Objectives . . . 19
- 4.4 Challenges . . . 19

5 Conducting the Pilot Program 20
- 5.1 First Phase - Grade 5 students (9 Nov 2009 - 13 Nov 2009) . . . 20
- 5.2 Second Phase - Grade 6 students (23 Nov 2009 - 15 March 2010) . . . 20
- 5.3 Learnings . . . 25
- 5.4 Challenges . . . 27
- 5.5 Errors . . . 28

6 Analyzing the data 30
- 6.1 The nature of the data collection process . . . 30
- 6.2 The sources of data . . . 30
- 6.3 The Primary observations . . . 31
- 6.4 Secondary observations . . . 36
- 6.5 Maria Niketan's end of term examinations . . . 41
- 6.6 Limitations of this study . . . 42

7 Conclusions 44
- 7.1 Primary issues . . . 44
- 7.2 Addressing the original objectives . . . 45

List of Figures

1.1 Frequencies of Normalized gain for Experimental Group - pre and post test comparison 5
1.2 Difference of Term 2 and Term 1 School examination scores - Experimental and Control groups 6

3.1 Vidyashilp Academy, Bangalore using first generation infra red feedback devices in 1999 13
3.2 Children evaluating the film produced by the HAAP project 13
3.3 Students using wired parallel feedback devices 14
3.4 Killing creativity through disciplinary processes 17
3.5 Attentivity in an AGEM classroom 17

5.1 Reflection session anomalies 24
5.2 Sixty three students in the AGEM classroom 25
5.3 A section of the experimental group during a re-run of a Topic Session 26
5.4 Correlation between communicative ability and Math scores in Topic Plans 2 and 5 27

6.1 Representation of group statistics for pre-test and post-test experimental group and the control group 32
6.2 Frequencies of Normalized gain for Experimental Group 33
6.3 Progression of scores and extremes of scores across Topic Plans 34
6.4 Individual student comparison of normalized average of Topic Plan Question Scores against average of Topic Plan 2,3,4 and 5 question scores 35
6.5 Post-test breakup per student into C, I, O and R categories and Post-Test Score 36
6.6 Correlation of age of student with post-test scores across question categories 37
6.7 Representation of trends of socio economic levels for male and female students using post-test results for experimental group 40
6.8 Topic Plan 5 Average score per student for Questions 2 to 52 41
6.9 Difference of Term 2 and Term 1 School examination scores - Experimental and Control groups 42

List of Tables

1.1 Nature of Program and Time spent 4
1.2 Comparison of group statistics for pre-test and post-test experimental group and the control group 5

5.1 Nature of Program and Time spent 20
5.2 Improvements in communicative expression 27

6.1 Comparison of group statistics for pre-test and post-test experimental group and the control group 31
6.2 Comparison of pre-test and post-test scores for the experimental groups across question categories 37
6.3 Comparison of control and experimental groups across question categories 38
6.4 Teacher estimate of ability compared to post test scores of experimental group 38
6.5 Gender comparison of question categories using post test results 39
6.6 Comparison of gender and Socio-economic status using post-test scores of experimental group 39

2 Introduction

The purpose of the AGEM Pilot was principally to quantify any gains that may occur in using AGEM technologies and methods to deliver instruction in the classroom.

2.1 The AGEM process

The principal levers of the AGEM process are:

1. Feedback systems: These imply a device - one per student - which allows each student to provide alphanumeric inputs in parallel to questions.
2. Pre-planned Topic Plans which are usually multi-session areas of content which include embedded questions, multi-media materials, and instructional content presented in screens.
3. Analysis mechanisms which are used to provide immediate feedback of the performance of the class in various graphical formats including histograms immediately after a question has been answered, and subsequent short term, medium term and long term analysis.
4. There is no public display of deficient responses although outstanding responses are occasionally identified.
5. Supplementary mechanisms such as activities, projects and reflections.

The AGEM process has been designed to cater to:

- large classrooms as a minimum group of about 20 is needed to provide representative feedback. The technology has no limit of maximum numbers.
- the needs of a large population of classrooms so that teachers can be relieved of the tedium of preparing materials and of making assessments.
- the needs of students so that differentiated instruction becomes possible without protest from the more talented students in the same classroom, and without making an individual child's errors a public issue.
- the needs of parents for objective evidence on how their child is faring in school and how to identify the talent that lies within the child.
- the needs of school managements to know how to enhance the learning levels of the school as a whole.
- the needs of society to be able to identify what works and what does not and to remediate that from a central resource.

2.2 Caveats

1. The AGEM process is not designed for schools which are well resourced, with teachers who are well trained, well paid, motivated, caring, constantly able to upgrade their skills and knowledge, and supported by a management that is dedicated to an education process which meets the needs of academia and the workplace.
2. While we did experiment with activities, simple projects and reflection sessions, these were not within the ambit of this pilot and although the learnings from those were used for designing our Topic Plans, these instruments are not analyzed in this report.

3 Setting the stage

3.1 The need for reform of the educational system

The Internet has changed many things, including how we live and what we value. At one time, the access to information was often equated to knowledge. Today, the value of a unit of information has reduced drastically.

This dramatic change has taken place in a very short time. Many careers have almost ceased to exist such as airline ticketing agents and other forms of brokerage. Many new disciplines have surfaced from Remote Sensing to Nanotechnology. Above all is the widespread perception that we live in a rapidly changing world. That perception is least apparent in our schools.

In India, this problem is intensified by the inability of many of our children to get access to any type of formal education. The Right To Education Bill is a recent enactment, and at this time, that right is far from realization.

3.2 A sad anomaly

Strangely, the one million schools in our country which represent a source of immediately available infrastructure, are not open to opening their doors for an after-school shift for those who have no access to education. Even more strangely, the most elite schools with the best infrastructure are least amenable to any such idea.

3.3 Prelude to this work

Twenty five years of research, investigation into maritime casualties, doctoral work on cognitive engineering, training school teachers in computers in British Schools in Hong Kong, and years of working with maritime simulators, engendered five principal realizations:

- There is a need for classroom learning to transfer into the academic, vocational and workspace domain.
- Immediacy of feedback of performance is a vital factor in learning, as is reflecting on, or debriefing that performance. One could say that feedback delayed is learning denied
- Grades 3 to 8 are the most cost effective means to remediate educational systems in the first instance. Of course the most promising years are birth to Grade 2 but solutions in that segment seem difficult on a large scale.
- Interaction with students, and performance measurement during delivery in the classroom, are the keys both to learning and to the measurement of learning. It may be unrealistic to expect school teachers to perform both these functions.
- Competencies are a function of education and training and only rarely of individual ability. Ability does of course impact the development of creativity and expertise. Hence, measurement must be a tool for remediation not filtering.

3.4 Personal milestones of development during the last twelve years

3.4.1 The Vidyashilp Academy experience in Bangalore, 1998-2000

We introduced our first generation of feedback technology into 6 classrooms at Vidyashilp Academy, Bangalore when I collaborated with Innomedia to create infra-red “Clickers” in 1999. In a parallel development we introduced cable based systems to display Topic plan materials in the classrooms.

We were also inspired by an MIT media labs project to produce the students desks shown in Figure 3.1 to generate collaborative learning environments. These desks and the “Clickers” are visible in the figure below:

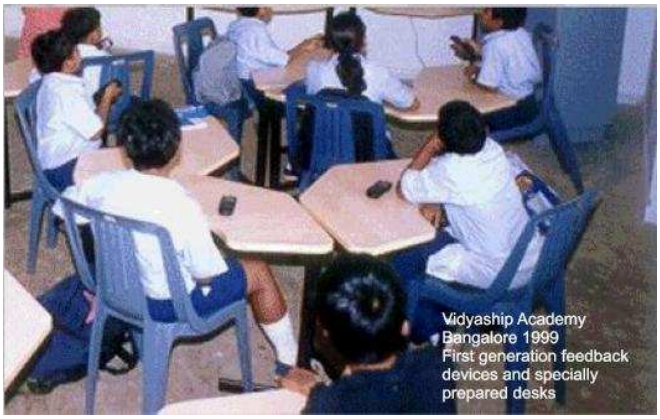


Figure 3.1: Vidyashilp Academy, Bangalore using first generation infra red feedback devices in 1999

We also introduced the idea of Projects with detailed instructions for both the conduct and the measurement of projects. These measures were popular with students and parents but not with teachers, who saw a loss of control. That control had already been lost with the idea of interactive classrooms. On the positive side, teachers were so affected by the impact of the technology on making classrooms easier to manage and the ease with which student interest was maintained that there was almost no resistance to the introduction of technology, as opposed to other reforms. In fact, when one teacher was asked, “Are the students learning from the system”, she replied, “Never mind what the students are learning. Ask how much I am learning”.

Topic plans at Vidyashilp were made through teachers working in groups researching content and providing a template and Internet based content to a computer operator who put them together to generate proprietary markup scripts which were rendered on TV screens in the classroom. I never got around to analyzing the data gathered before I parted ways with Vidyashilp in 2000.

3.4.2 The HAAP experience - 2006

Shonu Chandra and I, conducted the HAAP project with an NGO. During this project, we used the assistance of Karthik Ayyar, a member of the AGEM team to build a set of radio frequency wireless Clickers in 2006. The HAAP project required measuring the response of children in five states to a film produced by children to reduce stigma and discrimination against those infected or affected by HIV. The results of this work are in the document DistractorAnalysis.pdf on the AGEM website under Resources.

Figure 3.2: Children evaluating the film produced by the HAAP project

This demonstrated the power of giving children a say in initiatives that involve children. The film had been criticized by some experts on various accounts, almost all of which were negated by the results of the exercise. Children answered a set of questions prior to the film, again during the film through stopping the film at the relevant points, and finally by answering all

questions again at the end of the film. The children also commented on various aspects that had been the subject of critical comment. The pre and post responses show an important increment of learning with different children in four different states of India.

Another important learning was the value of measuring simple instances of distraction. These helped us understand the parts of the film that were least acceptable to the children. Interestingly, these parts were introduced after the children's version of the film had been made, to please HIV/AIDS administrators who felt the need to use the film to introduce HIV related government services. It is possible to conclude that such decisions should be taken in the wider context of making the film attractive to children as it deals with an important issue of stigma and discrimination for those affected or infected by HIV. There should be some other medium for creating awareness of government services.

3.4.3 The Trio School experience - 2007/2008

It became clear to me by this time that the clickers were inadequate for a complete learning experience. We had started work earlier on connecting wired USB keyboards to a laptop. This was problematic as USB connections and wires in a classroom are problematic additions to infrastructure. There were also problems with the number of such keyboards we could use at any one time.

Figure 3.3: Students using wired parallel feedback devices

With the help of another team member, G Simpson, we commenced the development of wireless keyboards to seamlessly create the environment that we sought. That process was completed by mid 2008 and we were preparing to spread this through the school when differences with the owners of the school led to my departure.

3.4.4 A progression

Over the years we have moved from infra red clickers to full keyboards. There have been several issues that have influenced our decisions:

- infra red was fundamentally "messy" with restrictions on distance and problems with robustness and reliability.
- the move from clickers to full keyboards had several influencing factors such as:
  - While multiple choice questions are better for many applications such as training and data gathering they tend not to be so effective for our educational applications where we need a broader spectrum of data for analysis. In the HAAP project for instance the clickers were ideal but being unable to use full alphanumeric input when designing teaching content is onerous and limiting.

- Even where we are looking for discrete answers, the excursions that some children make gives us insights into how they think. That helps us design better materials. In multiple choice questions excursions are constrained. For instance, when we had the picture of some fruits on a plate (See Figure 5.1), we would never have imagined the responses we got. Those responses gave us valuable insights.
- free form questions have been very valuable for us. Even though they are not the highest frequency of question types, they do have a very important part to play. Most importantly, they give children an opportunity to express themselves. The greatest gain is the opportunity for individual expression which many children never get in our classrooms. On the other hand, a gain that we cannot ignore is the development we saw in the increasing fluency with which children expressed themselves in written English.

3.5 Experiences and Learnings

The learnings during this time came from several experiences in addition to those mentioned in the earlier parts of this section which specifically outline the evolution of our feedback devices.

3.5.1 The experiences

Balajanaagraha

This is a program run by an organization focusing on citizen participation in governance, Janaagraha which was created by two founders who still give it ideals, breadth and depth.

The program works with children in about 40 Bangalore schools and instills through a limited number of sessions run by volunteers, the essence of citizenship by using classroom teaching, projects and exhibitions.

I worked with them for several years as a volunteer, preparing materials, conducting classes and trying to establish measurement mechanisms. This involved delivering lectures on democracy to children living in slums close to the school or at the far end, working with an elite ladies college to create field activities for the school program.

CIE

For several years, I had the pleasure of working as a consultant with The University of Cambridge International Examinations (CIE). The work primarily involved visiting schools that had been accredited by this body to conduct CIE courses and examinations through spot checks when CIE examinations are taking place. Less frequent was the need to inspect schools for accreditation. However, in the course of this work, I was able to visit a large number of schools throughout India and realized the wide spectrum of quality that comes under the heading of “International Schools”.

One was able to speak with teachers in a professional setting and have access to the difficulties that many face as they struggle to come to terms with the reality of school teaching. This is frequently perceived as being boring and uninspiring but is balanced by the reality of a high level of non working days in the year.

Shonu Chandra

Shonu Chandra is a documentary maker who probably knows more about the problems of disadvantaged children than anyone I know. For many years he worked with an NGO who gathered children from rural and sometimes urban areas to create a film that documented some problem in their lives. More than 50 of these films have been produced and a collection of 34 of these films won the “One World” award.

Shonu took me under his wing, and we did several projects together. These included the project on Child Led Social Equity Auditing called the PFL Project , in Tsunami struck Nagapattinam and Earthquake struck Kashmir. We worked on the HAAP Project and projects with children in Vietnam. We reported on the NGO’s progress with Tsunami efforts in Sri Lanka, Malaysia, India, Indonesia and Thailand through a visit to these countries, and we wrote a significant number of child project proposals for different countries. It was a tremendous experience both working with the diversity of children and working with NGOs.

Mind, Brain and Education

In the summer of 2008, a chance communication with Prof. Kurt Fischer of the Harvard Graduate School of Education (HGSE) led to my attending the Mind, Brain and Education Workshop at HGSE in Boston. This was a formidable experience

and many of the learnings from that have found their way into the design of our Topic Plans. The whole concept of Differentiated Instruction, and Kurt's work on the non linear manner in which learning ideally proceeds, were new to people like myself who are enthusiastic amateurs in the K-12 domain.

System Dynamics

I came across this subject and the wonderful community of system thinkers worldwide some years ago. A 2008 meeting with Jay Forrester, Emeritus Professor at MIT who created this science, led me to the realization that this was an essential necessity in the K-12 arena. Our technology represents an excellent medium for this, but we are in the early stages of developing our systems and methods. However, the promise of multiple students developing or exploring a model collaboratively is seductively alluring and we will create the needed software.

3.5.2 Learnings

Broad learnings

Simply put, the learnings were:

- We underestimate children's capacities on the one hand and overestimate the rate at which they can imbibe learning.
- Learning has to be contextual for children
- rural kids have more exposure to analogue processes. They often have less access to formal education and information, but may have greater scope when given access
- urban kids who are poor often lose out on both counts. They have less exposure overall and through poverty are denied access to meaningful education and information sources.
- classroom management is a challenge across the country. Classes are large, teachers domain knowledge is frequently limited by what they themselves learned during their educational process, and the young are often bright and inquisitive. As a result the need to maintain discipline often generates much abuse - mental, physical and verbal.
- teachers who are new are often bright-eyed, keen and nurturing. Their principal protagonists at that stage are often the management/owners in particular, and the system in general. Eventually teachers succumb, as they need to keep their jobs and preserve their careers. There appears to be little they can reach out to.
- decision making is the skill that differentiates the adult from the infant. Yet our schools which should provide a safe environment for learning this vital skill for adulthood abjure the whole idea of children making decisions. Families often do not even consider this necessary, unmindful of the costs of children learning decision making after becoming adults.
- boredom is the primary attribute, sadly, of most classrooms I have visited. Usually the teacher has been most motivated to show his/her skills at the time that a visitor is in attendance - yet children sit glassy eyed.

Classroom management issues

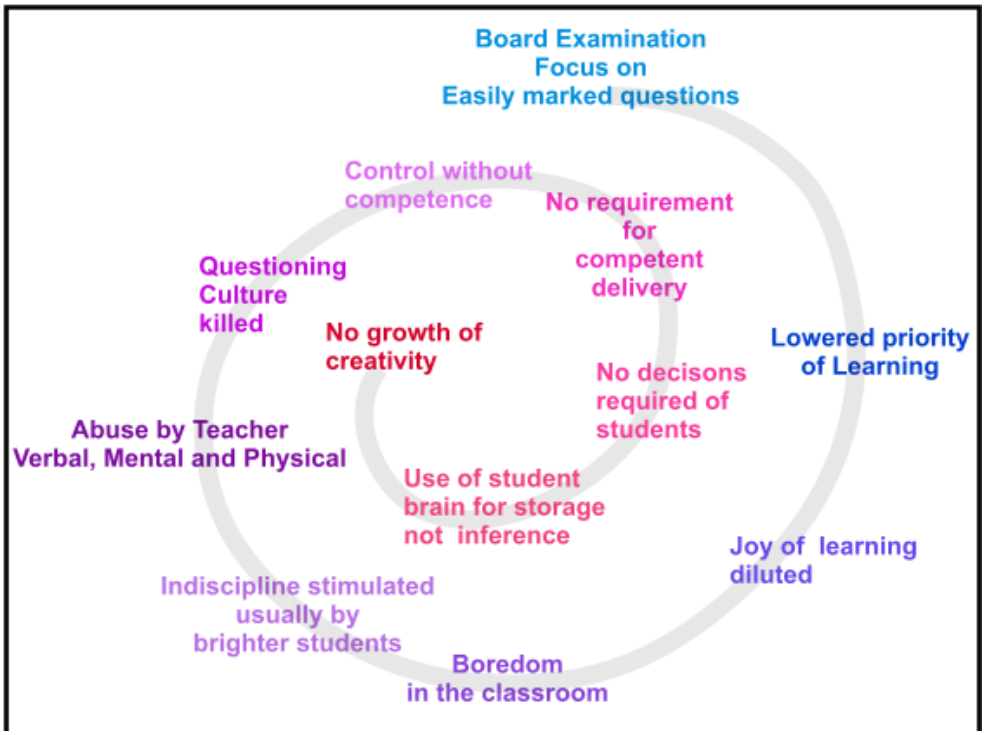


Figure 3.4: Killing creativity through disciplinary processes

Figure 3.4 is an endeavor to capture the process by which the classroom experience degenerates for the student, in the process killing a population's greatest resource - the creativity of the next generation.

The AGEM processes of centrally generating content and measuring performance release teachers from the need for disciplinary measures for classroom management, or so it would appear over these last ten years.

Figure 3.5: Attentivity in an AGEM classroom

While it is not impossible to have indiscipline, there are several factors that militate against it in an AGEM classroom as the picture from the Pilot shows in Fig. 3.5:

- the interactivity created by the large numbers of embedded questions. Students wait for the questions and tend to pay attention in order to answer them.
- the awareness of the performance of the class as a whole rather than of the individual through seeing the graphical feedback
- the quality of the multimedia presentation
- the self esteem acquired by the teacher as the students ascribe the presentation to the teachers contribution
- the students perception that measurement is now taking place continuously, rather than, only in high stakes examinations.

That has the potential for the release of stress, as teachers give up the focus on discipline. Learning conversations can then ensue.

4 Designing the Pilot Program

4.1 The rationale for maths

The choice of the use of Maths as a topic for the Pilot was the result of studying some of the large amount of research available from the United States which points both to the significance of Maths and the difficulties that teachers have in delivering maths concepts in the classroom.

We also noted the literature that points to the relationship between communication ability and mathematical progression and the potential for the growth of inferential ability through mathematical instruction.

We were also influenced by the thoughts communicated by my close friend, colleague on the AGEM team - and much as he would hate to admit it, a social activist at heart - Teyjas Schae. Teyjas put forward the idea of using our feedback systems to create courses for the 90% of Indians who work in the unorganized sector to acquire basic skills as well as certification in Mathematics and English. This meant that we would have to develop these topics, particularly maths from first principles to school-leaving standard but in such a way that this part of the population would find the skills acquired, relevant to the workplace.

It then seemed there was no alternative but to cover the topics end to end, culminating in a Grade 10 competency. The need to relate content to practice is inevitable in such a development.

4.2 The overarching goals of the Maths Program at Maria Niketan

At the first Topic session the students were told of four abilities we would try to inculcate during the Maths program within the scope of the available time. These were cited as:

- To observe and measure carefully
- To calculate correctly and reliably
- To be able to apply previous knowledge in new situations. We call this inference.
- To be able to describe the results of calculations in written and spoken English.

4.3 Objectives

Our objectives in conducting this program were to:

- establish the robustness of our hardware, software and processes in a school environment
- acquire performance data and determine the efficacy of our systems in a specific discipline i.e. maths
- determine on a prima facie basis, how the system impacts other aspects of the classroom e.g. classroom management.
- gather research data to answer other questions such as the differences in the nature of responses from students with different attributes, e.g. gender, age etc.

4.4 Challenges

Our first challenge was finding a school to conduct the Pilot in. This proved a formidable endeavor. Even charitable schools that catered to children from a nearby slum refused to entertain our request. It became increasingly apparent that schools were not open to any competing initiatives - for that is how our submission was seen - in the school.

Through the kind offices of Mrs. Claire Patel and the Archbishop of Bangalore, we were introduced to the Maria Niketan School headed by Father Gregory whose personal involvement in making this work has been a considerable contribution to the success of the program.

5 Conducting the Pilot Program

The program was delivered principally by Manu Patel and Samar Singh - both from the AGEM team with assistance from Jyotsna Singh and Kiran Singh.

What would possibly be labeled Lesson Plans are in this case often labeled Topic sessions, as our plans are made for a coherent learning process for a specific topic. We have found this approach successful as we are not constrained by the limitations of a specific school day. Instead the topic continues until concluded. We found little disadvantage in practice and considerable advantage in design with adopting this principle.

5.1 First Phase - Grade 5 students (9 Nov 2009 - 13 Nov 2009)

As we believed we were preparing Topic Plans for a Grade 5 standard we enrolled students from that Grade at Maria Niketan School initially. We discovered that the children were generally too young for the type of material that we had created due to a rather limited exposure socially. This restriction meant that our Topic Plans would not be contextual within their perspective. We also had other difficulties during this time which may have made matters worse which were related to our own systems and hardware. We subsequently asked the school to provide us a group from the Grade 6 Class which they agreed to do.

5.2 Second Phase - Grade 6 students (23 Nov 2009 - 15 March 2010)

5.2.1 Profile of the student group

The school assigned us the Maths remedial group. In this school, students join different remedial classes based on the weakness of the individual student - from 1400 to 1445. There were 26 students assigned to us, of whom 4 left the class during the early stages of the program. Hence, the analysis data is based on the performance of 22 students.

The profile of the group of 22 is as follows:

Composition: 13 boys and 9 girls

Math ability: The school provided the following on the basis of a Likert scale referenced around the perceived performance profile of students in Grade 6.

- Level 2: 7 students
- Level 3: 9 students
- Level 4: 5 students
- Level 5: 1 student

Average age: 11.8 years on 1 January 2010

5.2.2 The time distribution

Time in hours	Topic
3.75	Pre-test with experimental group
18.75	Topic Plans 1 to 5
11.25	Other activities
2.1	Post test for full group
35.85	Total Program

Table 5.1: Nature of Program and Time spent

5.2.3 The Pre-tests

During the pre-tests 26 children answered 113 questions over a period of 4 sessions of 45 minutes each. These questions were obtained from the Number World's website

The rationale for choosing this was that it would be an independent and competent measure for benchmarking progress while having no sourced relationship to the content that was to be taught. There were slight modifications made to the questions such as changing names of cities and people so participants could have a level of familiarity. No attempt was made to consult the question bank when Topic Plans were being prepared.

General response statistics - experimental group

- The 26 students who undertook this test took took 4 sessions of less than 45 minutes to complete 114 questions. One question (No 163) was rejected as a typo had crept into the correct option.
- Four participants who left the course have been removed from the analysis.
- During the tests, 24% of the instances of answers provided were correct, while in 5% of the instances, no answer was provided. This included instances where no answer was given due to a student being absent.. Question 163 which was rejected constituted 1%, while 70% gave the wrong answer.

5.2.4 The Post-tests

During the post tests, 63 students including the experimental set of 22 answered 65 questions. These 65 questions were a subset of the original set of the Pre-test questions. The rationale for using the subset was that we would have a referential measure of complexity in the two tests. While there does remain the question of the set of 22 students being subject to a training effect, this was sought to be diluted by the following measures:

- Rapidity with which the questions were delivered e.g. less than 2 minutes per question on average.
- The absence of any form of individual recording during the pre-test stage.
- The numbers of questions. There were 113 during the pre test stage and 65 during the post test stage
- The students not being given feedback in either pre or post tests of the correctness of their response
- a time period of 103 days that ensued between the end of the pre-test and the start of the post-test
- No notice to the experimental group that there would be a post test.

General response statistics - experimental group

- 22 students took the test along with 41 students from the control group
- The tests were completed in two sessions of 45 minutes
- Sixty five questions were answered
- The average score of correct answers went up from a mean of 24% to a mean of 75%

General response statistics - control group

- 41 students took the test at the same time as the 22 students from the experimental group
- The tests were completed in two sessions of 45 minutes
- The average score of correct answers was 61%

5.2.5 The Program delivered

An introductory topic plan to show the relevance of maths was followed by 4 topic plans that dealt with:

1. "The concept of number" referred to as TP2 or LP2
2. "Percentages" referred to as TP3 or LP3
3. "Decimals" referred to as TP4 or LP4
4. "Fractions" referred to as TP5 or LP5

5.2.6 Broad goals of the Topic Plans

In all the Topic Plans, barring TP2, the goals were cited in the form of questions. At the end of the Topic Plan students were invited to indicate on a scale of 1 to 5 whether they believed the goals had been achieved for them personally.

Those responses have not been analysed as they tended to be overwhelmingly positive. This could be due to the traditional belief that one does not criticise the teacher, which is hard to overturn in the short term.

TP2: The concept of number

1. to understand why we need symbols to represent quantities
2. to understand that we can develop a system to represent quantities using any number of symbols

TP3: Percentages

1. what are the relationships between integers and percentages when we talk about quantities?
2. what are the characteristics or properties of percentages?
3. how can percentages be added?

TP4: Decimals

1. What are the relationships between integers and decimal numbers?
2. What are some of the ways we use decimals in real life?
3. How can decimal numbers be added and subtracted?
4. How can decimal numbers be changed into fractions?

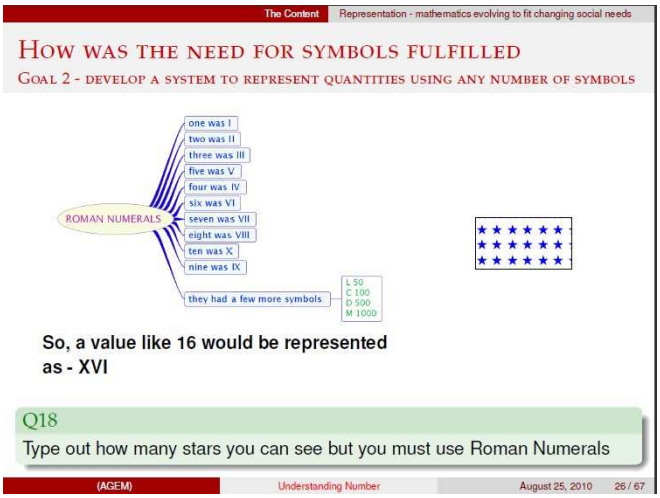
TP5: Fractions

1. Why do we need fractions?
2. How do we represent quantities using fractions?
3. What are the different ways in which fractions are used?
4. How do we add and subtract fractions?

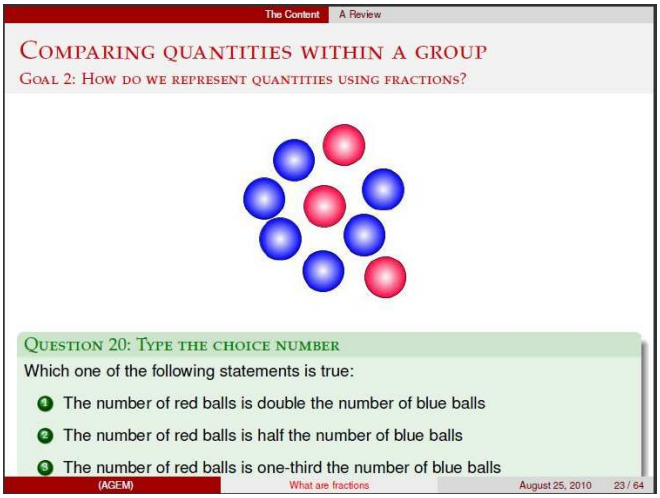
5.2.7 Typical forms of questions

Questions in our lesson plans proceed by small increments of learning to the fulfillment of the specified goals, hence it is difficult to find questions that will be meaningful in an isolated context. However, samples of the different types of questions are provided below:

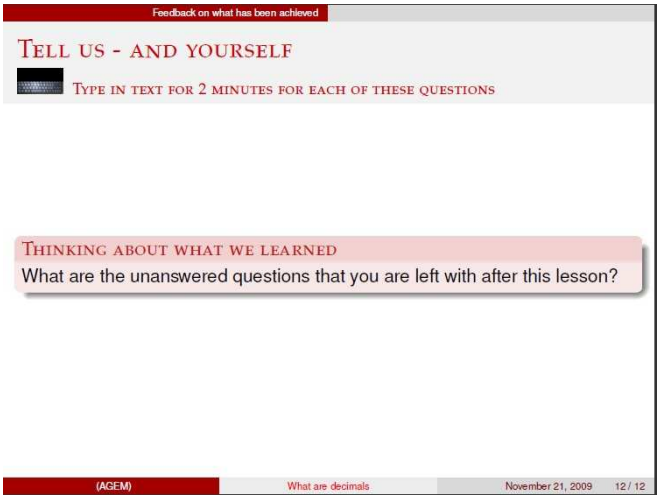
Questions requiring a discrete free form input



A simple multiple choice question



A question requiring an extended free form input



The dates shown at the bottom right end show the date of compilation of the latex files into pdf form. We often store our Topic Plans in latex source rather than pdf. Hence these dates may not conform to the date on which the Topic Plan was run

if compiled again later.

5.2.8 Projects, Activities and Reflections

Projects

Two projects were conducted. The first was designed for students to work out the time period of a pendulum of 3 different lengths so that students could interpret the beginnings of a relationship of time period with length.

The second project was designed to enable students to identify the declining rate of rotation of a CD around its centre. The non-linearity of the rate is what students were expected to infer.

Although a significant number of questions were part of these projects, they were principally designed to start a thinking process. There was considerable progress from the first activity to the last project in terms of the responsiveness of students.

Activities

One activity was conducted, which turned out to be unrealistic in terms of its expectations. This was centered around Polya's problem, and was conducted in two sessions. Students had great difficulty in responding to any questions.

Reflections

This was a useful exercise. It was conducted after Topic Plan 2 had been concluded. The results of that showed some surprising inabilities in calculation, which led us to introduce a maths drill. When compared against the syllabus as exemplified by the textbook for Grade 6, the children were considerably behind the grade. One example of the weaknesses was the response to a set of 4 questions 9 - selected from the maths drill session - that had been framed to test their observational and counting capabilities. These are reproduced in Fig. 5.1.

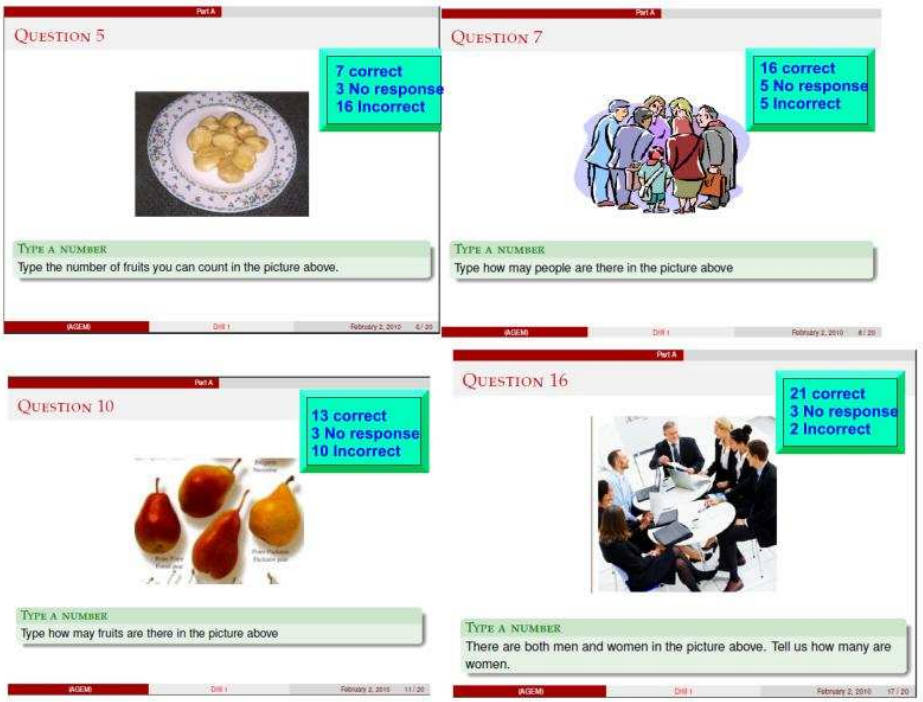


Figure 5.1: Reflection session anomalies

During the Reflection session, the students cited some problems with viewing the screen which was a function of size of screen, quality of eyesight, and distance from the screen. We tried to remediate this by getting a bigger screen and changing the room so the children could sit closer.

However, it also became apparent that children had difficulty in applying simple mathematical processes including counting, to fairly basic real life circumstances, as is shown by Fig. 5.1. We felt this could possibly reflect an inferential disability.

This changed the nature of our design for future Topic Plans to include a lot more imagery on which mathematical processes were required within the embedded questions. We also reduced the increment of complexity between questions. None the less, it appears that this increment could have been further reduced, as Fig. 6.8 shows.

During the Maths drill we learned that the entire group of 22 failed to correctly answer the question, "What is one-eighth of 32". Hence, there was apparently a fundamental weakness in mental mathematics and of course the whole idea of fractions. This may explain why the class sat through a 2.5 hour session on fractions on the last day without complaint. This was necessitated by the shortage of time.

Feedback from teachers The graphic in Fig. 5.1 was also shown at the Presentation for the teachers held some weeks after the reflection session. No real insights emerged from that session which could explain the results. We are left with the view that in a classroom of more than 60 students, it is hard for a teacher to be aware of the precise voids of understanding for individual students, even when that is shared across a significant portion of the class.

5.2.9 Data acquisition

Data was acquired principally in three ways:

1. Computer based record of student's keyboard inputs.
2. Audio and Video record directed at visual expressions of students.
3. The team's personal notes.

5.3 Learnings

5.3.1 Classroom management

Classroom management was only an issue for a few minutes in one session, when we endeavored to take the full group of 63 students for a session prior to the post test and were ill prepared logistically for the exercise. In particular, each student brought in haversacks in which Indian students have historically carried several kilos of textbooks. These took up substantial space in the room which is visible at the back in Fig. 5.2. Another mass, not visible, was in the front of the class. None the less, the students adapted within a remarkably short time, took their keyboards and settled down as the video frame below shows:

Figure 5.2: Sixty three students in the AGEM classroom

The behavior of the experimental group was exemplary without exception. On one occasion, a part of the group was directed by the teacher to other activities. At this time, one of the topic plans was re-run with the remaining group of 12. A single frame from the video recording of the session, shown below, depicts a moment of intense concentration. However, a review of the many hours of video recorded, also illustrates an unusually high level of consistent attentiveness.

Figure 5.3: A section of the experimental group during a re-run of a Topic Session

Our team has been pleasantly surprised by the enthusiasm of the students and their capacity to learn. It remains moot to what extent the nature of the environment created by the feedback devices was a factor in this process.

5.3.2 The issue of textbooks

Our experience in schools, colleges, workplaces and universities in India has shown a clear void in the realm of writing, and the associated communication ability that should accompany it. This is in sharp contrast to the spoken word where even the students in the Pilot group were quite fluent and some cited English as their mother tongue. It would appear to us that there is a clear need for developing the ability for expression.

It is not clear however, if textbooks contribute to developing that ability. To the contrary, children are greatly encumbered by the need to carry heavy schoolbags from home to school and vice versa.

In our Pilot, there was no necessity for textbooks, making it possible to demonstrate that at least for teaching maths effectively, there is little need either for textbooks or homework.

5.3.3 The issue of communicative expression

It may be a fair generalization that in most classrooms in India, it is difficult for all students to participate interactively in the classroom. Perhaps, for this reason, we found the students capacity to articulate thoughts in writing was particularly limited especially in the earlier stages of the Pilot. Initially, we thought this was due to unfamiliarity with the keyboards. However, during one session devoted to this topic, we found that the writing performance was quite outstanding when a sentence was written out on the blackboard but less so when the sentence was spoken out. There was an even more significant decrease in quality when the students had to compose their own writing.

However, there appears to have been a slight improvement in quality as the course progressed. During Topic 2 on the 10 December, 2009, a question answered by the class was:

Please type in a short phrase which describes how we could make the counting more efficient by reducing the number of marks we need to make.

Questions are read out and explained before the capture from the feedback devices or keyboards is activated.

The context of the question is that students were shown a short film of cattle leaving a cattle pen. They were required to mark off each cow as it left the pen, using the tally method, and thereby determine how many cows were missing.

The responses from the first ten students in order of Student Identity (SI) are given below. Also given is the delay before starting to provide that response.

By contrast, the last question answered by students on the last Topic session was:

What are the things you would have done to make the course better?

The responses from the same ten students are given below along with the delays in presenting the answers.

Sl	TP2 Question Response	Delay	TP5 Question Response	Delay
1	coibh erty mht	81	i like themovie	14
2	cost of a cow	59	mats question	35
3	we count count if there is less lins	115	more question	18
4	4d4	117	more questions	19
5	one line half wecan	130		
6	half	77	mor taking and your bber	12
7	we know not do	127		
8	we can crach the linnnnne then we can	119	mor exerises	24
9	it is half	128	by doing sums and studing	19
10	we can draw 10 linechach	179	more play	28

Table 5.2: Improvements in communicative expression

While it is important to note some change in the quality of the response, it is even more telling to see the change in the delay. The delay describes the time period in seconds between the time when the question had been displayed and explained followed by activation of the input process, and the time when a student commenced to respond i.e. the first key-press. In the initial stages of the Pilot we were consumed by the difficulty that students had in commencing the “thinking” process. This appeared to reduce with time.

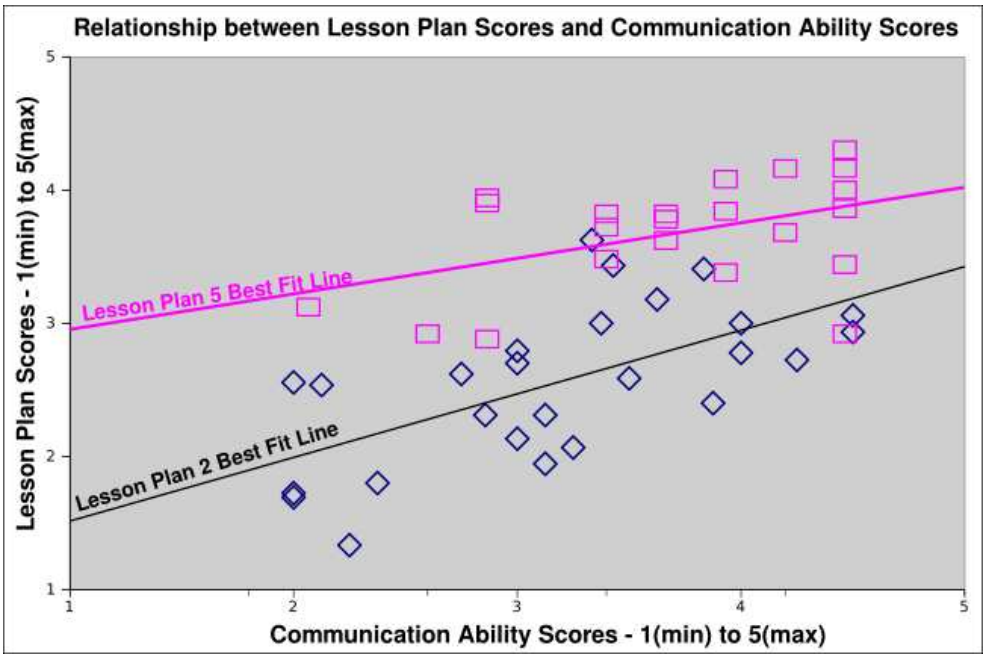


Figure 5.4: Correlation between communicative ability and Math scores in Topic Plans 2 and 5

Fig. 5.4 shows the correlation between scores in maths for Topic Plans 2 and 5 and scores in communicative ability. The scores in communicative ability were obtained by grading the students on a 1 to 5 level on those aspects of Topic Plan 2 where textual answers had been called for.

5.4 Challenges

There were several challenges during this period.

5.4.1 Mains power

The power situation grew progressively worse as the course progressed. In the final stages we installed a UPS which would keep the projector working for about 20 minutes. A projector requiring less power would have sufficed but as the rooms were unusually bright, it would have been difficult to shield the projection screen from ambient light, to the extent needed.

5.4.2 Ambient conditions

Initially, we used the schools laboratory as a classroom. This had very high levels of ambient noise though it was protected from much of the noise on that floor as a result of being at the end of a wing of the school building on the first floor. After January, 2010 we moved to the library room, where levels of ambient noise were lower although the room was a bit too bright for the projector. However, there was less likelihood of it being taken over for other work such as examinations.

5.4.3 Staff response

The early response of existing teaching staff to this pilot, tended to vary from a polite curiosity to very occasional hostility. This changed dramatically after we conducted a seminar on 24 February 2010 where we explained what we had done and requested assistance in understanding some of the responses we had received. The audience listened intently but there were few questions or answers. Most members of the audience were provided with one of our feedback devices and they were taken through a section of one Topic Plan. The cooperation of the staff increased significantly after this seminar, and there is the inevitable reflection that this should have been done at the start of the session. We were not aware in the early stages that our presence had been noticed to the extent it had been.

5.4.4 Competing school activities

To the credit of the school, it must be said that the children have a large number of activities they are engaged in especially around the Christmas season. However, this often resulted in unannounced diversion of the class to another activity or the removal of children from the class during classroom sessions.

5.4.5 Inappropriate seating

Seating arrangements provided were significantly better than what prevailed in the regular classrooms. However, as our system required children to operate keyboards and be able to see the projector, the seating arrangements were not ideal for the work we were doing. There is no evidence that the children felt limited by this, although we have no way of objectively determining the impact of this.

5.4.6 No display on keyboards

Our feedback devices have no display that the student can look at to see what s/he has typed. The instructor is however able to view this in real time. Occasionally, when the instructor notices a typing error students are instructed to backspace and re-type their response. Typically this happens when a key has been held down too long and repeats the character. While this would seem to be a limitation, it was also noted that there was no instance of a student indicating this as a problem. It does however make it difficult for students to see what some one else has typed. A review of the videos captured during the classes does not provide evidence of this form of cheating though students were sometimes seen to be discussing between themselves, which is a practice we did not particularly discourage as neither the frequency of such instances, nor the duration was disturbing. In general, students tend to be increasingly focused on answering the questions rather than seeking the correct answer by other means. This is supported by the consistent decrease in the amount of delay recorded, as shown in Table 5.2.

5.5 Errors

5.5.1 Control group

We failed to appreciate that this group of students was a subset of the larger population of one Class of Grade 6. Our early conversations with the children seemed to indicate that they were a subset of a group of classes and constituted the remedial

group from that population. Hence, the idea of a true control group was a little difficult to visualize. In practice these were the subset of a single class of about 70 students.

6 Analyzing the data

6.1 The nature of the data collection process

6.1.1 Components

There are three parts to the system.

1. Feedback device: Currently this looks like a standard keyboard but with nothing attached to it. The keyboards are adapted to contain a wireless transceiver that passes data to the second part of the system - the coordinator device.
2. Coordinator: This is responsible for managing the communication with up to 128 feedback devices. It interfaces to a laptop via a USB interface. We have never had the opportunity to test the system with more than 76 devices.
3. Laptop: This has the necessary software and is also used to store the data sent from the feedback devices, as well as to commence capture of input data from the feedback devices.

6.1.2 Operation

Historically, the laptop has been used to display the pre-prepared Topic session with embedded questions. This is used in extended desktop mode on an Ubuntu Linux distribution running on the laptop. This configuration permits any window to be dragged between the projected and laptop displays.

Our Topic sessions have largely been prepared using Latex Beamer which allows us to call any type of application from within the PDF file e.g. a movie file.

Typically, the capture software runs on the laptop screen. At this early stage of development it permits start and stop of capture, monitoring of data as it is typed in, and the ability to generate histograms of data as well as a concordancing of text input by the students as a whole. These displays ensure that no student's response is individually identified in the histogram or the concordance. At any stage this grouped graphical response can be moved to the projected display for students to see. All data that is input by students is automatically logged into a Comma Separated Variable (.csv) file. That file becomes the performance archive source.

6.2 The sources of data

6.2.1 Logged student inputs

Most of the data that has been used to do the analysis in this section has come from the inputs provided by students on the keyboard. The data that is captured is:

- the question number
- the student identity
- the input typed in by the student
- the corrected response where the instructor has corrected, say a spelling mistake in a students input. This is particularly important where concordance is being used
- the date and time
- the delay in seconds between starting capture and the student activating the first key-press after the instructor initiates the capture process

6.2.2 Video and audio record

We tried to maintain a continuous video record of all sessions, as it had been our intention to conduct a Distractor Analysis of the sessions to see which parts of the Topic Plan needed improvement. This has not been possible as there is not an adequately distributed range of Distractors both in terms of intensity and in terms of individuals within the group.

The sessions before 11 January 2010 do not have an audio component. These were recorded using a webcam, and audio quality was too poor to be usable. After the start of the new term on 11 Jan 2010, we were able to deploy a Flipvideo recorder which provides High Definition video and the quality of audio recorded has made background noise quite acceptable.

6.2.3 Examinations conducted by the school

Given the nature of the performance increments, we chose to examine the impact that the Pilot may have had on the experimental group in terms of the results of the examinations conducted by the school. We received the results of the exams for the first term and the second and final term. We chose to analyse the results for Maths, English and General Science, as we felt that these are areas that may have benefited.

We were provided the results for each of the 76 students. However, one student did not appear to have taken the examinations.

Our experimental group was composed ot 22 students. An additional 43 students took the post-test, but the total number of students who took both Term examinations was 75. Hence, for the purposes of the analyses based on Term examination results provided by the school, the group of 53 has been taken to be the control group.

Fortunately, we had started our Pilot in November, 2009. At this time, the Term 1 examinations had already been completed. We ended our Pilot on 15 March, 2010 and the Term 2 examinations took place shortly after. Hence, this provided an independent and unrelated measure of performance. The analysis is provided in Section 6.5.

6.3 The Primary observations

6.3.1 Questions

Over the period of the Pilot, students answered more than 400 questions using the feedback devices. These comprised the 4 Topic sessions, activities, projects and reflections sessions but does not include the 113 questions the students answered in the pre-test and the 65 questions answered in the post test.

6.3.2 The Statistical comparison

	Test Score Statistics (%)			
	Mean	**Max**	**Min**	**Std. Dev.**
Experimental Group - Pre Test	24	50	5	9
Experimental Group - Post Test	75	89	46	12
Control Group - Test at end of session	61	82	35	9

Table 6.1: Comparison of group statistics for pre-test and post-test experimental group and the control group

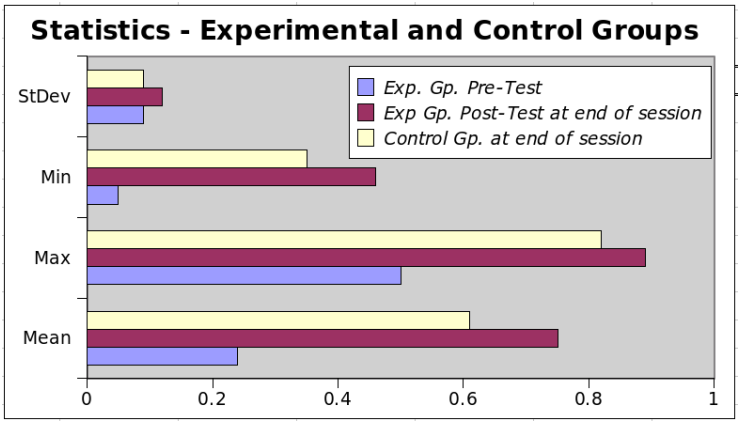


Figure 6.1: Representation of group statistics for pre-test and post-test experimental group and the control group

The notable elements are that the experimental group started off at a very low level, through being a remedial group, and at the post test scored higher than the control group in terms of the mean score, the maximum score as well as the minimum score. Unfortunately we have no data for the start in terms of the control group so there remains the possibility that the control group made similar progress during this period.

The counter argument would be that the experimental group was already assigned to a remedial class program, and it is unlikely that the control group would have performed to the very low levels the experimental group showed at the pre-test time.

Nothing in this should be taken to indicate that the experimental group would necessarily do well in assessments not based on a conceptual understanding of the subject e.g. school examinations. We are endeavoring to get this data but do not have it at this moment.

Potential sources of error

We have used the spreadsheet Average function to determine average score. This can artificially enhance the scores where there are missing values, as these were ignored. Where a student was absent from the class, we have assumed it is legitimate to ignore the missing values. Where a student who was present in class failed to answer a question, it is desirable that the sum of the scores should be divided by the total number of questions to obtain the student's average score.

However, we found that for the experimental group post test, this latter effect amounted to 1.5% of the cases, while for the control group it amounted to 7% of the cases. As any recalculation would tend to further increase the difference between the experimental group and the control group we have chosen not to conduct this re-calculation.

For the experimental group pre-test, we found that there were 26 instances where students were present and failed to enter a value. This amounted to less than 1% of the total opportunities to provide responses.

We have therefore ignored the level of nil responses in all instances.

6.3.3 Normalized gain

On the recommendation of Derek Bruff, Assistant Director, Center for Teaching, Vanderbilt University, normalized gain has been included, as shown in Fig. 6.2.

Normalized gain, usually denoted <g>, equals (post-pre)/(100-pre). So, for instance, if a student got a 40 on a pre-test and a 70 on a post-test, he would have a <g> of (70-40)/(100-40) = 30/60 = 0.5. That means his change in scores from pre-test to post-test was 50% of what it could have possibly been. This measure is a useful way to represent improvement since it factors in the pre-test scores.

As indicated earlier, the pre-test results are only available for the experimental group. The statistical characteristics of the distribution of the experimental group for normalized gain are:

- Mean: 66%
- Standard Deviation: 16%
- Minimum: 27%

- Maximum: 84%

The histogram for the experimental group is depicted in Fig. 6.2 and shows the highest frequency to be in the range of 70% and 80%.

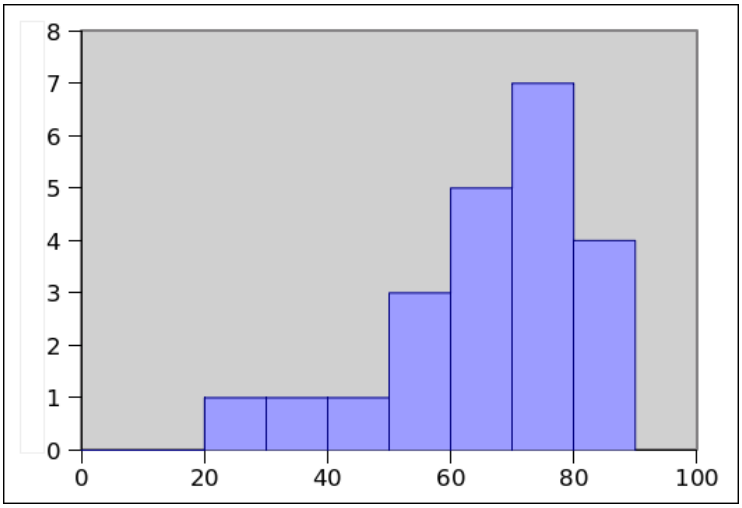


Figure 6.2: Frequencies of Normalized gain for Experimental Group

6.3.4 Measuring progress of the experimental group across Topic sessions

Progression of Topic Plan scores

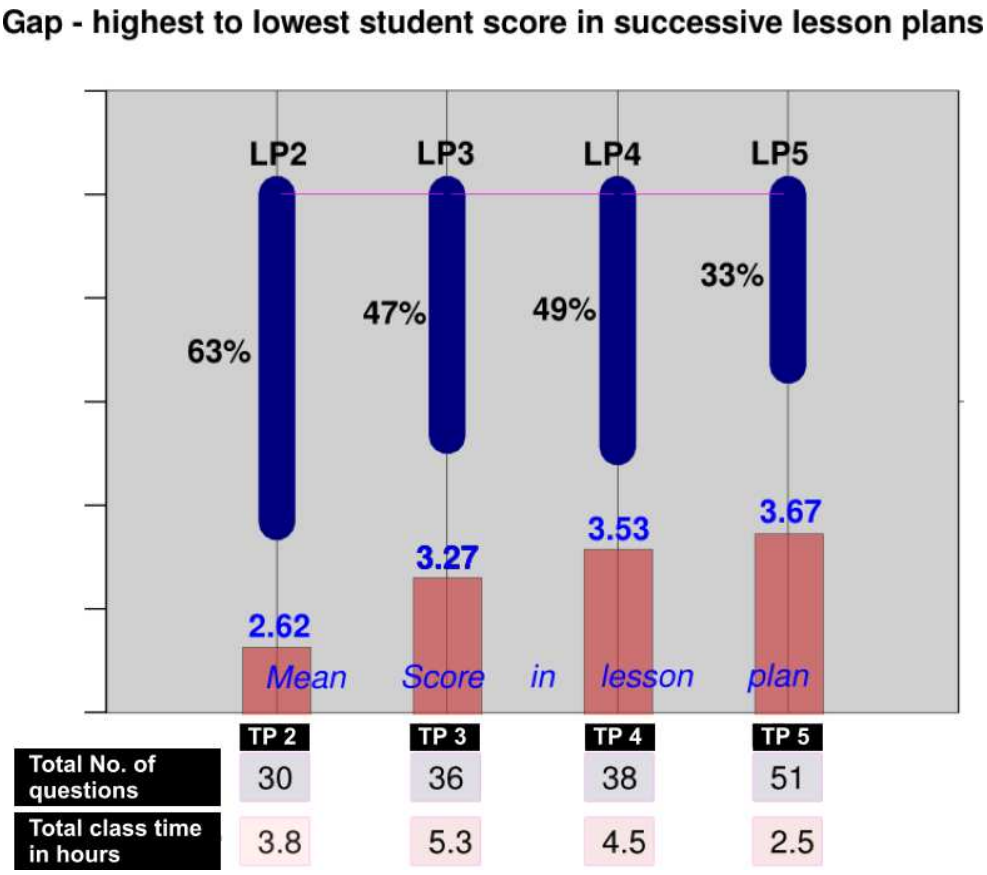


Figure 6.3: Progression of scores and extremes of scores across Topic Plans

In Figure 6.3 two sets of data are visible. The bar graphs in red show the progression of mean scores on successive Topic Plans. These show a gently increasing trend.

The blue bars show the extremes of scores within the experimental group. These were obtained by equating the highest score obtained in a Topic Plan to 100%. The lowest score then represents the lowest score obtained for that Topic Plan. In such a normalized scheme the highest score is always 100%. Hence the gap between the highest and lowest score can be compared between Topic Plans.

We see a slight increase in the mean score which could also be a function of our increasing familiarity with the student group leading to better design of the Topic Plans, as the analysis of the Topic Plan showed us where the steps in learning were too steep (see Figure 6.8).

More welcome however, is the decreasing gap between the highest and lowest scores taken on a normalized basis. Together these give some indication of the significant increment in performance over the duration of the pilot.

6.3.5 The potential impact of inferential ability

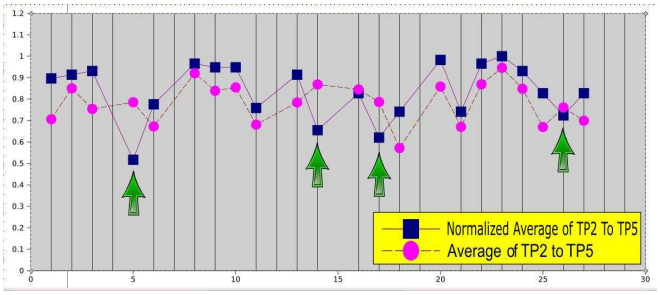


Figure 6.4: Individual student comparison of normalized average of Topic Plan Question Scores against average of Topic Plan 2,3,4 and 5 question scores

The normalized Topic Plan scores, averaged for TP2 to 5, for each student as described in Section 6.3.4 are shown in blue squares for each student. The average scores for Topic Plans 2 to 5 are similarly reduced to a 0 to 1 regime.

If each student had contributed to the increasing compression seen in the upper bars of Figure 6.3, then each instance of the normalized score for each student would have been above the average score obtained over TP 2 to 5.

However, 4 students - indicated by the green arrows - seem to have violated this trend. We chose to investigate if this was related to weaknesses in inferential ability.

Figure 6.5 shows that students 5, 14,17 and 26 represent the lowest results for inference with Student 5 having the lowest which is consistent with the greatest difference in the figure above. Student 18 has a score closely matching that of Student 26 in terms of the Post-test score but there is a relatively high score in the Inference category which is putting the normalized score well above the TP2 to 5 average score.

Against this we need to point out that Student 11 has an inference score which is the same at Student 17 but does not create the same effect in Fig. 6.4. Practically speaking, any correlation with such a small sample does no more than raise the possibility of the need for more research. Student 16, on the other hand, only differers in terms of intensity but has a significantly higher Inference rating.

We would also state our inability to say that these categorizations have adequate rigour as explained in Section 6.4.

Average - Result	QuestionType				
StudNo	Calc	Inference	Obs	Reporting	Total Result
1	0.92	0.64	1.00	0.86	0.80
2	0.96	0.68	1.00	0.71	0.82
3	1.00	0.68	1.00	0.71	0.83
5	0.50	0.43	0.50	0.43	0.46
6	0.67	0.61	1.00	0.86	0.69
8	0.96	0.75	1.00	0.86	0.86
9	0.88	0.79	1.00	0.86	0.85
10	0.83	0.86	1.00	0.71	0.85
11	0.75	0.57	1.00	0.57	0.68
13	0.96	0.68	0.83	0.86	0.82
14	0.75	0.50	0.33	0.57	0.58
16	0.83	0.64	1.00	0.57	0.74
17	0.62	0.57	0.17	0.57	0.55
18	0.67	0.64	0.83	0.57	0.66
20	0.96	0.82	0.83	0.86	0.88
21	0.62	0.64	0.83	0.71	0.66
22	1.00	0.75	0.83	0.86	0.86
23	1.00	0.79	1.00	0.86	0.89
24	0.96	0.68	1.00	0.86	0.83
25	0.75	0.71	1.00	0.57	0.74
26	0.83	0.46	0.83	0.57	0.65
27	0.92	0.61	0.83	0.57	0.74
Total Result	0.83	0.66	0.86	0.71	0.75

Figure 6.5: Post-test breakup per student into C, I, O and R categories and Post-Test Score

6.3.6 Classroom management

The exemplary behavior of the students exposed to this program was in contrast to the prevailing environment in the school, where very high levels of background noise persisted throughout, indicating a significant amount of distraction. Although our frequent observations of other classrooms for the three month period of the pilot did not show unruly behavior in general, high levels of distraction were evident. It was also not possible to identify the relatively intense levels of concentration that prevailed in the Pilot classroom as is evident in Fig. 5.2 even when more than 60 students were present, 41 of whom had no prior experience of the AGEM system.

We feel it is possible that the the use of the feedback devices created the focus that has been evident right through the Pilot.

6.4 Secondary observations

These observations are based on the Post test measurements. We gave each question one of the four categories below depending on how well it could test the following abilities:

1. Calculation (C): To calculate correctly and reliably
2. Inference (I): To be able to apply previous knowledge in new situations.
3. Observation (O): To observe and measure carefully
4. Reporting (R): To be able to describe the results of calculations.

We are not confident that this was a rigorous classification. We asked two members of our team to provide these classifications and where there was a difference a third member arbitrated to provide a final classification. Hence, we feel that this is a reasonable basis for a tentative assessment.

The results below are indicated in terms of the percentage of instances cited against the total number of instances of that particular category across all participants.

6.4.1 Experimental Group - Pre and Post test comparisons

	Scores in %				
	Question Category				
Group	**C**	**I**	**O**	**R**	**Total Result**
Pre-Test (Experimental Group)	24	24	26	24	24
Post-Test (Experimental Group)	83	66	86	71	75

Table 6.2: Comparison of pre-test and post-test scores for the experimental groups across question categories

In general, we note that students abilities were minimal at the pre-test level with little differentiation across categories. The greatest increase took place in the Observation category, with the least increase in the Inference category. This was consistent with our observations during the class and also with the results of the individual Topic Plan analysis. However, there were marked increases in each category between pre-test and post-test scores of the Experimental Group.

6.4.2 Correlation with age of student across question categories

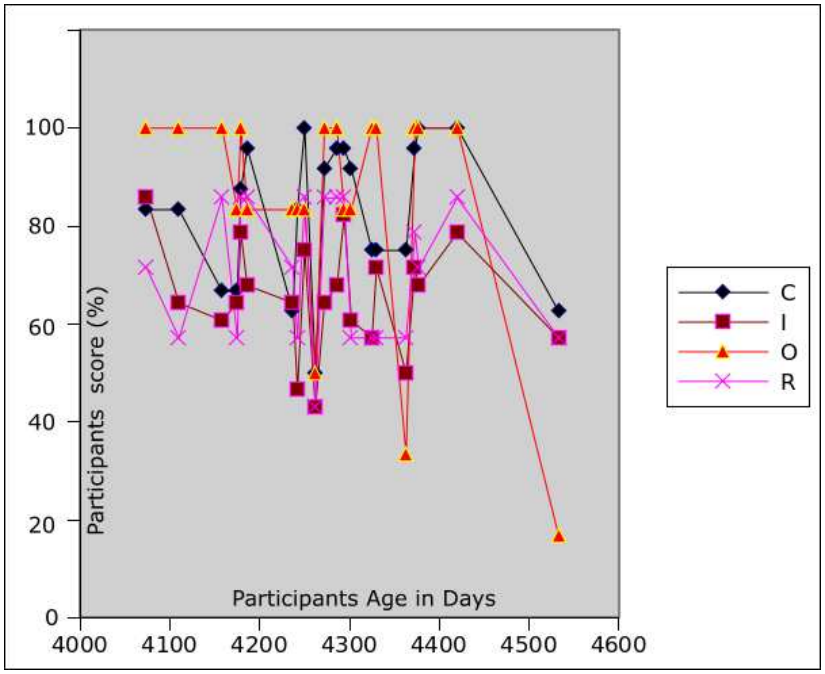


Figure 6.6: Correlation of age of student with post-test scores across question categories

There is little evidence of any correlation between age and post test scores. If there is a tendency it tends to be inverse, although such assessments could be biased by the performance of one individual at the higher end of the age scale.

6.4.3 Comparison across Categories between experimental and control groups - performance on post-test

The table below shows the difference between experimental and control group performance

	Scores in %				
	Question Category				
Group	**C**	**I**	**O**	**R**	**Total Result**
Experimental (E)	83	66	86	71	75
Control (C)	68	51	76	60	61
Difference (E-C)	15	15	10	11	14

Table 6.3: Comparison of control and experimental groups across question categories

It is notable that the common characteristics between the experimental group and the control group are related to the ordinal progression from being weakest in Inference to increasingly improved performance in terms of Reporting, Calculation and Observation.

We were surprised by the capabilities of the students when projects were conducted that required taking readings off the screen. For instance, good results were obtained when they were asked to determine the readings of time differences between successive instances of the pendulum reaching maximum amplitude on the same side. However, calculation ability was a relatively severe problem with children often being unable to divide by ten to obtain the time period from ten observations.

This observation stands in stark contrast to the anomalies displayed in Fig. 5.1. The best explanation that we can provide is that it took some time for students to enter into a "observing-thinking-reporting" process, and the the Maths Drill where these results were obtained preceded the delivery of Topic Plans 3,4 and 5, where the greatest learning may have taken place.

The experimental group seems to have made the greatest strides in the area of Inference and Calculation compared to the control group. Both results are to be expected as the participating children got considerable experience in answering questions many of which also honed inferential skills and some developed calculation skills. The lower increments in Observation could be based on the initial preponderance of Observational ability. The Reporting results are better than expected as there was no overt attempt at specifically building reporting ability.

6.4.4 Comparisons across Categories between the maths teacher's estimates of ability and scores - experimental group

At the end of the course but before these results were computed, the maths teacher was invited to provide an estimate of the abilities of participating students of the experimental group in maths on a Likert Scale of 1 to 5 where 1 represented the lowest level of ability and 5 represented the highest level of ability in the full class. The table below provides the scores against the order of ability cited.

		Q Type				
Teacher's Estimate of ability	No of students	**C**	**I**	**O**	**R**	**Total Result**
Likert Level 2	7	0.74	0.63	0.9	0.65	0.7
Likert Level 3	9	0.83	0.67	0.78	0.71	0.75
Likert Level 4	5	0.95	0.69	0.93	0.8	0.82
Likert Level 5	1	0.92	0.61	0.83	0.57	0.74
Total Result	22	0.83	0.66	0.86	0.71	0.75

Table 6.4: Teacher estimate of ability compared to post test scores of experimental group

From this it would seem that while the estimates were generally correct the one student cited as being a Level 5 was probably closer to Level 3.

It is interesting that this student identified as Level 5 scored well in the Calculation category but was weakest of all levels in the Inference score. In India, it is not uncommon for calculation ability to be used as a proxy for competence in maths.

6.4.5 Comparisons across Categories between male and female student scores - experimental group

	Question Category				
Gender	**C**	**I**	**O**	**R**	**Total Result**
Female	0.84	0.65	0.83	0.67	0.74
Male	0.83	0.66	0.87	0.74	0.75
Total Result	0.83	0.66	0.86	0.71	0.75

Table 6.5: Gender comparison of question categories using post test results

There appears to be little difference between Female and Male student performance in terms of the post-test scores.

6.4.6 Comparison across categories between socio-economic levels - experimental group

We requested and received from the school a Likert scale indication of the socio-economic levels of the experimental group within the framework of the class comprising the experimental and the control group. The groupings are as follows:

	Question Categories	**C**	**I**	**O**	**R**	**Total**
	No of questions	24	28	6	7	65
		Scores in %				
No of stud	**Female students**	**C**	**I**	**O**	**R**	**Mean Value**
4	SocioEconStatus 2	95	71	96	79	83
2	SocioEconStatus 3	83	66	92	57	74
3	SocioEconStatus 4	71	56	61	57	62
0	SocioEconStatus 5	No female student				
No of Stud	**Male Students**	**C**	**I**	**O**	**R**	**Mean Value**
2	SocioEconStatus 2	88	59	67	64	71
4	SocioEconStatus 3	76	59	88	71	69
4	SocioEconStatus 4	80	69	96	75	76
3	SocioEconStatus 5	92	79	89	81	85
Total		**C**	**I**	**O**	**R**	
22	Average Score	83	66	84	69	74

Table 6.6: Comparison of gender and Socio-economic status using post-test scores of experimental group

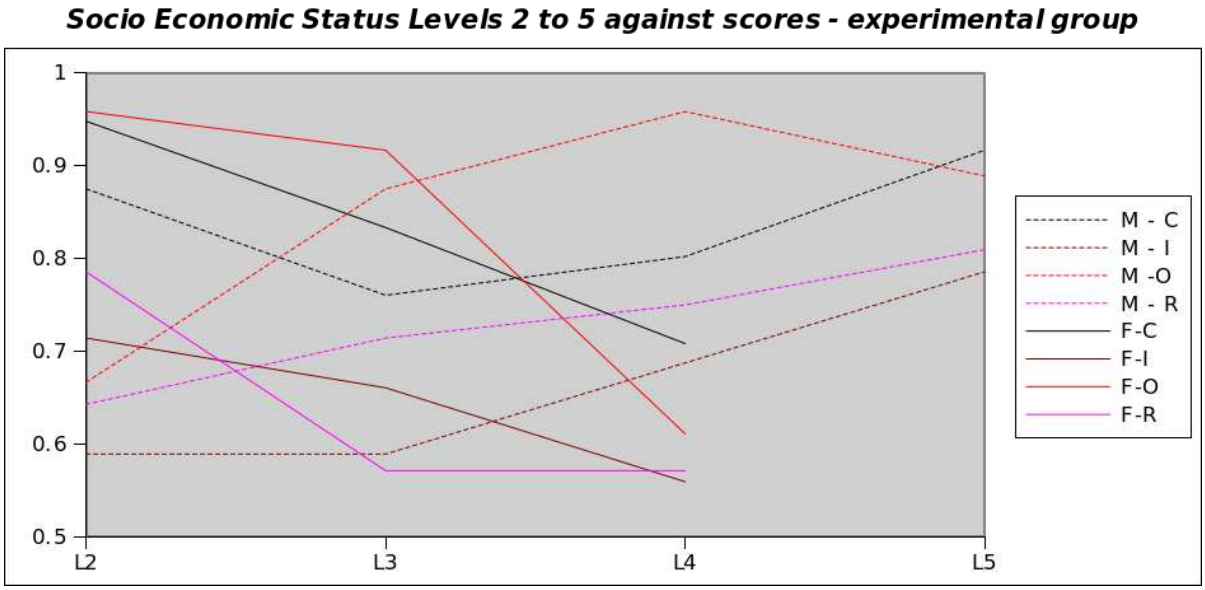


Figure 6.7: Representation of trends of socio economic levels for male and female students using post-test results for experimental group

In the graph above, the scores for female participants are in solid lines and the scores for male participants are in dotted lines. The different colors represent the four categories that question types were broken into viz. Calculation, Inference, Observation and Reporting.

The general trend for the scores of female participants seem to indicate a decrease in performance with increase in socio-economic status while the scores of male participants seem to generally indicate an inverse relationship.

At the prevailing socio economic levels in the school it is possible that the less well off see education as a necessity as compared to those at higher socio-economic levels, particularly among female children. Among the female population a higher socio-economic level may provide a greater sense of security and therefore a lower level of motivation. With male participants this is likely to be less of an issue as security is not a prime motivator. However higher levels of resources resulting from a higher socio-economic level may support performance e.g. access to books, computers or the Internet.

6.4.7 Detecting errors in Topic Plan design

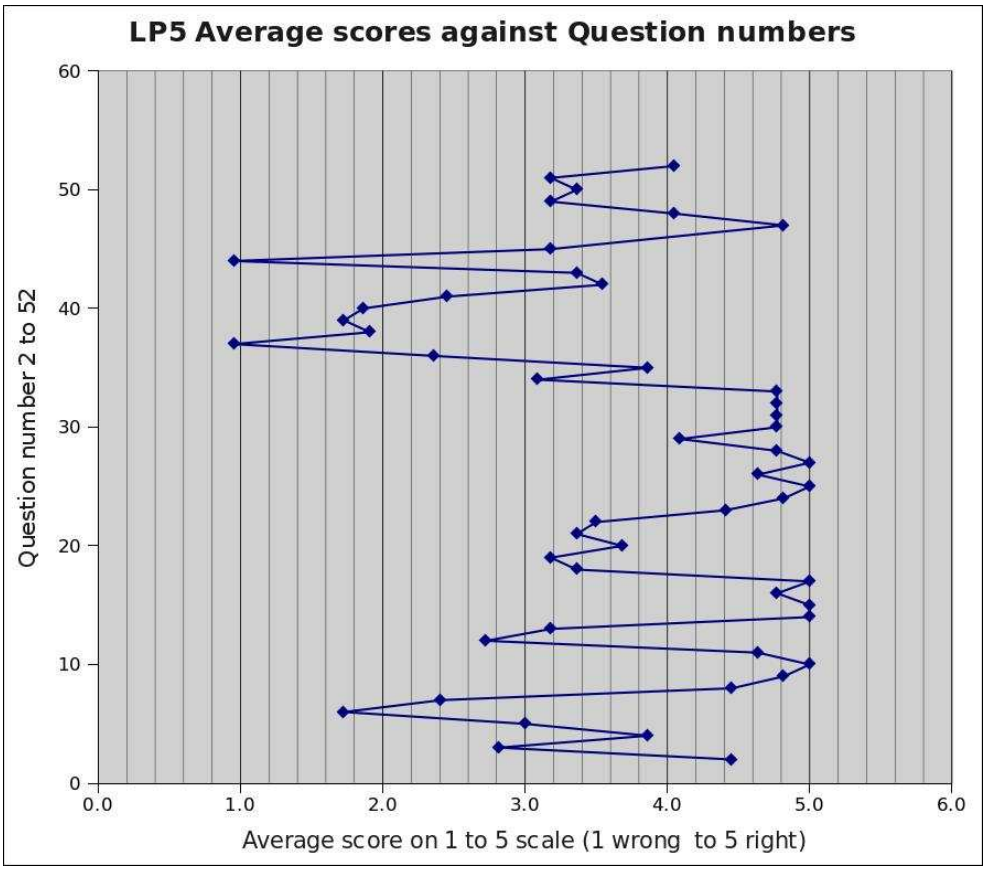


Figure 6.8: Topic Plan 5 Average score per student for Questions 2 to 52

In Fig. 6.8 we can see that the average score of the class in Topic Plan 5 for Question Numbers 35 to 40 was less than 2.5. This is a very low average and indicates a likely fault in the design of the Topic Plan, possibly through the content being taken at too fast a pace. A similar problem is evident in Questions 5 and 6. Such indicators give scope for careful consideration of the related content of the Topic Plan. That carries with it the scope for making better Topic Plans in the future and for revising existing Topic Plans.

6.4.8 The level of exposure of the experimental group

One of the surprising insights was the very low level of exposure the experimental group had to measurement of any type. When asked to type in their own height on one occasion, we realized only 1 student of 22 knew his height in terms of length units.

On another occasion, no student was prepared to hazard a guess about the length of a car parked and visible from the classroom window. When pressed, figures as high as 200 ft. were cited.

6.5 Maria Niketan's end of term examinations

The school conducted one set of examinations internally at the end of Term 1 around October 2009. These concluded before we commenced our engagement with the school. Term 2 examinations were conducted shortly after we completed our program in March/April 2010. We were of the opinion that our program may impact the results of the Term 2 examinations in Maths, English and Science. We asked for and received individual student results for both Term 1 and Term 2 exams. We

have no idea of what was taught or examined or the nature of the examination process. We do know that the three subjects being considered had 2 components each - Written and Oral.

It appears there are 75 students who took the Term 1 and 2 examinations out of a class strength of 76. Only 65 students attended our post test examinations which form the basis for our pre and post test assessment described earlier. It is assumed for this exercise that the control group comprised the 53 students who were not part of the experimental group.

Change in Scores between Term 2 (After AGEM Program) and Term 1 (Before AGEM Program)

Characteristic	Group (No. of Students)	Change (Term 2-Term1) scores		
		English	Math	Science
Average Change	Experimental Group (22)	9.9	7.5	13.8
	Control Group (53)	9.4	2.7	10.3
No change instances	Experimental Group (22)	2	1	1
	Control Group (53)	6	10	4
Negative change	Experimental Group (22)	3	4	3
	Control Group (53)	13	16	12

Figure 6.9: Difference of Term 2 and Term 1 School examination scores - Experimental and Control groups

Figure 6.9 is based on the differences for each student between the Term 2 results and the Term 1 results in the three subjects, English, Math and General Science.

The first two rows show the average change which averages the change in respect of each student on the difference of scores. Result are given in the upper row for the Experimental Group of 22 students and the lower row shows the results for the control group of 53 students.

The next two rows indicate the number of instances where no change took place for the experimental group and below that for the control group.

The final two rows indicate the number of instances where the change was negative. This means the Term 2 exam scores were lower than the Term 1 exam scores.

6.5.1 Conclusions of the comparison of Term exam scores

- The average increment shown by the experimental group was more than 2 times that shown by the control group i.e. 7.5% as opposed to 2.7%
- The number of instances where there was no change was limited to 1 for the experimental group while for the control group this was 10 times higher althogh the control group size was less than 3 times larger.
- There were students who got a lower percentage of marks in Term 2 compared to Term 1. There were 4 instances of that in the experimental group and 16 instances in the control group.
- The experimental group recorded a somewhat better performance in the English and Science examinations also, although this was not as large as it was in Maths.

We would wish to state that any improvements in the schools end of term examinations were unlikely to be related to commonality of content between their syllabus and the content we taught. The textbook for Grade 6 in the latter part which is likely to be the subject of the Term 2 examinations deals more with issues such as factorisation, Algebra, and Geometry.

6.6 Limitations of this study

Sample size

A larger sample would have served the Pilot better. However, we were provided a segment of the class selected for remedial teaching.

Control Group Pre-test

The study would have benefited from a pre-test provided to the control group. This would have been possible if our interactions with the students had not led us to believe that they were from several sections of Grade 6 rather than one class of that Grade. This problem was further compounded by the fact that due to software issues we were not able at that time to use more than 28 keyboards simultaneously.

Pilot of Topic sessions

We would have preferred to pilot our Topic sessions with a separate group but our attempts to get students from other sources were unsuccessful.

Absence of prior studies with multiple keyboards

While numeric clickers have been in use for some considerable time, the capacity to enter alpha numeric text has been more limited. We were unable to find previous studies of this sort of work. We believe further studies may be needed to see if similar results could be obtained by using only multiple choice methods.

However, it appears that the capacity for written expression is a necessary element for most students in India. We are not sure therefore that clickers with limited alphanumeric capabilities would be suitable for that purpose.

7 Conclusions

7.1 Primary issues

The primary outcome of this Pilot was a great sense of personal satisfaction for the AGEM team. The team that delivered the sessions found this to be an enriching experience once TP2 had been conducted and our software problems had been resolved. More specifically:

- While it would have been better to have pre-test results for the control group also, the significant improvement in the experimental group was unexpectedly large.
- The feedback from the children indicated they would prefer more questions. We are also of the view that an even slower progression to allow the weakest to learn better would be desirable. We never had a situation where some students appeared to be impatient at too slow a pace of progress.
- We address again as we first did in Section 3.5.2 the issue of contextual learning. We are left with the realization that context is important for young children particularly the urban poor whose exposure to real life events is more limited than it is for the rural poor in many ways.
- The issue of classroom management has been a non issue although the team that delivered the program was not specifically trained in managing young children. The keyboards kept the children engaged in a relatively permanent state of anticipation for the next question, but the quality of materials on the projected display, though not particularly well done, were an improvement on what prevailed. It is possible both factors conspired to make classroom management an easy task.
- The development of decision making skills is another issue we had targeted at the start. Many of the questions embedded in our Topic Plans did not have a clear cut right or wrong answer. However, all called on the child to make a decision. The very low frequency of cases where children did not respond to a question was possibly due to our focus on keeping individual decisions confidential. It was possibly also a factor in developing confidence, and hopefully, decision making abilities.
- We touched upon the issue of boredom in classrooms in Section 3.5.2. It is hard to go through the videos of the sessions and see signs of distraction that would point to boredom. On the contrary we were impressed with the high levels of attention.
- There is scope for the Topic Plans to be greatly improved as a result of the analysis. We have been pleasantly surprised however when we delivered Topic Plan 2 to a combined set of Grade 9 and 10 students at an elite school in Chennai that the written feedback provided was positive - sometimes unusually so - without exception.
- At the end of each Topic Plan we got responses which were invariably positive to the achievement of goals that we had stated at the start of the Topic Plan. We are not convinced however that this approach of declaring goals in advance has value, as it seems doubtful whether the children understood the goals at the beginning. We are also not certain that declaring the goals did not help them to map out an internal coherence of understanding as we progressed through the Topic Plan. We hope to explore this aspect further.
- We are led to the view that building communication skills in the children - an area where our keyboards have considerable potential - should be a parallel activity in any future project.
- We are also of the view that there needs to be a slower progression in terms of building conceptual understanding and in terms of creating inferential skills.
- We are left with the insight that just as it is difficult for most teachers to discern voids in understanding in the classroom (See Section 5.2.8), it was also difficult for us to envisage the extent of the gains in the experimental group. Hence, we do feel that the AGEM mechanisms are useful accompaniments to the learning process.

7.2 Addressing the original objectives

The objectives of the Pilot we cited in Section 4.3 were to:

1. establish the robustness of our hardware, software and processes in a school environment: Although we did identify some embedded software problems, we were successful in overcoming these at a very early stage and proving successful operation for several months.

2. acquire performance data and determine the efficacy of our systems in a specific discipline: we believe this has been accomplished.

3. determine on a prima facie basis, how the system impacts other aspects of the classroom: We have seen a positive impact on classroom management, as well as on attentive behavior in the classroom, and finally on the increment of learning performance.

4. gather research data to answer other questions such as the differences in the nature of responses from students with different attributes, e.g. gender, age etc.: While we have obtained and analyzed this data it is not clear if very valuable insights have emerged.