2010

FINAL REPORT FOR THE CONSULTANCY TO DETERMINE THE STATUS OF THE CONTINUOUS ASSESSMENT PROGRAMME (CAP) IN THE SIXTY (60) FULL TREATMENT SCHOOLS UNDER THE SES PROJECT

JEROME DE LISLE
CONSULTANT
8/31/2010

Contents

Executive Summary

1) This final report describes a theory-driven evaluation of the Trinidad and Tobago Continuous Assessment Programme (CAP) implemented in the primary school system. Evidence was collected using a multiphase mixed methods research design, with information gathered on the fidelity and strength (intensity) of the programme in a sample of 60 schools in the seamless project and 40 schools in the original CAP pilot project.

2) Phase I was an exploratory qualitative study of eleven sites, Phase II was a quantitative modelling study based on a multi-instrument survey of 378 teachers in 35 schools, and Phase III was an explanatory qualitative study of two schools with different levels of implementation and two principals at relatively high implementation sites.

3) The evidence collected suggests that programme strength was variable across sites. In schools reporting high programme strength, most of the original CAP activities were still being done. These schools usually had strong leadership by principals who were originally trained under the CAP Pilot programme. However, in schools with low programme strength, some teachers were not even aware of the CAP.

4) Overall, the concerns based profiles suggest that the system was dominated by *nonusers* of CAP, although there might be lower resistance in schools that reported high implementation.

5) The most frequently implemented activity was the CAP project, but this activity was not conducted in a way that provided the full benefits of formative assessment. Neither was there efficient use of the data collected even at high implementation sites. Thus, programme fidelity was judged to be universally low, with inadequate formative assessment, a lack of feedback to students, and poor or inappropriate data use. The quantitative modelling study confirmed that fidelity outcomes, such as providing feedback and using multimodal assessment, were very different to "doing" CAP and, as such, the independent variables in the model were much less predictive of fidelity measures.

6) The integrated meta-inferences confirmed that several contextual, organizational, and teacher variables were important, with user variables more important in fidelity.

7) Most students in schools had very positive views of assessment, but tended to see classroom assessment as a mechanical recording of grades or used as an accounting procedure, rather than as a tool for improving learning in the classroom.

8) The key recommendations focused on the need to develop a coherent and balanced comprehensive assessment policy and a simplified, targeted and restructured CAP, with an emphasis on formative assessment.

9) Caution should be exercised when developing new policy that seeks to integrate classroom assessment with the placement function currently held by the SEA. The wide variation in practice, universal absence of training, and general lack of support at building sites suggested that teacher measurements and judgements would lack validity and integrity within a high stakes context.

Glossary

ANOVA-stands for analysis of variance, a statically technique used to evaluate whether there are differences between the mean across different subgroups.

Continuous assessment - Ongoing holistic assessment in the classroom designed to produce data that leads to improvement in teaching and learning.

Evaluation - A careful, retrospective, and systematic assessment of the merit, worth, and value of the administration, output and/or outcome of some government intervention policy, or programmes. The evidence may be used to in future action.

Evaluand- that which is being evaluated (e.g., program, personnel, product, policy, proposal, procedure)

Innovation/Intervention - A new policy, service, or programme that is introduced into an education system in an attempt to solve a problem. The innovation requires incorporation into existing practice.

Implementation - The process of assessment change beginning with initiation and ending in routinization.

Interim Assessment - A standardized assessment administered by a school, cluster of schools or an education district designed to provide feedback into the existing standard of learning.

Meta-inference - A final mixed method conclusion based on both quantitative and qualitative findings.

Mixed Methods Research Design - A research methodology that combines quantitative and qualitative approaches in an integrative fashion. There are several design types and variants with different emphases in terms of weighting and sequencing.

Multiphase Mixed Method Design - A particular type of mixed method design often used in evaluation studies in which the researcher implements component projects within multiple

phases. These phases may be sequential and/or concurrent g over the course of the single programme.

Formative Assessment - Assessment that is embedded with teaching and learning, involving providing various sources of feedback to students and designed primarily to provide high quality information for, improving teaching and learning, ultimately leading to improvement in student learning outcomes.

Path analytic model –a diagram showing hypothesized causal relationships between a set of variables with arrows based on the significant regression coefficients.

Programme- An organized, planned, and ongoing effort designed to ameliorate an educational or social problem or improve education or social conditions.

Regression coefficient (Beta)-Beta is the standardized regression coefficient sued to measure the impact of an independent variable.

Summative Assessment - Assessment designed to provide a summary measure of the state of learning among students.

Scoring Rubric-A qualitative-evaluative scoring device which provides criterion referenced information on the nature and levels of performance required on a given assessment task.

Implementation Fidelity - The trueness of implementation of an innovation based on the expectations and description of the programme designer or program theory.

Implementation Strength/ Intensity - The extent to which different elements of the innovation are present at a high level at different implementation sites.

CBAM - Concerns Based Adoption Model - A comprehensive theory of education change that focuses upon the needs of the user developed in Canada and used worldwide.

List of Figures, Tables, & Boxes

Introduction

Why evaluate local CAP?

In the last two decades, increasing attention has been placed on the role of student assessment in the improvement of education systems across the world. Much of the initial emphasis has been on the use of assessments external to the school, commonly in the form of accountability testing. Kellaghan and Greaney (2001) noted that these assessments for monitoring standards (called national tests in Trinidad and Tobago) have had a critical role in improving the quality of education in several nations. Much of the recent growth in educational assessment, then, has been in the implementation of these monitoring and accountability systems. This has occurred in both developed and developing countries, as they respond to the expanded vision for meeting learning needs and the requirement to improve and assess learning achievement first outlined in the World Conference on Education for All (EFA) in Jomtien in 1990. Some have called these measures, *national learning assessments* (UNESCO, 2008), but even when these external assessments are instructionally supportive and used formatively by schools, they provide limited information on student learning.

The globalization of assessment and assessment practice has also meant some uncritical transfer of assessment practice and policy (Sebatane, 2000). Policy transfer in education refers to the adoption of an education practice in another country. However, increasingly in countries of the South, uncritical transfer from Western countries has been replaced by "gelling", in which indigenous knowledge is incorporated into the imported education policy (Johnson, 2006). Nevertheless, developed countries also use transfer to analyze international best practice when constructing new education policy. However, there are limitations in transferring policy in assessment because different countries vary in the weight placed on the three primary purposes of assessment, *learning, certification, and accountability*. As Black and Wiliam (2007) have argued, tensions arise from the conflict between these different purposes, with various countries employing different strategies to resolve tensions. In some countries, such as Germany, the summative function dominates, with students required to face high stakes assessments frequently. In New Zealand, both accountability and certification measures include teacher assessments in the secondary school and thus the role of the teacher and classroom assessment are enhanced. In the US and the UK, assessments external to the school predominate with a focus on constructed response in the UK and multiple-choice formats in the US. In the UK, new policy has introduced data from teacher assessments into accountability measurement at the primary school level.

The external summative function in the form of public examinations has traditionally dominated Caribbean assessment at both the primary and secondary level, but in recent years, external assessments have changed, making even greater use of the traditional UK constructed response rather than the multiple-choice format, popular in the US. At the secondary school level, teacher assessments have been incorporated in the certification function through the school-based assessment used extensively in CAPE and CSEC. At this

stage in its developmental cycle, uncritical policy transfer might not be the best strategy for education reform in Trinidad and Tobago. At the same time, incorporating indigenous knowledge must not ignore international best practice or include misunderstandings and distortions of effective policy in assessment. It seems important, then, for Trinidad and Tobago to gather benchmarking and other information in pursuit of evidence-influenced policy. For example, the decision to incorporate teachers' assessments into the high stakes selection for secondary school is a critical decision that should be based on evidence of the quality of teachers' assessments and overall assessment literacy of teachers and principals. Useful information on current assessment practice can be provided through both policy research and rigorous evaluation studies on assessment use and practice assessment. Such indigenous knowledge can point in the direction of further and appropriate refinement in policy.

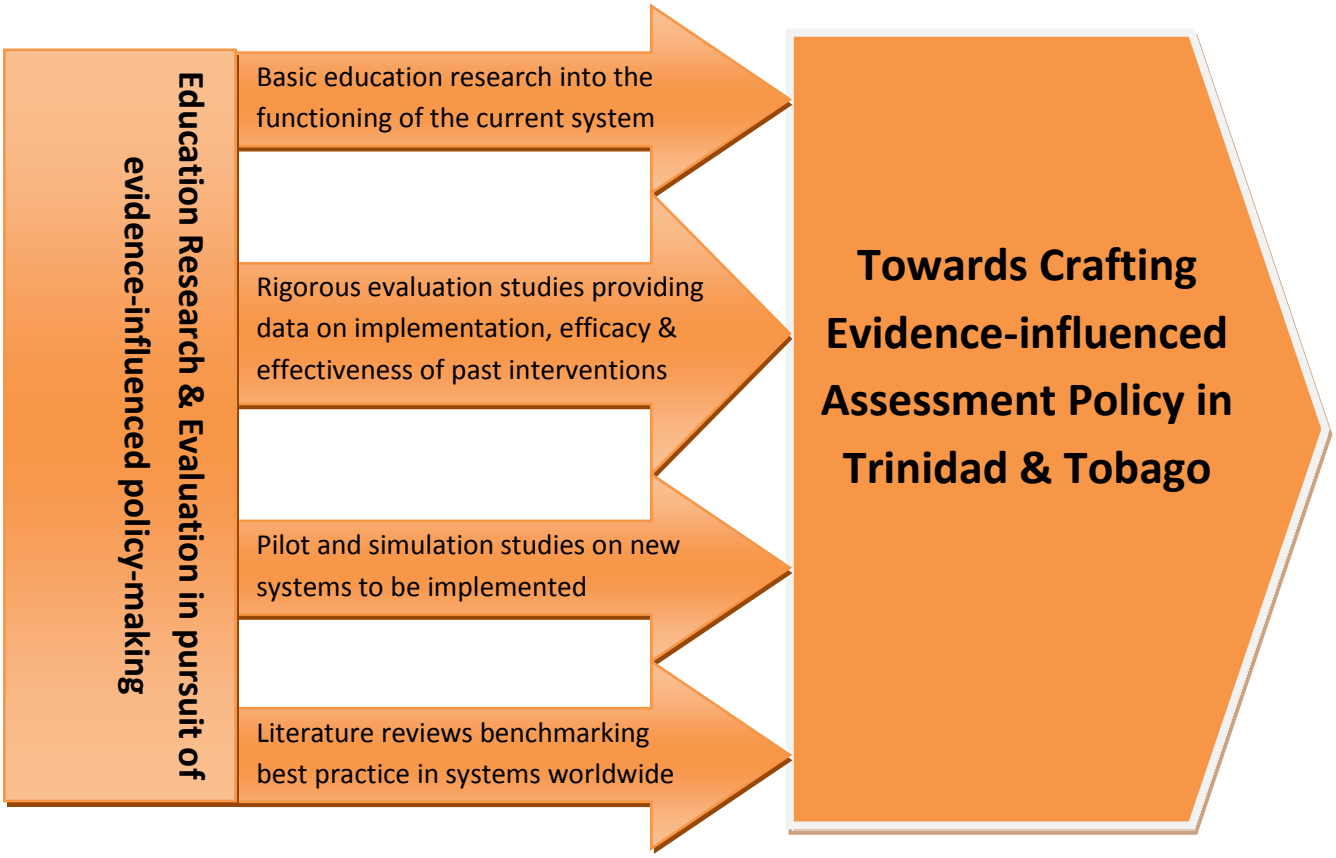


Figure 1: Sources of information for evidence-influenced policy in Trinidad & Tobago

Figure 1 shows the various ways education research and evaluation might influence the creation of evidence-influenced assessment policy in Trinidad and Tobago. As shown, benchmarking best practice might be coupled with basic research and evaluation of current assessment schemes. There is surprisingly little basic research into assessment issues in Trinidad and Tobago. Certainly, there is little research into systems such as the CAP, NCSE, and even CSEC and CAPE school-based assessments, although the latter has been in existence for some time.

Internationally, the increasing role for large-scale or external assessments within accountability systems and the resurgence of public examinations in Commonwealth and European systems have been complemented by a renewed focus the role of classroom

assessment in promoting student learning (Black & Wiliam, 1998a, 1998b; Vlaardingerbroek & Taylor, 2009; Wolff, 2004). Classroom assessment has received much greater scrutiny in recent times, both in the developed and developing world. In the US, for example, despite the increasingly important role of large–scale assessment used for accountability in recent education policy, there are now strong calls for a balanced assessment policy and increasing emphasis on formative assessment practice (Redfield, Roeber, & Stiggins, 2008; Stiggins, 2008).

Africa has had a long history of using continuous classroom assessment, dating back to the 1970s, with reform sometimes associated with political and social movements. Thus, in many cases, there have been attempts to use continuous assessment to replace or supplement high stakes certification examinations, where the failure rates have been quite high (Ajidagba, 2004). Even in Africa, well-designed Government-sponsored and independent evaluations of such practices are increasingly shedding light on the tensions between formative and summative purposes in continuous assessment practice in Africa. Since apartheid, for example, South Africa has experienced two assessment reforms, continuous assessment in the early 1990s and assessment for outcomes based education in 1997. Proposed reform beginning 2011 also gives a significant role to continuous assessment (Motshekga, 2010). Vandeyar and Killen (2003) examined the outcomes based reform which encouraged radical change in classroom assessment practice but found very little change in the traditional focus on producing marks for reporting.

The lesson for Trinidad and Tobago is that studying and improving classroom assessment must be of increasing concern for policymakers as they enact new general policy in curriculum and teaching-learning. It may be that in traditional teaching systems, much of what comes in as innovation in teaching, learning and assessment is not easily adopted. It is perhaps significant that two decades after Jomtien, then, some of the international emphasis has shifted to the key role of classroom assessment as a notable factor in improving learning outcomes. Packaged as assessment as/for learning or simply as formative assessment, classroom assessment has the ability work in the interests of improving schools and learning. Indeed, meta-analytic evidence provided by influential organizations such as the OECD confirms that formative assessment might be one of the more powerful interventions in educational reform (OECD, 2005).

Placing greater value on classroom assessment

However, it is perhaps wrong to believe that much of current classroom assessment is formative. Indeed, classroom assessment practice in the Caribbean classroom might contribute little to promoting learning among students. A formative-summative dichotomy governs the way teachers make use of assessment in the classroom. Thus, Black and Wiliam (2007) have reminded us that:

> The terms “classroom assessment” and “formative assessment” are often used synonymously, but the fact that an assessment happens in the classroom, as opposed to elsewhere, says very little, either about the nature of the assessment, or about the functions that it can serve. Classroom assessments may provide a sound basis for summative

> assessments, and those conducted outside the classroom may provide valuable insights into how to take learning forward (pp. 22)

Along that dichotomy, term tests and quizzes would have limited formative use because they cannot provide detailed information to help the learner improve. They cannot inform on and highlight criteria and expectations. Moreover, in the Caribbean classroom, tests are traditionally of poor quality with limited use in providing feedback to students.

Black and Wiliam (2007) have argued that school accountability can come only after student learning in the classroom. Therefore, as happens now in Trinidad and Tobago, a system for monitoring standards cannot improve education unless there are adequate opportunities to learn within classroom settings, and this includes application by the teacher of strategies for formative assessment. Along this same vein, the reform of classroom assessment must come before further refinement of the national assessments of educational achievement. Teaching in the Anglophone Caribbean often minimizes the role of classroom assessment in student learning, with internal assessments often mimicking external assessment in intent and form. Teachers come to see the assessment purpose in the same way as the external agency, not to promote learning but rather to measure it and judge the worth of the student. Teachers function as measurers and even as judges but rarely s promoters of learning. This is true even in the secondary school despite the traditional presence of school-based assessment at the CSEC and CAPE levels.

A major factor in this distortion might be the misguided training that perpetuates misuse and mistrust of some assessment formats. Some teachers have come to attribute goodness and badness to different assessment formats. It may be advisable in our context to avoid the dualisms that exists in some US academic communities, which pit traditional against alternative forms of assessment. The practicing Caribbean teacher will need skills in all assessment formats and must be keen to use the right format for the right purpose. Multimodal assessment must become part of the assessment policy of all schools. It does seem advisable, however, to focus on performance, authenticity (realism), and feedback to students, key elements in most modern formative assessment systems. In Trinidad and Tobago, misunderstanding of the role and nature of classroom assessment is pervasive and therefore much of the possible positive impact on student learning might be lost. Quite simply, the evidence suggests that classroom assessment is most useful when it is formative; designed to promote learning by providing clear standards, multiple sources of feedback, and opportunities for students to work on meaningful, challenging, and authentic tasks. This kind of assessment will be integrated with modern constructivist teaching but not with the traditional transmission-type pedagogy that now dominates Caribbean classrooms.

While some have denied the existence of any tension between formative and summative intentions, the reality is most evident in the misuse of school-based assessment for CSEC and CAPE (Griffith, 2009). When does the teacher provide feedback and when does he or she measure learning; and can he or she do both at the same time; and if so, how effectively? Denying that such a tension exists may lead to implementation failure. It also leads planners and policymakers to reject the argument that high assessment literacy is fundamental to the implementation of both formative assessment and data use. The uniqueness of the Caribbean situation in terms of assessment practice provides an important backdrop to the issue of

implementing and evaluating a multi-component intervention such as the Continuous Assessment Programme in the primary schools of Trinidad and Tobago. The truth is that formative assessment is a difficult practice for teachers to implement in all parts of the world (Webb et al., 2004), but placed against the fragile foundation of traditional practice, beliefs about learning, poor or misguided training, and the lack of professional development in schools, the difficulty of installing quality assessment in the Caribbean classroom is magnified. Thus, any future reforms must be closely tied to increased professional development opportunities in general and the support and encouragement of coaches and professional learning communities at each building site.

In this regard, the experience of assessment reform in Asia has been instructive. Countries such as Hong Kong and Singapore are equally examination-oriented like those Caribbean. There is now evidence that traditional beliefs and practices do act as significant barriers to improvements and contribute to continued distorted practice. For example, Hong Kong's attempt to change assessment practice dates back to 1990 and is reinforced in a series of policy papers, most notably 2001 (Yu, 2006). Notably, Chan et al (2006) and Kennedy (2007) identified the many challenges to implementing the 2001 "Learning to Learn" reform in Hong Kong resulting from societal factors and teacher belief systems. Despite clearly written goals, Chan (2007) found that teachers continued to emphasize summative tests. As one respondent in her study said:

> We have various forms of assessments, like formative assessment and summative assessment. For example, we have developed observational records for each student to observe students' behaviours during their participation in classroom activities six or seven years ago. These records would be sent to the students and parents for their information… However, the most important aspect of assessment in my school is summative assessment (p. 9).

It may be important, then, for Caribbean systems to begin to study assessment reform and to construct theory that will help them change and align their systems with the trend towards improved classroom assessment practice.

Describing the Evaluand:
Continuous Assessment in Trinidad & Tobago

Evaluations are investigative studies designed and conducted to assist a target audience to assess an object's merit and worth (Stufflebeam, 2001). Thus, the intention of an evaluation is to obtain credible evidence necessary to judge the integrity and or impact of some programme or intervention. The object of the evaluation *the evaluand*. In this study, the evaluand is the Continuous Assessment Programme in Trinidad and Tobago. Details of the programme were found in two MOE published documents, *"Integrating Continuous Assessment into the Teaching and Learning Process Operational Manual*", obtained from the Ministry of Education and the *Operational Manual for the Pilot Phase* obtained from the Programme Designer.

The Continuous Assessment Programme (CAP) was installed in the Primary School system of Trinidad and Tobago in 1998, initially as a pilot followed by "full" implementation in 2000. The programme was developed as part of the Fourth Basic Education reform project and its recommendations for upgrading nationwide testing, assessment and evaluation (World Bank, 1995). These guidelines were included under Section D8, *Upgrading Nationwide Testing, Assessment and Evaluation*, and Annex D-2, *Terms of Reference for National Assessment Plan, Assessment and Evaluation Unit Structure, Terms of Reference for the Establishment of Computerized System.*

The framework referred to the monitoring of national standards and continuous diagnostic testing identified in the Task Force Report (1993-2003). This would require an enhanced assessment system operating within schools and an external monitoring system. There would also be need for data warehousing to ensure longitudinal monitoring of standards. In the model chosen, national tests were to be constructed externally but teachers were to be involved in adapting and analyzing the results for their classes. The Continuous Assessment Programme would fulfil a continuous diagnostic testing function, ensuring early identification of students with difficulties. Data from the CAP would be forwarded to the Division of Educational Research and Evaluation, Ministry of Education. This division had the responsibility of building a comprehensive database of student attainment.

BOX 1: Excerpt 1 from fourth Basic Education Plan

(8) Upgrading Nationwide Testing, Assessment and Evaluation

2.21 The CEE is a one-time placement test for access into secondary school. The CEE provides information on learning attainment levels for only Standard 5 students. There is no on-going nationwide monitoring of student achievement at lower grade levels. As a result, children with learning difficulties are not identified and supported on time to improve their performance. The NTFE has recommended the development of a continuous diagnostic testing and remediation support system, supported by two earlier national diagnostic tests in Standards I and 3. On a sector wide basis, the MOE lacks relevant, longitudinal information on student achievement to timely monitor and evaluate its services and develop appropriate policies and improvement plans (p. 6).

ANNEX D-2: Sector Assessment and Evaluation Reform
* Objectives and Outputs
* TORs for National Assessment Long-Term Plan
* Assessment and Evaluation Unit Structure
* TORs for Establishment of a Computerized Testing and Student Tracking

System
Objectives
The proposed assessment reform would support the improved delivery of the primary school curriculum by providing information on student achievement to (i) policy makers, (ii) education managers and supervisors and (iii) principals and teachers. Based on the generated information of student performance (i) policy makers would make necessary adjustment to sectoral strategies and propose investment projects(; ii) education managers would plan their support strategies to serve the varying needs of performing and non-performing schools; and (iii) principals and teachers would use feedback on student learning to improve classroom teaching assessment and remedial support, as well as to identify teacher training needs and are as to be addressed in their school improvement plans.

Outputs
* A Clearly Stated Plan for National Testing, Assessment and Dissemination of Student Achievement
* A Fully Established and Strengthened MOE Division for Research and Evaluation
* An Introduced and Fully Operational Computerized Testing and Student Tracking System

ANNEX D-2

Terms of Reference for the Development of A Long-Term Plan For National Assessment

The Ministry of Education and the sector stakeholders, through the National Task Force on Education, have called for a national plan to improve national testing in Trinidad and Tobago. The reform plan would include:

a) A clear strategy to develop, administer, core and report diagnostic assessment and student achievement in standard I through III, and as capacity and experience develops, it will encompass the Common Entrance Exam and school leaving exams for the primary and secondary programs;
b) The strategy must be driven by the instructional curriculum and be subjected to broad teacher review;
c) Quality control and security must be an integral part of any test development strategy, as well as reliability, validity, collaboration and sustainability;
d) Based on information generated by diagnostic testing at standard I and III, the MOE will build a comprehensive data base of continuous national attainment and develop a dissemination strategy on student performance to the school and classroom level; and
e) Develop a support system for student support and remediation. Curriculum facilitators would be trained to assess deficiencies identified in student test results and would provide assistance to improve classroom teaching and assessment methods. Principals and teachers would be trained in the new assessment methods and in follow-up strategies for teaching and assessment reforms, as well as in student remediation support.

The overall system is illustrated in Figure 2. As shown, the continuous assessment system would address, enhance and incorporate the majority of educational assessment within the primary school, including perhaps the exit examinations.

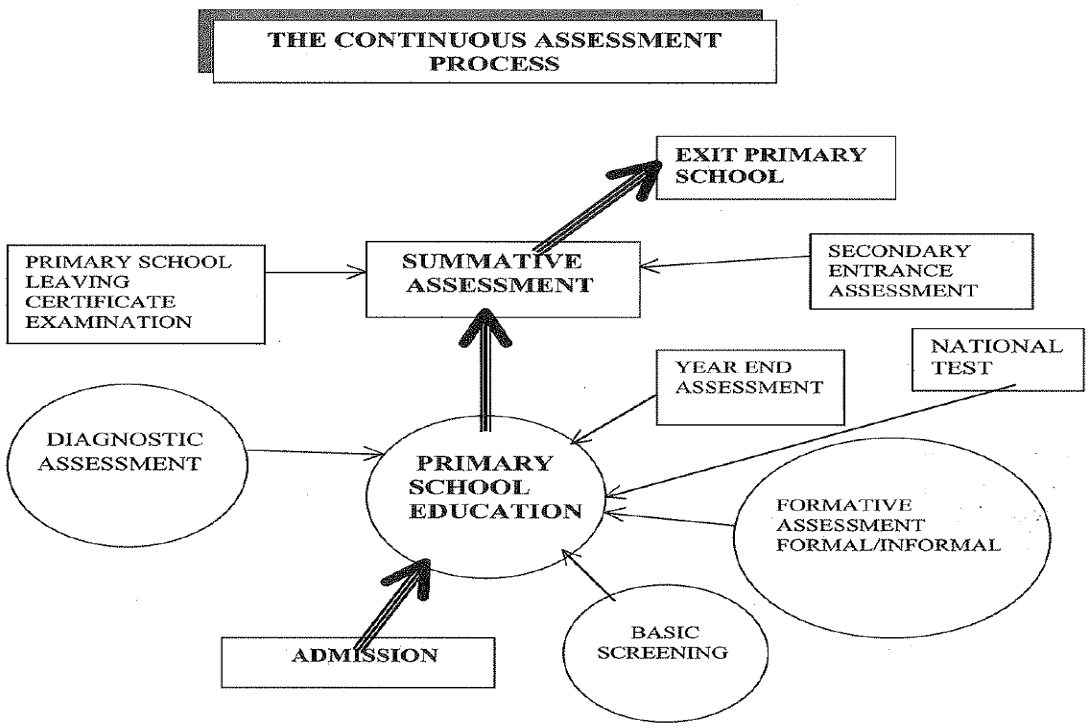


Figure 2: The role of the Continuous Assessment Programme in the system-wide assessment of students in the primary school

The CAP was, therefore a multi-component programme consisting of several separate components such as summative assessment, formative assessment, screening, and remediation. In practice, the CAP required teachers to conduct diagnostic (screening), formative (projects) and summative (weekly and monthly tests) assessments and to make use of data to inform teaching and learning. It included a significant "remediation" component in the form of systems for screening and intervention. The multiple elements, some of which had never been installed in the system prior, made the overall programme somewhat cumbersome.

Figure 3 illustrates the multiple components of the CAP programme, ranging from diagnosis to records and use of data. Considering the current level of professional development, the implementation of such a multi-component programme would require substantial support and resources, including training.

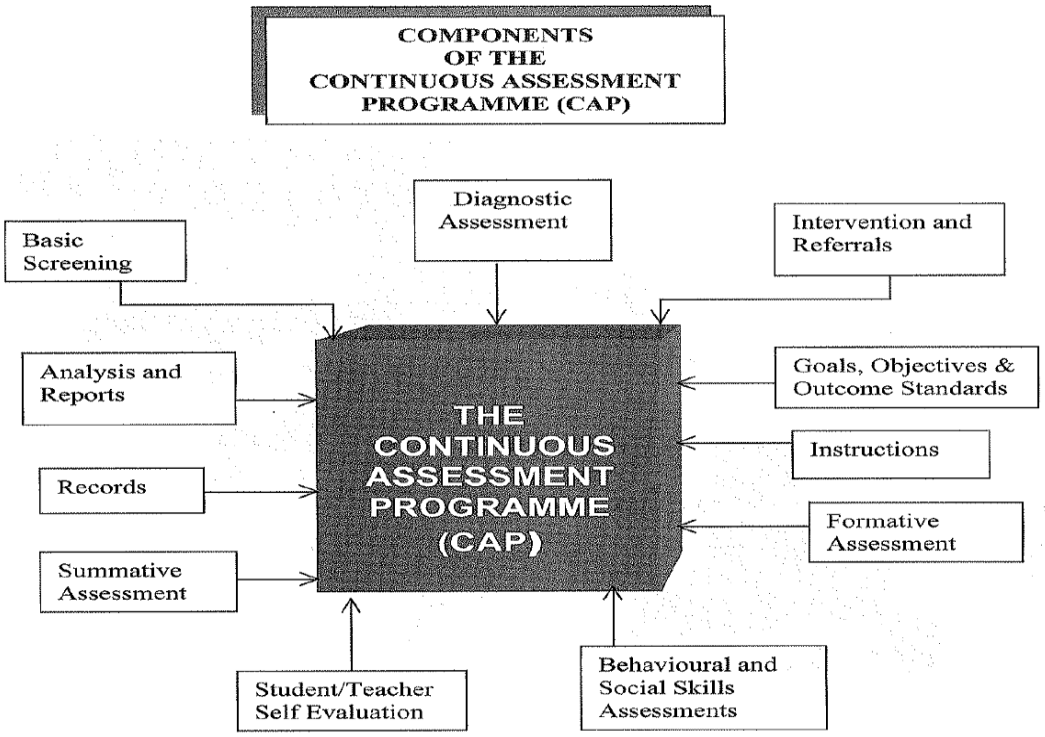


Figure 3: Multiple components in the Continuous Assessment Programme

Thus, the CAP turned out to be a rather ambitious but comprehensive programme of classroom assessment and data use, linked to the critical elements of the special education needs provision. The programme designer, Dr Janet Stanley-Marcano, in one of the elite interviews conducted for the study, suggested that the major purpose of the programme, in her mind, was to improve students' readiness to learn. With hindsight, the complexity of the multiple-component design might have increased the resistance experienced. Considering the unhappy history of education change in Trinidad and Tobago and with much hindsight, it might have been more advisable to focus, at least initially, on the core elements and delink it from other embryonic sub-systems in need of further development.

Nevertheless, it could be that if supported by training, resources, and appropriate leadership, the CAP might have been implemented with some degree of fidelity. Such high quality implementation could become an important lynch pin in education reform, able to change the quality of classroom interaction and student learning in the traditional Caribbean classroom. Several products were included in the outcomes expected from implementing the CAP programme. Teachers were expected to complete the student admission record, the student cumulative record card, performance records and journals. These products are listed Figure 4. However, it is notable that the documentation is very light on the use of the information gathered, except in the area of screening and intervention.

DOCUMENTS	USE	EXAMPLES
Student Admission Record	Official records of students' entry to the school system. Reference for assigning students' unique identifier	Standard Ministry of Education document
Student Cumulative Record Card (SCRC)	Official recording for individual students' school life experiences, behaviour and achievements. These cards will move with the student through the Primary and Secondary system	Standard Ministry of Educational document
Class Register	Recording student's daily Attendance	Standard Ministry of Education document
Teacher Work Plan and Assignment Schedule	Blue print for goal achievement standard for teaching/learning. Instructional strategies, Assessment Schedule, benchmarks	▪ Weekly Record and Forecast Scheme of Work for the term ▪ Daily Plan Book In school and Home Work Assignment
Students' Performance Records	Maintenance of data on student's challenges, strengths, weaknesses, intervention and impact. Tracking growth, preparation of reports and updates.	▪ Assessment, Tests ▪ Progress Cards ▪ Test and Assessment scores Anecdotal Records
Parent/Teacher Conferencing	To inform parents on student's challenges and progress and to formulate partnerships in order to promote student learning.	▪ Open Day
Teacher's Journal	To note day by day successes and challenges and to stimulate reflective thinking	

Figure 4: Products of the Continuous Assessment Programme

Managing & Implementing CAP

According to the Programme Designer, the directions for the management and implementation of the nationwide assessment programme were also contained in the Fourth Basic Education Proposal. These directions mandated the set up and restructuring of the DERE, with an Assessment and Evaluation Unit and a Computerized Tracking System. These latter recommendations were not incorporated into the local innovation. The Operational Manual for the Pilot Phase does, however, contain great detail on managing the assessment change. Several teams were to have a role in implementation and monitoring including the CAP secretariat and the Secretariat Monitoring team (Pilot Operational Manual, p. 24). Despite these directions, the document does exclude some practical suggestions that one would expect in a viable operational plan, especially concerning the change strategy.

BOX 2 Excerpt 2 from the Fourth Basic Education Plan

ANNEX D-2 Proposed Structure and Organization for an Assessment and Evaluation Unit (AEU) within the Division of Research and Evaluation

8. Proposed functions and responsibilities for the AEU include:

a) Conduct national student assessments and testing and distribute generated information for diagnosis, remediation and placement of students;
b) Coordinate information gathering in the performance of the education sector and establish and regularly update research and statistical data;
c) Provide relevant and update data to support sectoral studies, policy formulation and projected investments;
d) Ensure the input from special education agencies, guidance units, and the MOE Division of Curriculum Development (DCD) for test design, administration strategy, and student performance feedback to schools and other education agencies; and
e) Inform the public on the anticipated benefits of the testing and assessment reform, the implementation steps and the expected impact on school-level processes and outputs.

Organizational Structure and Staffing

9. Staffing for the Testing and Assessment Organizational Structure would include:'

a) An independent Teaching and Assessment Committee and
b) An experienced Director and other professional staff with teaching and subject experience to include:
c) An Administrative Examination Officer responsible for external and national examinations;
d) Four Examination Officers responsible for the development of assessment design, statistical analysis and examination papers in the field of (i) Science and Mathematics, (ii) English and other languages, and (ii) Social Studies/Humanities.
e) A Quality Control Officer to be responsible for work flow, security and accounting; and
f) Four Test Technicians - to collect, enter, and maintain each of the data bases associated with English/Foreign Languages, Math, Science and Social Studies /Humanities. 'The proposed staffing description does not necessarily represent additional personnel but redefinition of functions and responsibilities of existent staff m the DRE.

Computerized Student Tracking and Testing and Assessment System Objective

10. To use the latest computer technology, utilizing cost-effective PC desktop hardware and software, for the construction of professional quality tests, performance assessments, surveys, and lesson plans' publishing.

Technical Background Needed by DRE Staff

11. The following technical areas would be developed within the DRE through an initial six-month internship for selected staff and by in-service training and technical assistance for two-three years during project implementation:

a) Principles and theory of measurement
b) Testing and student tracking computer programs and systems
c) Use of scanning, scoring, graphics and file management programs
d) Principles and operations of databases and the management of algorithms and retrieval strings
e) Student performance data bank system including programs for answer sheets, statistical analysis and report writing, capture of concomitant variables by surveys to extend the value of the achievement data, item analysis of field trials and dissemination of test results.

Terms of Reference Interpreted

As formally defined for this report, an evaluation is a study **designed to provide a credible retrospective assessment of the merit, worth and value of the management, output and outcome of a government intervention that is intended to play a role in future, practical action situations** (Vedung, 1997, p. 13).

This evaluation consultancy was designed to:

(1) Determine the degree to which the Continuous Assessment Programme has been effectively implemented in the 60 sample SES schools
(2) Investigate factors that hinder implementation of the CAP
(3) Develop recommendations for improving the system of continuous assessment to support improved teaching and learning as well as student academic performance.

To achieve these objectives the consultant was required to:

(1) Work closely with the Directors of the DERE, DCD, and Seamless Programme Coordinating Unit
(2) Conduct a literature review on CA and assessment as learning use.
(3) Develop an evaluation research design for conduct of the study
(4) Develop data collection instruments for the schools
(5) Conduct data collection, inclusive of key informant interviews
(6) Review MOE policies and develop recommendations based on the evaluation.

The deliverables included

(1) A work plan
(2) A midterm report
(3) A draft final report at the end of July

The work plan and midterm report included some details of the evaluation design, which have been expanded in this final report. The midterm report also included a summary of the design and use of the major data collection methods and a review of MOE policies and recommendations. These were also presented orally to the DCD and DERE.

The results are significantly expanded for this final report and the recommendations are further streamlined. The final report also contains a comprehensive literature review of continuous and formative assessment, including assessment reform and change. These will also be presented in workshops to the DCD and DERE.

Literature Review

Defining Continuous Assessment

Le Grange and Reddy (1998) defined continuous assessment as *"the assessment of the whole learner on an ongoing basis over a period of time, where cumulative judgments of the learner's abilities in specific areas are made in order to facilitate further positive learning*" (p. 11). To varying degrees, then, continuous assessment might include holistic, ongoing, or multiple assessments by teachers in the classroom. This kind of assessment is likely to be integrated with the curriculum and with the teaching-learning process. The assessments may be informal or formal, formative or summative, producing data for decision-making, ensuring feedback to students, and providing information for setting targets and standards for classrooms.

Nitko (1995) has argued that the multiplicity of perspectives makes continuous assessment confusing for stakeholders and encouraged policymakers to develop an organizing framework to provide a common platform for understanding what is meant by the term. I tend to agree with this approach. Nitko in establishing his own framework distinguished between several aspects of the process, such as formal vs. informal; formative vs. summative; and instructional vs. official. Lack of clarity might have been a problem in the CAP documentation, with little indication of the emphasis to be placed on either formative or summative assessment. Interestingly some conceptualizations of continuous assessment in the African context appear to stress continuous testing as captured in the early work of Crooks (1988). It may be that this kind of approach to continuous assessment differs from approaches that emphasize informal and formative assessment.

Table 1: Differences between traditional and continuous assessment (adapted from Le Grange & Reddy, p. 11)

CRITERIA	TRADITIONAL ASSESSMENT	CONTINUOUS ASSESSMENT
FOCUS	Is mainly made up of written examinations that take place in formal settings.	Is made up of a variety of assessment methods that can be formal or informal.
USE	Is used to decide whether or not the learner is promoted to the next grade.	Is used to inform the learning process through which learning outcomes are required.
WHEN	Takes place after the learning process at dates and times previously decided on (Summative).	Takes place during the learning process when it is considered necessary (formative).
SCORING	Is mostly norm referenced rather than criterion referenced	Makes use of criterion referencing than norm referencing.
DATA	Provides isolated marks or percentages to show how learners have changed	Provides information in context as feedback on how learners are changing.

Le Grange and Reddy (1998) distinguished between traditional and continuous assessment using the criteria of focus, use, timing, scoring, and data. These differences are summarized in Table 1. Continuous assessment, including both written as well as performance measures,

makes use of criterion-referenced benchmarks and provides information in context as feedback to learners. For this evaluation, I consider continuous assessment to consist of both formative and summative assessment practices, although I will argue that it is the formative, which provides the greater impact on learning. In the definition, I also consider the range of assessments used over a period and the data used to both capture and inform the progress and learning of students.

Elements of Continuous Assessment in Trinidad and Tobago

The Operational Manual for the Pilot Phase listed and described in great detail the 9 procedural steps of CAP illustrated in Figure 5. These were: (1) admission, (2) recording, (3) screening, (4) intervention, (5) diagnosis, (6) referral, (7) classroom activities, (8) analysis and reports, and (9) evaluation. There is considerable overlap in these "steps" because admission and record keeping also involves data collection activity and referral is one of several intervention strategies. As defined in the Pilot manual, these steps also involve the conduct of formal and informal assessments such as observations.

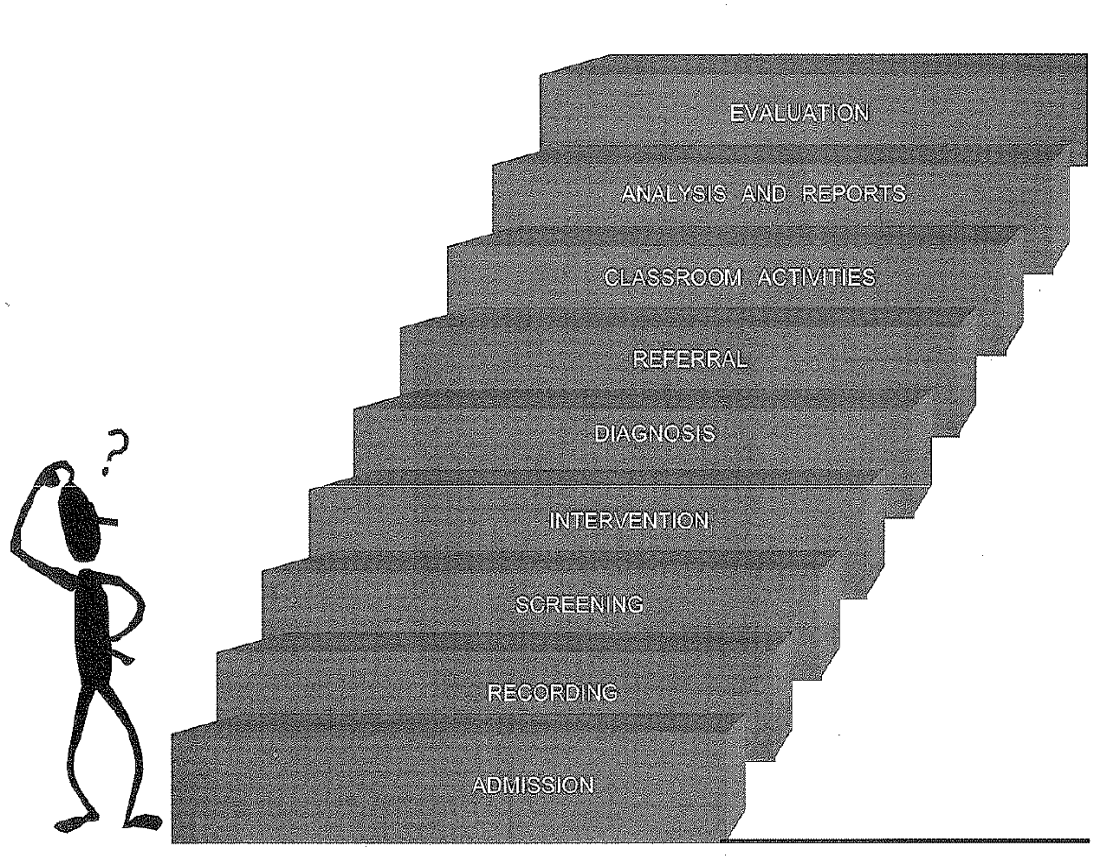


Figure 5: Nine Operational Procedures in CAP.

The process as outlined is somewhat cumbersome because diagnosis, intervention and referral, in reality, constitute one connected event. Likewise, record keeping is a data collection process leading to the use of that data in referral or evaluation. "Classroom activities" is a vague term that might include critical tasks such as the conduct of formative assessment and regular summative assessment, including goal setting, which is a component of formative assessment. Analysis and reporting along with evaluation are also data use steps.

Thus, there are in fact two key stages, the conduct of the assessment and the use of the data from that assessment, the latter including analysis, reporting, and the action or decision that is taken. This reclassification of the CAP procedures is illustrated in Table 2.

Table 2: Reclassification of procedures and activities in the Trinidad and Tobago CAP

CAP Procedure	Sub-Activity	Classification
1) Admission		• Data Collection
2) Recording		• Data Collection
3) Screening	▪ Screening ▪ Observation ▪ Action	• Conduct Screening Assessment • Data Use Action/Decision
4) Intervention	▪	• Conduct Diagnostic/Formative Assessment • Data Use Action/Decision [Teacher]
5) Diagnosis	▪	• Conduct Diagnostic/Formative Assessment • Data Use Action/Decision
6) Referral	▪	• Conduct Diagnostic/Formative Assessment • Data Analysis & Reporting (N=1) • Data Use Action/Decision [SSS Team]
7) Classroom Activities	▪ Goal Setting ▪ Instruction ▪ Formative Assessment ▪ Summative Tests used for Formative Purposes ▪ Summative Assessment	• Conduct Dynamic, Formative & Summative Assessments
8) Analysis & Reports	▪	• Data Analysis & Action
9) Evaluation	▪ Formative Evaluation ▪ Summative Evaluation ▪ Action Research	• Data Use Action/Decision

Screening, Diagnostic, Formative and Summative Assessments

Essentially, CAP specified four different types of assessments: (1) screening, (2) diagnostic, (3) formative, and (4) summative. However, it is possible that *developmental screening assessments* as distinct to *measures of readiness* might have a role in the early primary school, even in admissions process; however, the CAP documentation does not encourage this kind of activity in admissions, perhaps for fear of labelling or other negative action. Instead, it narrowly focuses on the use of screening assessments meant to detect indicators of non-cognitive problems that can be ameliorated with early intervention. For educational intervention, diagnostic assessments administered by a teacher or SEN specialist are recommended. In reality, such diagnostic assessment would come after a cycle of formal and informal formative assessment which would provide a picture of the difficulty the student has

in the classroom. This process, however, is not made clear in the documentation and did not exist as regular special education needs (SEN) practice at the time.

The CAP operational plan implicitly emphasized the use of dynamic or flexible interactive assessment meant to help the student progress in learning and learning capacity. However, one might imagine that the pre-assessments in the formative assessment cycle might also be used diagnostically for targeted action in addressing issues of differential learning needs. Pre-assessments will provide instructors with information about student's prior knowledge and misconceptions before beginning a learning activity. They might also provide a baseline for understanding how much learning has taken place after the learning activity is completed. The CAP's understanding of monthly and termly tests, according to the Operational Manual for the Pilot, is that these tests are meant to inform the teacher about students' progress and therefore the intention is to use these summative tests formatively. However, according to the Programme Designer, student scores were also forwarded to the Ministry of Education (MoE), implying some other action.

Formative Assessment as the Centrepiece of Continuous Assessment

As already indicated, all classroom assessment may not necessarily be formative; however, the argument here is that much of it should be. Again, Black and Wiliam (2004) reminded us that:

> The terms classroom assessment and formative assessment are often used synonymously, but . . . the fact that an assessment happens in the classroom, as opposed to elsewhere, says very little about either the nature of the assessment or the functions that it can serve (p. 183).

The key element in enhancing student learning would be the formative assessment practice, which by definition would include the screening and some diagnostic assessments identified in the CAP operational manual for the pilot schools. For this and other reasons, this evaluation study regards formative assessment as the essence of the continuous assessment practice described in the CAP documents, although this might not be explicitly stated. Indeed, the OECD's definition of formative assessment is equally applicable to continuous assessment. They defined formative assessment as **"frequent, interactive assessments of student progress and understanding to identify learning needs and adjust teaching appropriately"** (OECD, 2005, p. 21).

The OECD's definition, however, does not fully capture the intent and best practice involved in the formative assessment process, including the interactive and dynamic nature of the process and the important provision of feedback. Indeed, formative assessment is distinct from continuous summative assessment or even using performance and authentic assessments. The latter is an important distinction because some teachers and teacher educators equate formative assessment with authentic and performance assessment (Ainsworth & Viegut, 2006; Bell & Cowie, 2001).

Recently, experts in the field have addressed the issue and considered useful working definitions of this essential educational intervention. The four definitions supportive of

formative assessment as visualized by this group are summarized in a 2009 position paper on assessment for learning (Davies et al., 2009. pp. 4-5). The four definitions in Box 3 highlight the fact that formative assessment is often a planned process that involves the collection of evidence on student learning, with that evidence used to guide and direct various aspects of teaching and learning. Critical to the process as evident in definition 3 is providing feedback to the student. For the purpose of this evaluation, then, the McManus (2008, p. 3) definition which includes the process of feedback and improvement in teaching-learning provides the clearest focus and is aligned to the understanding of the process as described in the documentation.

Box 3: Four working definitions of formative assessment endorsed by the 2009 position paper on assessment for learning.

1. '*Assessment for Learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there'*. Assessment Reform Group (2002)

2. *'Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited'*. Black & Wiliam (2009).

3. *'Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes.'* McManus (2008).

4. '*Formative assessment is a planned process in which assessment-elicited evidence of students' status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics.'* Popham (2008).

Bell and Cowie (2001) suggested that there were several unique characteristics of formative assessment. The listed nine, including (1) responsiveness, (2) sources of evidence, (3) tacit process, (4) use of professional knowledge and experiences, (5) integration within teaching and learning, (6) done by both teachers and students, (7) use of multiple purposes, (8) contextualized nature, and (9) the presence of dilemmas. Perhaps the most important focus in terms of the analysis of CAP is the responsiveness or dynamic nature of the process, the focus on evidence, and the integration with learning. The Operational Manual for the Pilot does list formative assessment under classroom activities, but perhaps more emphasis might have been placed on helping teachers understand that formative assessment and teaching-learning are not separate activities.

Bell and Cowie (2001) also distinguished between planned formative assessment and interactive formative assessment. The latter is an important part of the dynamic nature of assessment as teachers respond to the multiple and hard to predict interactions in a classroom. Several formats may be used in the process of formative assessment. Questions and answer lessons and conversations are commonly used in interactive formative assessment. Several formal methods may be used to gather evidence in the pre-assessment or formative stage; including quizzes, projects, and homework assignments using different assessment formats. There are also strategies like graphic organizers, which will help students interact with topics.

Carless (2007) built on the concept of interactive formative assessment and distinguished between whole class and individual formative assessment. Further, they conceptualized pre-emptive formative assessment or anticipatory feedback defined as "teacher actions which attempt to clarify student understandings before misconceptions have resulted in ineffective learning or performance and/or loss of marks in assessed tasks" (p. 176). Earl (2003) considered the three different purposes of assessments, (1) of learning, (2) for learning and (2) as learning. Assessment for learning and assessment as learning are both types of formative assessment; but assessment as learning is more fully embedded in the instructional process and is designed to enhance the students' self-regulating skills. Several new policy documents in developed countries include the distinction between these three purposes.

Students are at the centre of this kind of assessment when they use personal goals and external standards to judge their own work. This form of assessment is similar to what Stiggins (1994) earlier called student-centred or involved assessment. Another useful distinction made by Pyror and Crossouard (2008) is between divergent formative assessment and convergent formative assessment. Convergent formative assessment has behaviorist elements and is structured whereas divergent formative assessment extends the thinking of students. This assessment required flexible or complex planning and involved the use of open-ended authentic tasks. Centring formative assessment on the provision of feedback to learners is aligned with current theory on the nature of formative assessment. It is feedback which acts to enhance motivational and cognitive processes in the learner thereby ensuring higher student achievement (Shepard, 2000). Feedback is the critical element in forming learning because it carries the learner to the desired endpoint, providing scaffolding so that learners can accomplish more advanced activities and engage in advanced thinking and problem solving (Shepard, 2005, 2006). Formative assessment achieves these different goals by acting to (1) reduce the level of uncertainty, (2) reduce the cognitive load of the learner, and (3) allow the student to correct inappropriate task strategies, procedural errors, or misconceptions.

Black and William (2009) in explicating a theory of formative assessment in practice identified five essential elements in formative assessment as illustrated in Figure 6. For the teacher, formative assessment involves the following steps: (1) clarifying the learning intentions and criteria of success (for example, rubrics); (2) engineering tasks that elicit understanding (for example authentic performance assessment); (3) providing feedback in many ways to the learner so that they can move forward; (4) fostering peer learning and self and peer assessment; and (5) activating or empowering the learner to become self-regulating. Pre-assessments and goal setting are included under the rubric of "where the learner is" the conduct of the assessment is critical to determining where the learner is at this point.

<table>
<tr><th></th><th>Where the learner is going</th><th>Where the learner is right now</th><th>How to get there</th></tr>
<tr><td>Teacher</td><td rowspan="3">1 Clarifying learning intentions and criteria for success

Understanding and sharing learning intentions and criteria for success

Understanding learning intentions and criteria for success</td><td>2 Engineering effective class-room discussions and other learning tasks that elicit evidence of student understanding</td><td>3 Providing feedback that moves learners forward</td></tr>
<tr><td>Peer</td><td colspan="2">4 Activating students as instructional resources for one another</td></tr>
<tr><td>Learner</td><td colspan="2">5 Activating students as the owners of their own learning</td></tr>
</table>

Figure 6: Essential Elements of Formative Assessment in Practice (Black & Wiliam, 2009, p. 8)

Focusing upon the student primarily, Shute (2008) defined *formative feedback* as information communicated to the learner that is intended to modify his or her thinking or behaviour for the purpose of improving learning (p. 254). Feedback may vary by content, function, and presentation. Using the dimension of complexity, Shute also identified several different types of feedback as illustrated in Table 3. It might be that more complex forms of feedback are more effective. However, the evidence for this assertion remains limited. As Shute has indicated, there are several different dimensions to feedback function and presentation. Thus, feedback may vary in timing, quantity, mode, and audience. Feedback may also be response specific or goal directed. Chappuis (2009) noted that it is not just the presence of feedback alone that creates the enhanced achievement effect, but it is the quality of that feedback. High quality feedback is descriptive, occurring during learning, addressing misconceptions or partial understanding, encouraging the student to reason on their own, and focused on the amount of advice the student can use.

Chappuis (2009) makes an interesting distinction between *success feedback* and *intervention feedback*. Success feedback points out what the student has done well and encourages the student but intervention feedback focuses upon a feature of quality or a problem that needs further work. Both success and intervention feedback are effective because they provide verification and elaboration. Verification is the judgment of whether an answer is correct and elaboration is the informational aspect of the message, providing relevant cues to guide the learner toward a correct answer (Shute, 2008).

It must be stressed that continuous assessment is not a strategy or an event, but a process. There are several models of the formative assessment process in the current literature. Built into the steps of Black and Wiliam (2009), for example, is an implicit process of investigating (1) where the learner is going, (2) where the learner is right now and (3) how to get there. Chappuis (2009) has organized the process into seven sub-steps and three primary targets as listed in Box 4.

Table 3: Different types of feedback arranged in levels of complexity (Shute, 2008, p.

Feedback Type	Description
1. No feedback	Refers to conditions where the learner is presented a question and is required to respond, but there is no indication as to the correctness of the learner's response.
2. Verification	Also called "knowledge of results" or "knowledge of outcome." It informs the learners about the correctness of their responses (e.g., right–wrong, or overall percentage correct).
3. Correct	Also known as "knowledge of correct response." Informs the learner response of the correct answer to a specific problem, with no additional information.
4. Try again	Also known as "repeat-until-correct" feedback. It informs the learner about an incorrect response and allows the learner one or more attempts to answer it.
5. Error flagging	Also known as "location of mistakes." Error flagging highlights errors in a solution, without giving correct answer.
6. Elaborated	General term relating to the provision of an explanation about why a specific response was correct or not and may allow the learner to review part of the instruction. It may or may not present the correct answer (see below for six types of elaborated feedback).
7. Attribute isolation	Elaborated feedback that presents information addressing central attributes of the target concept or skill being studied.
8. Topic contingent	Elaborated feedback providing the learner with information relating to the target topic currently being studied. May entail simply re-teaching material.
9. Response contingent	Elaborated feedback that focuses on the learner's specific response.
10. Hints/cues/ prompts	Elaborated feedback guiding the learner in the right direction, e.g., strategic hint on what to do next or a worked example or demonstration. Avoids explicitly presenting the correct answer.
11. Bugs/misconceptions	Elaborated feedback requiring error analysis and diagnosis. It provides information about the learner's specific errors or misconceptions (e.g., what is wrong and why).

---------- **Box 4: Formative Assessment Steps**----------

Where am I Going?

1. Provide students with a clear and understandable vision of the learning target.
2. Use examples and models of strong and weak work.

Where am I Now?

3. Offer regular descriptive feedback.
4. Teach students to self assess and set goals.

How Can I Close the Gap?

5. Design lessons to focus on one learning target or aspect of quality at a time.
6. Teach students focused revision.
7. Engage students in self-reflection and let them keep track of and share their learning.

Tuttle (2009) provided a useful series of steps that includes pre-assessment to collect baseline data. These steps are compatible with new standards-based assessment systems used in the classroom.

1) **Pre-assessing students**
2) **Sharing learning goals with students**
3) **Sharing or co-creating of learning criteria with students**
4) **Employing quality classroom discourse and questioning**
5) **Using rich and challenging tasks that elicit students' responses**
6) **Identifying the gap between where the students are now and the desired standard goal**
7) **Providing feedback that helps students identify how to improve**
8) **Using self-assessment and peer assessment**
9) **Providing students with opportunities to close the gap between current and desired performance**
10) **Celebrating learning progressions**

Heritage (2010) recently constructed one of the more comprehensive descriptions of formative assessment in describing what she called the learning cycle. Heritage's model focuses upon collecting evidence for learning and like Shepard (2000, 2005, 2006), she considered the cognitive implications of scaffolding and other strategies designed to help students close the gap. This seems to be a useful model in the context of Trinidad and Tobago, where there are often significant achievement gaps on different factors. Heritage's model includes the learning goals and criteria for success as does Shirley Clarke (2001, 2005, 2008), but the steps in which evidence is elicited and interpreted follows closely the early work of Wiliam and Black (1998). As expected, feedback is a critical element in the loop followed by activities designed to respond to these learning needs. This feedback includes scaffolding, which helps students bridge the gap. The model is useful because it explains the way the cycle is repeated several times in an extended authentic assessment.

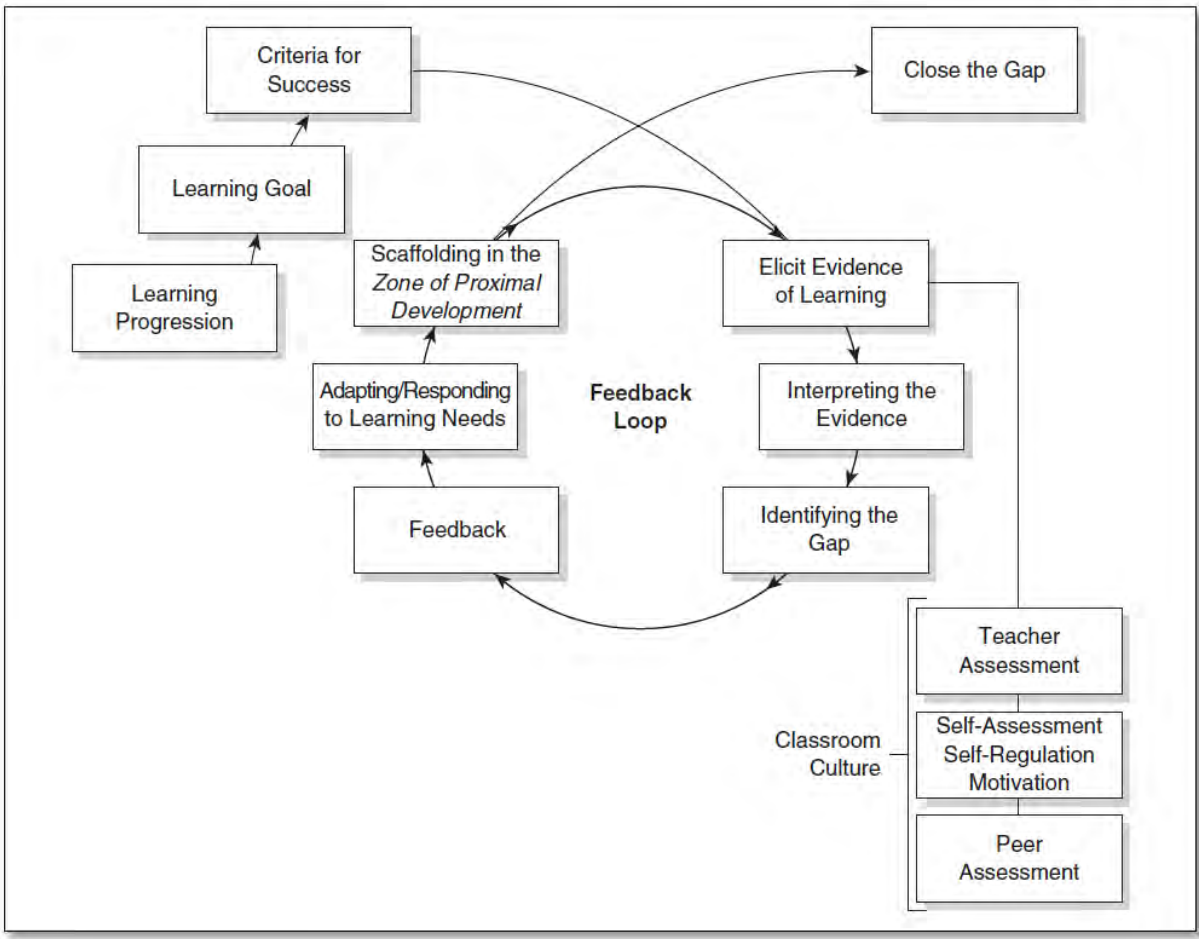


Figure 7: Formative Assessment in the learning cycle according to Margaret Heritage (2010, p. 11)

Multimodal Assessment & Assessment Literacy

The CAP assumes use of several different assessment formats, although it does not explicitly describe these assessments in detail. It was assumed perhaps that teachers had enough assessment literacy to construct different types of tests well. However, it was not considered that if teacher tests lacked validity and reliability the quality of the data would prohibit efficient data use. Certainly, invalid data will make an entire system of teacher led high stakes assessment inefficient and corrupted. Considerable attention was given in the CAP documentation to the role of assessment projects in the Operational Manual for the pilot. It is assumed that this emphasis was meant to indicate the role of authentic performance tasks in the formative assessment function and the ability to capture multiple intelligences when well designed. Although these assertions might be true, project based assessment will not by itself fulfil all of the complex criteria involves in true formative assessment.

Additionally, it might be useful to divorce the vehicle of the assessment from the tasks involved in constructing a performance assessment. The failure to do this in both training and certification assessment courses has resulted in a de-emphasis on task design and rubric development. Thus, in reality, most of the benefits of authenticity and performance might have been lost and formative assessment processes such as feedback have not been captured in local practice in the classroom. This is still a salient point because one of the targets in the Seamless Project is to obtain "technical assistance for training in performance assessment, mentoring and pedagogical skills". This assumes, of course, that the performance assessments

are integrated with teaching-learning, formative, and authentic. A more appropriate focus, then, would be on constructing and using performance based tasks for formative classroom assessment.

Table 4 provides a listing of the different assessments outlined in the CAP along with a possible role for new assessments such as benchmark and interim assessments commonly found in some assessment systems in developed countries. For each assessment, the different possible formats are listed along with purpose and role. In the future, it might be important to help teachers understand that their primary use of assessments must be formative and this should supersede the use of even national tests in determining the status of *individual learners*. The table does not include the use of the SEA to measure school performance, since this is quite an inappropriate usage. Public examinations scores should not be used to evaluate systems and schools because they contain information strongly related to factors external to the school (London, 1989).

The issue of multimodal assessment and the need for test construction skills among teachers points towards the need for the original CAP documentation to explicitly consider assessment literacy as part of the innovation. Stiggins (1991, 1995) originally coined the term "assessment literacy" referring to the necessary skills required for competent assessment practice in the classroom. Stiggins' focus was on the role of assessment in school improvement and he suggested that more had to be done in the way of professional development with US teachers in order for teachers to get the full value from assessment. Schaefer (1993) looked at the teacher education courses in the US and found that several programmes back at that time did not contain a measurement course and when such a course was situated, the content was not usually aligned with the required skills of practicing teachers. For Trinidad and Tobago, assessment training at all levels has been notably limited and this is not a problem to disappear in the near future.

The issue in 2010 is that the assessment requirements of teachers have expanded tremendously in Trinidad and Tobago and elsewhere. In the current climate, local primary school teachers must have competence in multimodal, formative, and accountability assessment as well as high stakes testing systems, such as the SEA. Popham (2009) noted a similar problem in the US with the increasing need for teachers to understand the role of different assessments, especially for accountability testing in the No Child Left Behind (NCLB) era. In the international community, a recent addition to the complex of needed skills and competencies is for teachers to use assessment data in data driven decision making. Thus, the problem of competent practice is enormously multiplied within the local context and it would be quite impossible to implement assessment reform without an elaborate training and retraining programme.

Table 4: Different assessments, assessment formats, purposes and roles

Type	Format	Primary Purpose/Use	Primary Audience	Assessment Role
Secondary Entrance Assessment	Constructed Response (Discrete & Extended)	Determine Performance of Individual Student	Parent	System Placement
National Tests	Multiple-Choice, Constructed Response (Discrete & Extended)	Determine Performance of School & Districts	MOE	Summative
		Determine School Improvement	MOE, District, School, Teacher Teams	Summative
		Develop Programmes/ Interventions	Teacher Teams, School	Formative
***Benchmark, Interim or School-wide Assessments**	Multiple-Choice, Constructed Response (Discrete & Extended)	Determine School Improvement Status	District, School	Formative
		Develop Programmes/ Interventions	District, School, Teacher Teams	Formative
Classroom Assessment	Multiple-Choice, Constructed Response (Discrete & Extended)	Determine Report Grades	Teacher Teams, Individual Teachers	Summative
		Diagnosis Provide feedback	Teacher Teams, Individual Teachers	Formative
	Performance	Provide feedback	Teacher Teams, Individual Teachers	Formative
		Determine Report Grades	Teacher Teams, Individual Teachers	Summative
	Observations/ Journals/ Anecdotal Records/Checklists	Screening Provide feedback	Teacher Teams, Individual Teachers	Formative

*Not included in the original CAP documentation

Webb (2002) defined assessment literacy as the "knowledge of means for assessing what students know and can do, how to interpret the results from these assessments, and how to apply these results to improve student learning and program effectiveness" (p. 1). The concept of assessment literacy has been built into *The Standards for Teacher Competence in the Educational Assessment of Students* published by the American Federation of Teachers, National Council on Measurement in Education, & National Education Association, 1990). The standards are listed in Box 5.

--Box 5: Standards for Teacher Competence in the Educational Assessment of Students--
STANDARD 1 – Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
STANDARD 2 – Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
STANDARD 3 – The teacher should be skilled in administering, scoring and interpreting~ the results of both externally produced and teacher produced assessment methods
S**TANDARD 4** – Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement.
STANDARD 5 – Teachers should be skilled in developing valid pupil grading procedures that use pupil assessments
STANDARD 6 – Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators.
STANDARD 7 – Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

These standards provide useful guidance for directing continuous professional development in assessment on a national basis. The focus is on choosing the right assessment, developing assessments, making decisions, communicating results and recognizing improper practices. Several new models of professional development in assessment have recently been developed and might also be considered in new training or certification programmes (Webb, 2009). There is also some useful work specifically on professional development for formative assessment (Chappuis et al., 2005; Wylie, Lyons, & Goe, 2009).

Geddes (2005) reported on the work on assessment literacy in New Zealand with the Assessment to Learn (AtoL) Project. The development model used in this project (discussed later) focused upon teachers using an action research iterative model of change embedded in a school culture of professional learning. Despite the theory-based nature of the programme, however, Geddes (2005) commented that:

> Although our data suggests the development was successful, informal conversations with teachers highlighted that the changes are fragile and a few teachers are still struggling to make changes at all. Research suggests that the acquisition and transfer of new knowledge and skills require a great deal of support (p. 51)

The General Teaching Council of England (GTC) funded a project designed to better understand ways to enhance the teachers' role in student assessment (GTC, 2004). The document calls for achieving the right balance between teacher formative assessment, teacher

summative assessment and measures of accountability. In the project paper, Harlen (2004) reviewed the work on teacher led assessment throughout the world. She concluded that:

> The evidence from the studies reviewed suggests that teachers are more reliable in their assessment when they have a good grasp of the criteria, which will help in identifying relevant evidence as well as in making judgements of it (p. 28).

She suggested several strategies for improving the validity and reliability of teacher assessments, including moderation, development of criteria, specification of item/task banks, and enhanced teacher training.

One of the important points made by the programme designer of the Trinidad and Tobago CAP is that the proposed system was to be strongly teacher-centred, with plans to involve teachers even further in the data collection process. However, the implementation framework did not include sufficient specification of tasks and criteria nor did it provide an elaborated framework for teacher professional development that would have allowed teachers to become proficient measurers and judges. The point here is that modern systems that make use of teachers in external assessment, such as New Zealand and Sweden, pay added attention to the quality of professional development and therefore can better rely on the data collected. This cannot be said locally and thus any intention to make use of scores created in poorly constructed teacher assessments in Trinidad and Tobago for accountability purposes or otherwise might be considered foolhardy at this time.

Data Use Processes

The evaluand is very heavy on the use of the assessment data and identifies several actions that teachers and others must take based on the results collected from the conduct of assessments (See Table 2). Admittedly, assessing data driven decision-making was not fully incorporated in the evaluation design even though DDDM is a key research area for the evaluator. This might have been short sighted because on repeated analysis of the Operational Manual for the Pilot it became clearer that data driven decision-making was central to the effective operation of CAP in the schools. The teacher was expected to use data collected from the conduct of continuous assessment to intervene in their own classroom or initiate the referral process (with the involvement of external agencies). However, the documentation did not fully elaborate on a system that would allow the teacher to move smoothly from the data to the action (decision). This use of data is not common in Trinidad and in the past little attention has been paid in the past to data-driven instruction. Data-driven instruction (DDI) is a method of collecting student data where teachers accurately assess student learning in order to take actions (decisions) that are based on that data in an attempt to improve instructional systems and to continually promote student achievement.

DDI in a school is part of a wider process of data-driven decision-making (DDDM) defined here as the process in which student assessment and other relevant data is collected and used to inform decisions related to planning and implementing strategies at the district, school, classroom, and individual student levels. DDI is different to DDDM in the scope of both the decisions made as well as the data used. DDDM more frequently makes use of data outside of student assessment and includes decisions that extend to school organization and school

improvement. DDI is focused on individual or groups of students. However, both DDI and DDDM can be collaborative as groups of personnel sit and plan.

The value of installing a formal DDDM or DDI system connected to the conduct of classroom assessment is that it makes clear procedures and resources to move from the conduct of assessments and collection of data to actions and decisions, as illustrated in Figure 8. As shown, the data must be converted into user friendly information before it can become usable. Analysis and inference from this type of information leads to knowledge and insight into the status of the student. Such insight may be used in decision making to lead to an informed decision or action with regards to the further training, re-teaching or other such instructional decision. Facilitators include organizational and system supports such as time for collaborative decision-making. An important facilitator would be technology since the data is hard to manipulate in the current form.

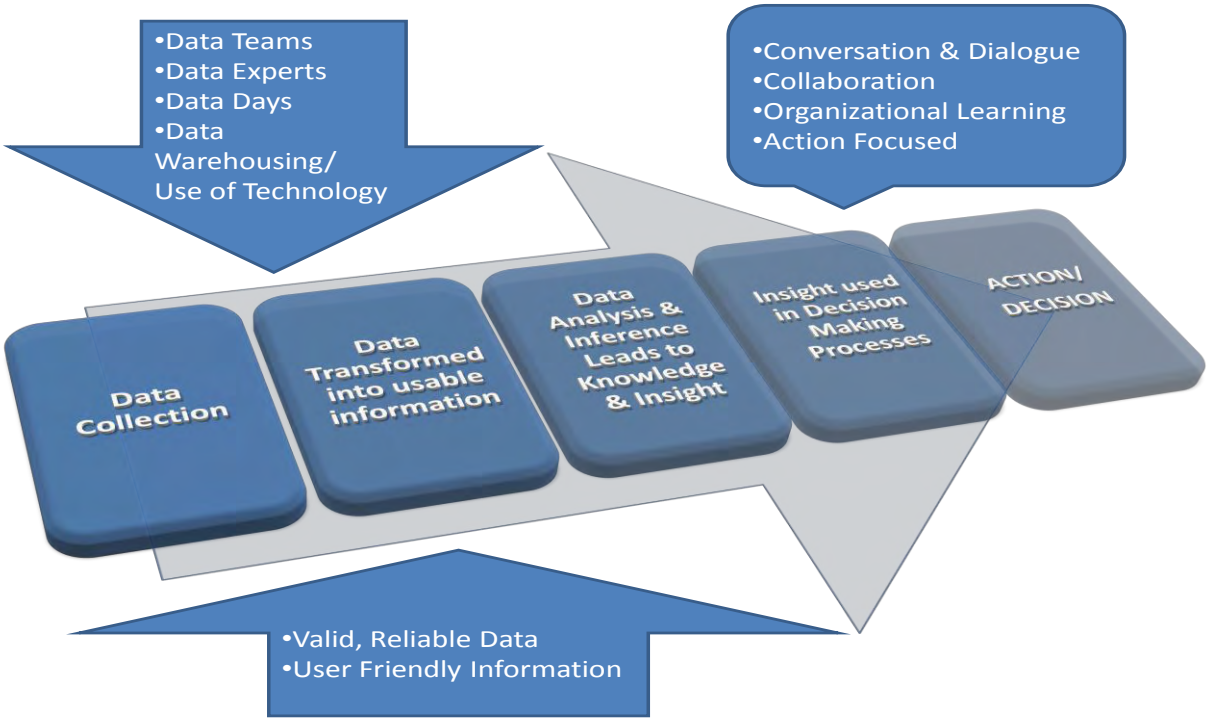


Figure 8: DDDM process with facilitators

Even in a system with external summative assessments such as Trinidad and Tobago, data driven instruction based solely on internal continuous assessment will have added value. This is because the data is real-time and can reveal vertical content and class level mastery in all subject areas, rather than only on the selected areas in a national test in which there is delayed feedback. It is continuous assessment data (assuming that it is of high quality) that must be used to make key decisions about individual students in the school and not information from national tests or the high stakes Secondary Entrance Assessment.

Differentiating between formative assessment and data use

It is important to draw clearly the boundaries around formative assessment, continuous assessment and data driven instruction. A formative assessment system is not equivalent to a

data-driven instructional system. The latter will make use of data from both formal summative and formative classroom assessments to inform and drive learning. Critical to such a process is the culture of inquiry and collaboration between teachers. Moreover, a data-driven system might also use perception and other types of data, including performance data from external assessments such as national tests (Bernhardt, 2004). Formative assessment is a unique process that focuses upon assessment embedded in instruction and designed to provide feedback. Much of the process can be conducted without the production of traditional data in the form of student scores and the like. Traditional continuous assessment practice, however, depends upon data from internal summative assessments and formative assessment practice and makes strong use of data-driven instruction to drive achievement. Data-driven instruction is part of a wider data-driven decision making system that may operate at several levels in an education system.

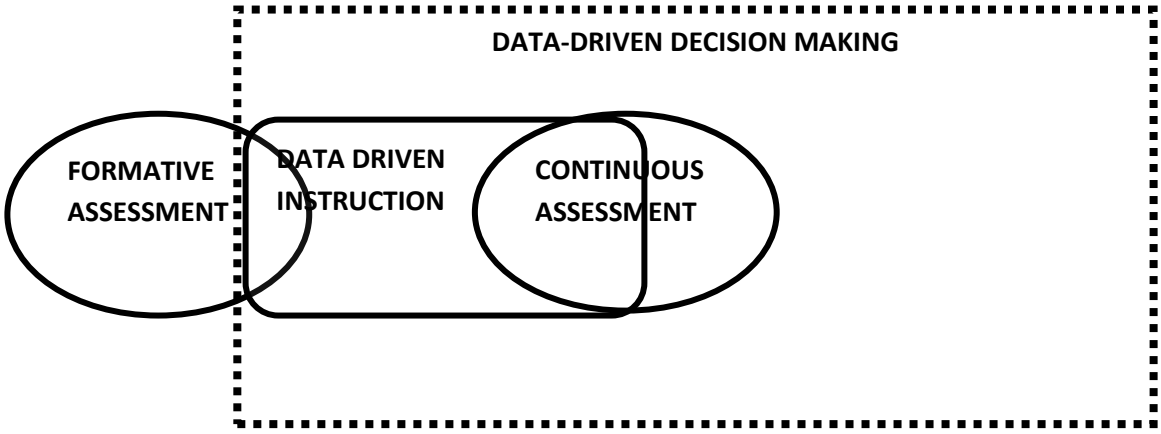


Figure 9: The relationship between assessments and data use

Campbell and Levin (2009) reminded us that since assessment for learning or formative assessment has a different purpose compared with accountability testing (which data driven instruction might also use), there is a need to balance the tensions between the two opposing systems. However, they believe that a systemic approach to both formative assessment and data-driven instruction leads to higher overall professional accountability. Instead, they proposed:

> Assessment for learning is an important strategy for connecting instructional strategies and classroom practices to the individual needs, progress and learning outcomes of students. To be fully effective, however, assessment for learning needs to be conceived of not only as a classroom and school level strategy, but just as importantly as a systemic strategy in which the schools, districts, and state are working together towards shared goals for student learning outcomes. This requires attention to the range of data required to inform improvement strategies, plus other student, school, and contextual data to provide a fuller picture of performance and outcomes at all levels of the system (p. 62).

Table 5 illustrates these ideas and compares assessment for learning (formative assessment) with a systemic reform that includes both assessment for learning and data-driven processes.

Table 5: Pure formative assessment compared with formative assessment & DDDM

Formative Assessment Practice	Systemic Formative Assessment combined with Data-Driven Decision-Making
Assessment for learning should be part of effective planning of teaching and learning.	Assessment of and for learning, plus relevant contextual, student and school data, is part of effective strategic and operational planning for state/province, district, school and classroom actions for improved teaching and learning.
Assessment for learning should focus on how schools learn.	Assessment for learning focuses on how systems, schools and students improve their progress, achievements, and outcomes.
Assessment for learning should be recognised as central to classroom practice.	Data-informed decision making, including assessment for learning, is recognised as central to educational practices at all levels of the system—province/state, district, school, and classroom.
Assessment for learning should be regarded as a key professional skill for teachers.	Skill and capacity to use, understand, and apply data to inform improved actions and outcomes are regarded as key professional skills for all educators, including teachers, principals, district staff, and state/provincial officials.
Assessment for learning should be sensitive and constructive because any assessment has an emotional impact.	Careful consideration of which data to use, by whom and for what purpose is required to ensure the sensitive and constructive use of data, including assessment, to support improvement for all not to rank or judge unfairly.
Assessment for learning should take account of the importance of learner motivation.	The use of data, including assessment for learning, to generate motivation to improve includes an emphasis on developing respectful partnerships to engage educators at all levels of the system to work together.
Assessment for learning should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed.	Shared goals for improvement, and indicators of success towards these goals, are developed and understood to generate a common commitment to improvement targets and learning outcomes.
Learners should receive constructive guidance about how to improve.	Building capacity for improvement involves providing feedback, strategies, resources, and supports to enhance both professional learning and student learning.
Assessment for learning develops learners' capacity for self-assessment so that they can become reflective and self-managing.	Data-informed approaches involve combining assessment of and for learning to balance external and internal accountability, while building the professional capacity for educators to implement self-evaluation, improvement planning, and monitoring strategies.
Assessment for learning should recognise the full range of achievements of all learners.	Assessment for learning, and the use of related data, recognises the full range of achievements of all learners and schools, including not only achievement results but also progress over time and equity of outcomes for closing gaps in performance while raising the bar overall.

Continuous Assessment in Context

To capture continuous assessment practice in different contexts, several studies were reviewed from both developed and developing countries. Table 6 summarizes 20 main studies that guided the review.

Table 6: Critical studies in continuous assessment practice worldwide

Critical Publications from Developing Countries	Critical Publications from Developed Countries
1. Chewang, K. (1999). *Continuous assessment in Bhutan: Science teachers' perspectives*. Unpublished M.Ed dissertation, University of New Brunswick.	1. Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Policy, Principles &Practice, 11*(3), 305-322.
2. Examinations Council of Zambia (2003). *Learner assessment for improved educational quality: An exchange of current ideas and best practices*: Sub-regional conference on assessment held on 30 June to 2 July, 2003 at Zambezi Sun Hotel Livingstone, Zambia	2. McManus, S. (2008). *Attributes of effective formative assessment*. Washington, DC: Council of Chief State School Officers.
3. Adebowale, O. F. & Alao, K. A. (2008). Continuous assessment policy implementation in selected local government areas of Ondo state (Nigeria): Implications for a successful implementation of the UBE program. *KEDI Journal of Educational Policy*, 5(1), 3-18.	3. Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability 21*:5–31
4. Israel, H. F. (2000). *The implementation and effects of continuous assessment in the English classrooms within the changing milieu of Education in South Africa*. Unpublished Ed.D dissertation. University of Baylor, Waco, Texas.	4. Tuttle, H. G. (2009). *Formative assessment: Responding to your students*. Larchmont, NY: Eye on Education, Inc.
5. Shandomo, H. (2008). *Continuous assessment in Swaziland: The predictable fate of western innovation in Africa*. Saarbrücken: VDM Verlag Dr. Müller Aktiengesellschaft & Co. KG	5. Stiggins, R. (2008). *An assessment manifesto: A call for the development of balanced assessment systems*. ETS Assessment Training Institute. Retrieved July 3, 2008 from www.assessmentinst.com/forms/AssessManifesto-08.pdf
6. Mulenga, M. B. and Kapambwe, W. M. (2008). *The implementation of school based continuous assessment (CA) in Zambia.* Paper presented at the 2008 IAEA Conference.	6. Heritage, M. (October 2, 2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89* (2), 140-145.
7. Kapambwe, W. M. (2010). The implementation of school based continuous assessment (CA) in Zambia. *Educational Research and Reviews, 5*(3), 99-107. Available online at http://www.academicjournals.org/ERR	7. Pryor, J. & Crossouard, B. (2005) *A sociocultural theorization of formative assessment*. Paper presented at Sociocultural Theory in Educational Research and Practice Conference, University of Manchester, September 8-9. Retrieved from http://orgs.man.ac.uk/projects/include/experiment/pryor_crossouard.pdf
8. Susuwele-Banda, W. J. (2005). *Classroom assessment in Malawi: Teachers; perceptions and practices in Malawi*. Unpublished Ph.D. Dissertation, Virginia Polytechnic Institue.	8. Chappuis, J. (2009). *Seven strategies of assessment* for *learning.* Portland, OR: ETS Assessment Training Institute.
9. Hayford, S. K. (2007). *Continuous assessment and lower attaining pupils in primary and junior secondary schools in Ghana*. Unpublished Ph.D Thesis, School of Education, University of Birmingham, UK.	9. Preuss, P.G. (2007). *Data-driven Decision Making and Dynamic Planning: A School Leader's Guide*. Larchmont, NY: Eye on Education, Inc.
10. Vandeyar, S., & Killen, R. (2007). Educators' conceptions and practice of classroom assessment in post-apartheid South Africa. *South African Journal of Education, 27*(1), 101-115.	10. Kowalski, T. J. and Lasley, T J., (2008). *Handbook of data-based decision making in education.* New York, NY: Routledge.

Continuous Assessment in Developing Countries

Continuous school-based assessment programmes have been implemented in several developing countries in Asia and Africa at both the primary and secondary school levels. Several research and evaluation studies have been conducted on these assessment approaches and are located in several sources. The documents reviewed ranged from theses and dissertations (ProQuest search) to conference proceedings and peer reviewed papers. Several Government documents were also reviewed. Wherever possible, direct contact was also made with researchers or Government officials. Although studies from the developed world often do not make strong use of assessment theory, increasingly, they are focused on addressing the more problematic issues related to implementation. Additionally, the evaluations often focus

on issues such as low assessment literacy among teachers and inadequate resources in schools, which are relevant to the Caribbean context.

In Africa, several countries have employed continuous assessment despite sharp differences in political philosophy. Nevertheless, political or educational ideologies have often been the driving force behind reform. Continuous assessment has often been applied as a "cure-all salve" to address multiple issues in the education system. The focus in this section is on several countries in Africa illustrated in Figure 10. However, Bhutan will also be considered in the context of a developing country within Asia.

Figure 10: Countries in Africa where CA systems reviewed for this document

Tanzania

There is a long history of continuous assessment in Africa, which is intertwined with social, economic, and political reforms. Tanzania has always had some form of school-based assessment because its system is based on the German model. In the early 1970s, continuous

assessment re-emerged as the answer to tensions created by English Secondary School external examinations (NECTA, 2003). Julius Nyerere had analyzed the basic features of the Tanzanian education system in 1967 and noted fundamental weaknesses in the formal system of education with its strong examination orientation.

> Government and Party themselves tend to judge people according to whether they have 'passed school certificate', 'have a degree', etc. If a man has these qualifications we assume he can fill a post; we do not wait to findout about his attitudes, his character, or any other ability except the ability to pass examinations. If a man does not have these qualifications we assume he cannot do a job; we ignore his knowledge, and experience. For example, I recently visited a very good tobacco-producing peasant. But if I tried to take him into Government as a Tobacco Extension Officer, I would run up against the system because he has no formal education. Everything we do stresses book learning, and underestimates the value to our society of traditional knowledge and the wisdom which is often acquired by intelligent men and women as they experience life, even without their being able to read at all (Nyerere, 1967).

Reorienting the curriculum for a socialist bent would require de-emphasizing these formal examinations that were external to the schools and the country (Kassam, 1994). Thus, Nyerere noted:

> Yet it is easy to say that our primary and secondary schools must prepare young people for the realities and needs of Tanzania; to do it requires a radical change, not only in the education system but also in many existing community attitudes. In particular, it requires that examinations should be down-graded in Government and public esteem. We have to recognize that although they have certain advantages—for example, in reducing the dangers of nepotism and tribalism in a selection process— they also have severe disadvantages too. As a general rule they assess a person's ability to learn facts and present them on demand within a time period. They do not always succeed in assessing a power to reason, and they certainly do not assess character or willingness to serve.

The Tanganyika African National Union (TANU) Musoma Resolution which focused on education for self-reliance was enacted in 1974. Central to this policy was the use of continuous assessment as captured by the following directive:

> The excessive emphasis now placed on written examinations must be reduced, and that the students' progress in the classroom plus his performance of other functions and the work which he will do as part of his education must all be continually assessed and the combined result is what should constitute his success or failure (NECTA, 2003).

At the secondary level, the continuous assessment package consisted of two components: the academic component comprising the continuous assessment and the character assessment and attitude towards work component. The 50% weighting of the continuous assessment

component was derived from exercises such as homework, class tests, and quizzes for 20%, terminal tests for 25% and projects in forms 3 to 5.

The revision of the CA programme took place in 1989. Challenges to implementation included the absence of articulate and comprehensive guidelines to help teachers, cumbersomeness of dealing with large numbers of exercises, academic cheating by teachers; victimization in character assessment, and the demands of the project work. Although CA is designed to empower the learner and the social system, it remains true that:

> The main teaching and learning approaches used in secondary schools in Tanzania is non – participatory with little attention to participatory methods. This is due to lack of teaching-learning materials and facilities. The system enables only students to participate in answering and asking questions during the lesson (Kalomba & Mpaju, 7-8).

In response to these shortcomings, NECTA published new guidelines, exercises were dropped, projects were reduced in number, and character assessment and attitude were no longer included for certification. The continuous assessment system in the primary school is not as elaborate and plans to further strengthen this area does not include combining scores from internal assessments with the external summative assessment scores of the Primary School Leaving Examination at Standard 7.

Nigeria

Early continuous assessment systems were also developed in Nigeria in the 1970s (Ajidagba, 2004). With the Government takeover of schools in the 1970s, nationally set standardized measures and public examinations such as the First School Leaving and Common Entrance Examinations were used for selection and certification. Dissatisfied with the high failure rates, however, continuous and school based assessments were introduced with the 6-3-3-4 system (Three years of Junior and Senior High School) in 1982. Thus, as with many of the other systems, continuous assessment appears to be a major education reform with ideological tenets rooted in improving both teaching-learning and student achievement (Ebelechukwu, 2007).

Although the CA programme was well structured, the reform in assessment practice has had limited impact. Ebelechukwu (2007) noted:

> Unfortunately, continuous assessment has been subjected to a great deal of abuse and misinterpretations by teachers because most of them appear not to understand the rational for continuous assessment in the school system. They have misinterpreted the policy to mean administration of monthly and bi-weekly tests in the cognitive domain. [These] suffer from various vices including lack of validity, and reliability. Studies have also shown that some teachers inflate pupils' scores, thereby making nonsense and negative the predictive comparability between such scores and those of central examination bodies. Consequently, disparity in assessment has been observed from school to school and from rural to urban locations (pp. 2-3).

Teacher misunderstanding and lack of support seem to be a common theme in CA implementation in Africa and suggests that the scores may lack validity and reliability.

Malawi

In Malawi, continuous assessment is a primary component of the Improving Educational Quality project (IEQ). Traditionally, progression in the Malawi primary school system was governed by annual examinations, with high failure rates pointing towards low internal efficiency. The goals of the CA pilot intervention in Ntcheu were listed as to:

1) Find out whether CA can stimulate the improvement of teacher development and pupil performance in the classroom.
2) Make teachers aware that CA is a system of providing feedback to teachers themselves, learners, parents, community and the Ministry officials.
3) Train the teachers in CA procedures in order to improve their own pedagogy skills which can eventually be implemented at a large scale (Kamangira, 2003).

The attempt to improve student performance and the practice of teachers has continued in later projects such as the Primary Curriculum and Assessment Reform of 2007 (Mchazime, 2003).

Susuwele-Banda (2005) captured some of the difficulty in implementing CA in the context of Malawi. The major challenges relate to the fact that in the best continuous assessment systems, there is a need for the assessment to be integrated with instruction and used to provide feedback. This will prove impossible if teachers see assessment simply as testing. He concluded:

> The data suggest that teachers perceive classroom assessment as tests that teachers give to their students at specified time intervals. What teachers said about their teaching was not reflected during their teaching. Since teachers perceived classroom assessment as tests, they showed limited ability to use different methods and tools to assess their students while teaching. The teachers' perceptions of classroom assessment have influence on their classroom assessment practices. Five of the six teachers perceived assessment as testing, and classroom assessment practices were not clearly embedded in their teaching.

The gap between intention and practice is a notable one, impinging upon the formative use of assessments.

Zambia

In Zambia, although mandated as early as 1977, continuous assessment practice was not effectively applied in the schools. The 1996 policy paper *Educating our Future* assessed the situation and pointed to the way forward as described:

> Although the 1977 Educational Reforms explicitly provided for the use of continuous assessment in determining borderline cases or difficult cases resulting

> from personal handicaps or circumstances, this provision does not seem to have been applied at the level of the Grade 7 composite examination. In future, however, more scope will be provided for the use of school-based assessment as part of the overall evaluation of pupils and for its inclusion as a component of certification and selection procedures. However, much preparatory work has to be undertaken before this can be done. Crucial elements are the readiness of teachers, the education of the public, and the ensuring of transparency. But what can be done at once is to have each school develop a comprehensive programme of school-based pupil assessment and feedback as an integral part of its teaching and learning processes. Each school, therefore, will be required to have a clear schedule of performance-monitoring activities that check pupils' progress. Prominent among these will be homework given to pupils on a regular basis, thoroughly marked, and quickly returned. Moreover, as noted already, the Ministry will develop procedures that will enable teachers to standardize their assessment methods and will develop standardized tests for use as an integral part of school-based assessment (Zambia Ministry of Education, p. 42).

As in Malawi, school-based continuous assessment has come to be regarded as an all-encompassing intervention through which, teachers will attend more to individual learners' needs, learning achievement will improve, and final assessment marks will more accurately reflect pupils' skills and competencies. Zambia envisioned using school based continuous assessment as a tool in the primary school for promotion to the secondary school. Unlike the case in Trinidad and Tobago, it should be considered, however, that only small numbers have the opportunity for progressing further at this fist selection point.

The current system has a dual focus of improving instruction and contributing to high stakes selection and certification at Grade 7 (Swallow, Nielson, & Chakufyali, 2009). Several pre-pilot and pilot activities lead to the 2007 system of combining school based marks and examination scores at Grades 5 to 7. The system operated in the following manner:

1) The annual average mark for each subject is calculated for Grades 5-7 based on a complicated scheme illustrated in Figure 9.
2) The cumulative average mark is calculated for 3 years
3) The CA marks and examinations marks are combined based on the weighting of 50% each. This is illustrated in Figure 10.

Kapambwe (2010; personal communication, August 29th 2010) has recently published an evaluation of the pilot and early implementation schemes (2005-2009). The implementation includes the development of materials and training for teachers. The evidence suggests that there was some improvement in the performance of students in the pilot tests although such an increase in scores might well be due to an inflation of grades, common in school-based continuous assessment schemes. However, some of the unique problems in implementation included large class sizes (of over 100), high teacher-student ratios, and high pupil absenteeism. More common issues include the reluctance of teachers to engage in remediation faced with "syllabus" pressures, the absence of resources, teacher collaboration, and monitoring and evaluation. In the face of these challenges, it seems foolhardy to put confidence in the final scores derived from combining CA with the high stakes test scores.

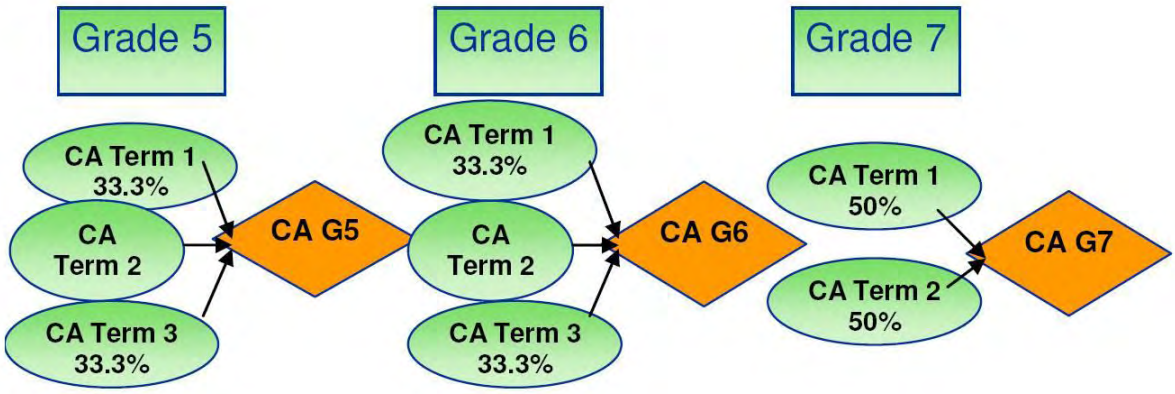


Figure 11: Continuous Assessment Grade Structure in Zambia

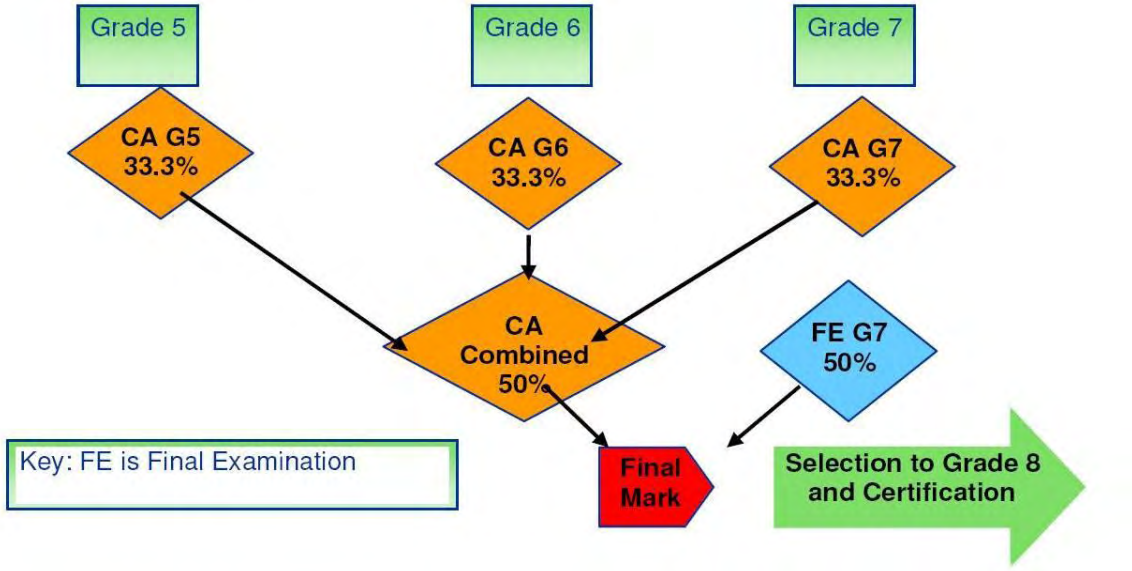


Figure 12: Combination of Continuous Assessment Scores with Scores from final examination at Grade 7 in Zambia

Namibia

In Namibia, several education policy documents[1] in the 1990s focused on using continuous assessment as a tool in education reform. The idea of CA was to develop a reliable holistic picture of each individual learner's progress and level of achievement in relation to the basic and life skills competencies. As The CA is used for promotion though the early grades, but in 2000 there was also a Grade 7 external assessment used for monitoring standards. The CA manuals differentiate between less structured and structured CA assessments, but it is not clear how these are to be used in generating an overall judgement by teachers (or whether they are to be used at all). Indeed one of the features of the documentation is its lack of clarity on format and purpose.

[1] Including *Toward Education for All* (MEC 1993) and the *Broad Curriculum* (MBESC 1996).

An analysis of one of the Continuous Assessment Manuals for Grade 4 shows a variety of discrete and extended response items, some of which are not truly authentic as illustrated in Figure 13. In theory, the system strongly encouraged the use of performance-based assessment and assigned letter grades, which that were supposedly criterion-referenced. These are illustrated in Box 6. Several issues have been reported in implementation, including the lack of clarity already mentioned and the tendency of teachers to focus on testing rather than assessment.

----------Box 6: Criterion Referenced System in Namibia CA----------
Grade A – Achieved Basic Competencies exceptionally well.
The learner is outstanding in the class in all main areas of competency.
Grade B – Achieved Basic Competencies very well.
The learner is above average in the class, and is more proficient than average in several areas, e.g. showing quicker mastery of some competencies, or being able to apply competencies to unknown situations or contexts, or showing new insight.
Grade C – Achieved Basic Competencies.
The learner has mastered the competencies satisfactorily in known situations and contexts. The large majority of learners should achieve this level.
Grade D – Partly achieved Basic Competencies.
The learner may not have achieved all the competencies, or may sometimes need help, but has sufficient mastery to go on to the next grade.
Grade E – Not achieved the majority of Basic Competencies.
The learner has not been able to reach a minimum level of competency for the year grade, even with extensive help from the teacher and is in need of compensatory teaching.

Swaziland

Swaziland developed its system of continuous assessment based on the principle of mastery learning. Again, the system was developed in response to persistent and notable examination failure. Since independence in 1968, Swaziland's intention has been to provide basic education for all; however, this has been foiled by high dropout rates in response to repeated failure. In response, therefore, to the need for quality improvements, in 1995, the National Education Review Commission (NERCOM), which was commissioned to examine the quality, efficiency and effectiveness of the formal school education system, recommended the introduction of Continuous Assessment (CA) in all primary schools, and for pupils in each class. Records to be kept included a learner achievement record, a report to parents, and a record of student progress. A certificate was to be awarded at the end of the 7th year.

As in the other cases in Africa, the CA programme has had multiple goals and multiple components (Mkhonta, 2003). The several evaluations of the Swaziland continuous assessment programme are instructive because it indicates that even though an innovation might be adopted system-wide, several years down the line, the system can become mechanical, routine and perfunctory as note by Mkhonta:

> However, ten years down the line since the introduction of the CA program, the goals and objectives of the program have not been fully realized. Teachers have and still do experience numerous problems and challenges in their efforts to

> implement the CA program. They have lost the enthusiasm and zeal with which they welcomed the program and most of them have given up on the ideals of the CA program (p. 38).

Nsibande (2007) found that teachers were mere technicians in the process and focused more upon procedures rather than on the essence of the process. Although materials were provided, teachers made little time to reflect or adapt the materials to the contexts. Consequently, teachers did not change in response to the imposition of CA. These reactions were caused by a lack of training and implicit beliefs, but also by the pressures and demands of teaching in difficult conditions. As noted in the following excerpt from Nsibande's study, teachers had little time to give one to one formative feedback.

> **MN: Their problem was that they could not respond to questions I gave them, then I gave them the same questions to work on and that is what CA requires.**
> R: How was doing this dealing with their inability to give you correct answers in the first instance?
> **MN: But there isn't enough time for me to look at each learner's problem so I concentrate on general issues. We were told to give learners the same test until they master it and that is what I am doing.**
> R: Don't you think their failure to respond to questions could be an indication of other challenges they have?
> **MN: Maybe [demonstrates a bit of irritation]**
> R: Now tell me, how did you take such a possibility into account when doing remedial work?
> **MN: I see what you are pointing at but if that is the case what should I have done?**
> R: What do you think you should have done?
> **MN: Maybe first find out why a wrong answer was given and then work on the remedial activities in another way but I am not sure how (p. 188)**

Such reflection and adaptation requires not only professional development but also the support of collegial groups and instructional coaches, which are missing in the continuous assessment reform of African countries.

English Second Language **Grade 4**

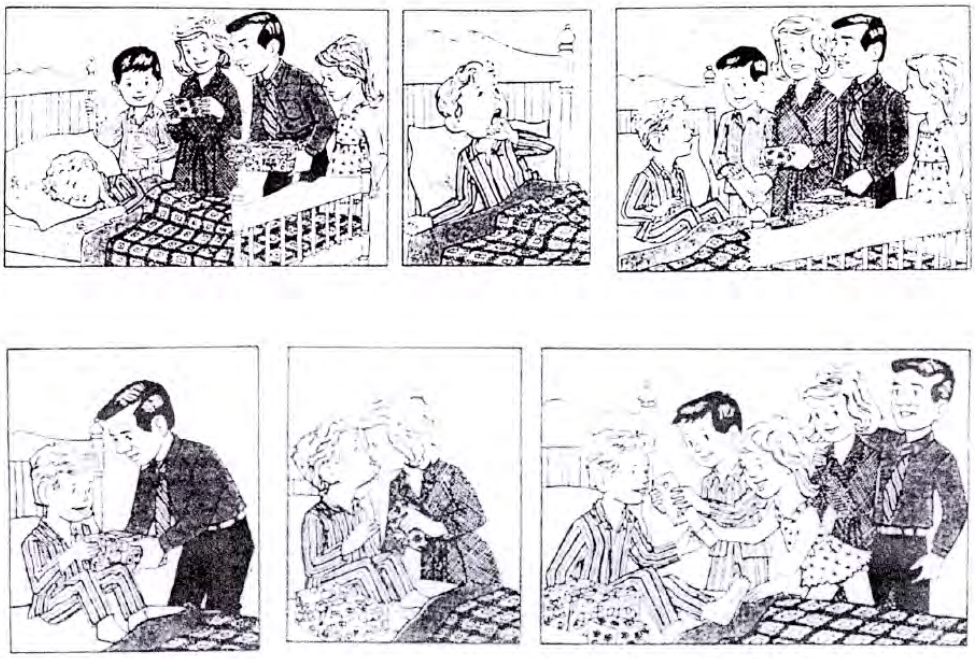
Continuous Assessment More Structured **Term 2** **Total: 20**

Assessment 3

Read the story about John's tenth birthday and answer the questions that follow.

MY TENTH BIRTHDAY

"Happy birthday, John!" Mother says. Slowly I open my eyes and look at the people standing around my bed. There is Father, big and strong. He has a smile on his friendly face. Next to him is my mother. She is a short, kind woman. In front of them I see Jack, my brother, who is twelve years old and Mary, my sister. She is fourteen and we all like her. "This is your present from Grandfather and Grandmother," says Father. "And this is our present!" Jack and Mary shout. I know that this is going to be a lovely day.



1.1 Whose birthday is it? (1)

1.2 Where is he when he wakes up?

_____ (1)

1.3 Complete: Father is stronger than _____.

Mary is older than_____. (2)

1.4 Which present does Father give him?

_____ (1)

CA Manual Grade 4, NIED, 2005 28

Figure 13: Sample items taken from the Namibia CA Manual

Ghana

From 2004, there has been substantial external funding for the development of the Basic Education Comprehensive Assessment System (BECAS) in Ghana. BECAS is a comprehensive national education system. BECAS focused upon focused upon supporting the development of tests that relate to grades one through six in the basic education system. The emphasis was on the educational relevance and appropriateness of the tests based on a comprehensive framework founded upon the essential understanding that underlies the curriculum. BECAS includes three separate assessments:

1) The National Education Assessment (NEA), which is a curriculum-based, competency assessment program measuring the entire curriculum with samples of school performances across the educational regions of Ghana. The system provides national indicators for grade three and grade six on educational achievement in English, mathematics, and two Ghanaian languages for a subset of regions. Although the NEA is designed for policy-level decisions, it makes use of an absolute arbitrary cut-off score for minimum competency (35%) and proficiency (55%).
2) The School Education Assessment (SEA) is a minimum-competency-based test that reflects the essential elements of the curriculum considered basic to the grade level performance expected for student continuation. SEA is given at grades two, four, and six and is now administered by all schools every two years.
3) The Continuous Assessment (CA) supplements the large-scale testing programmes with CA procedures and measures developed for grades one and three. The purpose of this systematic assessment is to indicate foundational problems in the classroom, where adjustments by the teacher can be accommodated during the process of instruction. The CA has been redesigned with a reduced testing load and a focus on essential knowledge and skills to provide formative information for classroom diagnostics to better monitor individual pupil performance

One of the more useful outputs of BECAS is the study of opportunities to learn (OTL). They defined OTL as the provision of adequate and timely instruction of specific content and skills prior to taking a test as measured by (1) time spent in reviewing, practicing, or applying a particular concept or by (2) the amount and depth of content covered with particular groups of students. Thus, the term focuses upon equal conditions for learning and the absence of barriers for learning. They understood that OTL indices are country specific and linked to performance in international assessments.

The OTL study, which focused on Mathematics and Language, found grave inequalities between schools, as well as districts, in (1) availability and adequacy of instructional materials like textbooks (2) instructional practices and management of instructional time and (3) teacher preparedness to implement the content standards. The committee correctly reasoned that OTL is a precursor to use of the NEA for judging schools and districts. Figure 14 shows the finding for percentage of the syllabus covered in schools. As shown, in more than one third of the schools, syllabus coverage was only 50%.

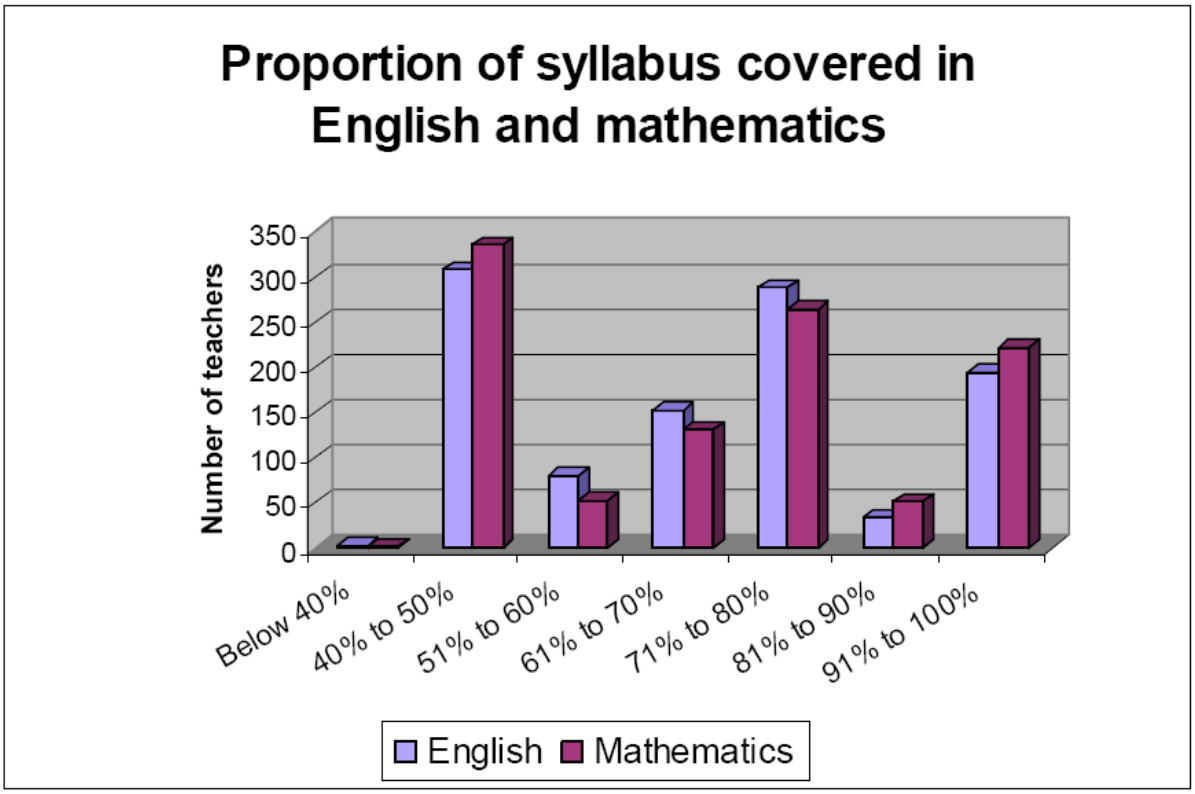


Figure 14: Proportion of Syllabus covered in Language and Mathematics as measured in Ghana OTL study

Hayford (2007), using a mixed method approach, found that large classes, national policy, and the lack of training and support were the primary barriers to implementing continuous assessment. Hayford's study of both the primary and lower secondary school continuous assessment programmes found that lower achieving students were alienated by failure in teacher made tests and they perceived a lack of support from parents and teachers. The data presented in the study suggested that part of the problem related to the lack of feedback from teachers.

South Africa

In South Africa, assessment has been an integral part of two education reforms: the adoption of continuous assessment in the 1990s, and the implementation of the 2004 Outcome-Based Curriculum reform. At the secondary school level, Israel (2000) found that the reliability of classroom assessment in the CA process was low with scores poorly correlated with external assessments. At the same time, teachers were overburdened by the workload and other demands and lacked needed professional development and support.

Vandeyar and Killen (2007) traced the changes in education policy aligned with the political, social and economic changes in Post-Apartheid South Africa, but noted the reluctance among teachers to embrace assessment change, which seem tied in to understandings of the curriculum and pedagogy. The authors argue that such conceptions develop from the context that teachers work in, but once developed they are quite resistant to change. Much of the resistance to the South African assessment reforms seems to come from the traditional beliefs

of teachers. As Pryor and Lubiski (2002) noted with regards to the continuous assessment reform in South Africa:

> Our data suggest that, even where there is a willingness to embrace new ideas about assessment, the enterprise is hampered by lack of training and, above all, by the fact that teachers hold tacit values which are in tension with the values underpinning the new assessment order. These factors, together with the complexity of the curriculum change, create serious problems of manageability and interpretation. Thus, far from empowering teachers and learners, the new requirements may have a tendency to make them feel even more isolated from control of their situation (p. 475).

New proposals for the assessment system in South Africa also make strong use of continuous assessment and combines scores with external tests, with the weighting of the external tests increasing in the higher grades as noted by the line minister (Motshekga, 2010):

> Secondly, Council agreed to regular, externally-set assessments at grades 3, 6 and 9 in literacy (in home language and first additional language) and numeracy/mathematics. It agreed on a weighting of continuous assessment and end of year examinations as follows: Grades R-3: 100% continuous assessment; Grades 4-6: 75% continuous assessment: 25% end of year exam; Grades 7-9: 40%: continuous assessment: 60% end of year exam and Grades 10-12: 25% continuous assessment: 75% end of year exam.

Bhutan

Bhutan is situated near to India, Bangladesh, China and Nepal. As with the African countries, external examinations have dominated the system at both primary and secondary levels (Chewang, 1999). There were two common external examinations in grades 5 and 8. Between 1972 and 1982, the examinations were administered from India. "Bhutinization" of the examinations have been accompanied by changes in both purpose and format. By 1994 the weighting of internal assessments for the Class VI assessment had increased from 10 to 20 and then to 50% in an attempt to reduce the impact of the final Primary School Certificate Examination (PSCE). This score is derived from projects, homework, classwork (30%) and two paper and pencil tests (10% each), called a trial and mock tests as illustrated. Ministry of Education evaluations suggest that some teachers want continuous assessment to be reduced. In his study of continuous assessment in the area of science at the primary level, Chewang (1999) found that the innovation was not clear to teachers. He listed six inhibiting factors, large class sizes, limited resources, syllabus coverage, lack of professional support, and the time factor.

Figure 15: Location of Bhutan

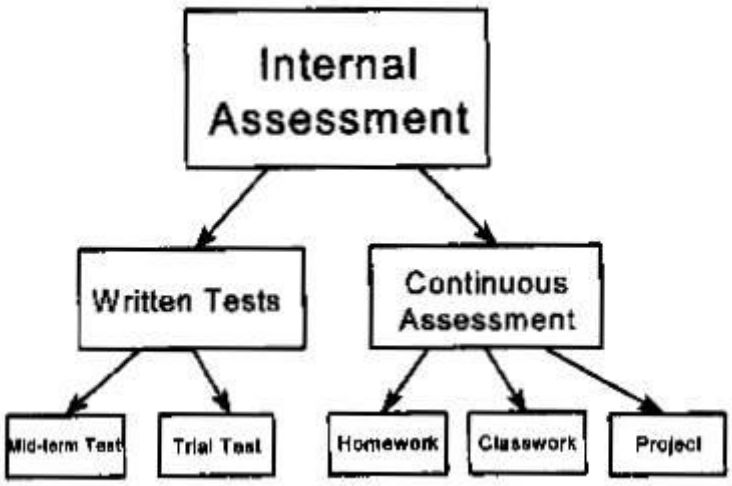


Figure 16: Internal Assessments in Bhutan at Class Level VI

Notably, as is usual in developing countries, professional development was promised, but the follow-up was not readily evident. Because the project was field tested, the Ministry of

Education did not monitor assessment literacy needs and some principals felt that the project failed because teachers did not know how to conduct this kind of formative assessment. Additionally, despite the reduction in the weight of the external assessment, the intention of teachers is still to conduct a race to finish the syllabus. Wangchuck (2000) had similar findings to Chewang (1999) and identified lack of training in the preservice and inservice sectors along with the burdens of assessment as the greatest barriers.

Summary

The analysis of these different country contexts provides useful insight into planning and evaluating continuous assessment practice in Trinidad and Tobago. Notable are the persistent problems of lack of training and teacher beliefs that are a significant barrier in implementing continuous assessment programmes. The political and ideological context also seemed relevant. In several of the cases reviewed, it appears that continuous assessment was being used as a tool to force instructional improvement in schools. This belief in a form of measurement driven change is also popular in Trinidad and Tobago. It would seem that the more appropriate approach would be to improve the quality of teaching and learning along with organizational elements and then use continuous assessment to further enhance the reform process. No assessment, large-scale or classroom-based, can be a magic bullet to alter the history of inequality and unequal learning opportunities. There also seems little value in giving the teacher power to assess and then recapturing that power by demanding return of marks to some central authority, assuming the capacity of the central authority to deal with the data and the value of the data for the purposes proposed. Teachers can certainly use scores from continuous assessment to inform their own teaching-learning and the Central Authority should stick to what it does best, namely constructing fair, standardized external assessments.

Continuous Formative Assessment in Developed Countries

The ideas of formative assessment and assessment for learning have dominated Western ideas on continuous classroom assessment since 1998. The actual term formative evaluation is derived from the 1967 work of Scriven; however, both this work and the work of Bloom focused on program evaluation (Cizek, 2010). The current understanding of formative assessment in the classroom might be attributed to the members of the King's College team, especially Paul Black and Dylan Wiliam. The impetus for development of the approach was driven by the 1988 changes to the curriculum proposed in the UK. The 1998 article published in Assessment in Education and written by Black and Wiliam summarized an extensive earlier review of 681 English-language articles related to formative assessment, including a number of controlled experiments. Their synthesis of the evidence from the 250 articles for the paper suggested that there were significant gains in student learning in classrooms using formative assessment.

Sadler (1998) wrote an article in the same journal volume, in which he argued that in truly "formative" assessment, (1) a student must come to hold a concept of quality roughly similar to that of the teacher, (2) be able to compare the current level of performance with the standard, and (3) be able to take action to close the gap. Further impetus for the idea of

formative assessment and its role in classroom assessment came from the group of education researchers and other professionals who formed the Policy Task Group on Assessment, under the umbrella of the British Educational Research Association. This policy task group set up the *Assessment Reform Group* (ARG) with funding from the Nuffield Foundation. The Nuffield foundation funded the King's-Medway-Oxfordshire Formative Assessment Project (KMOFAP), which began in January 1999.

The ARG is credited with some of the most significant publications in the field and the use of the term, assessment for learning. A number of researchers including Black and Wiliam have published seminal articles such as Inside the Black Box, which targeted teachers and the general public. In 2002, the group defined assessment for learning as:

> . . . the process of seeking and interpreting evidence for use by learners and their teachers to decide where they are in their learning, where they need to go and how best to get there (ARG, 2002).

This definition was accepted by several British Education agencies including the Department for Education and Skills (DfES), the Qualifications and Curriculum Authority (QCA) and the Office of Standards in Education (OFSTED). The focus of the definition was on students and using evidence from assessment to improve learning. It must be noted, however that the 2002 definition differs from the 2009 definition agreed to in the Position Paper on Assessment for Learning at the Third International Conference on Assessment for Learning. This definition states:

> Assessment for Learning is part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning.

In this definition, AfL is considered everyday practice with various sources of information available, but the goal is the still the same, to enhance learning.

Such has been the impact of formative assessment/assessment for learning that in the fourth Edition of the Handbook of Educational Measurement, the chapter on classroom assessment heavily focuses upon this theme. Its writer, and several other US authors have also explored formative assessment and assessment for learning. In a seminal article, Shepard (2000) had talked about the changes in pedagogical theory and that required changes in the use of assessment. Her vision of classroom assessment aligned with constructivist teaching including the use of innovative assessment formats, protecting classroom assessments from high stakes testing, and creating a learning culture that encourages new types of assessment practice. She highlighted the need for dynamic ongoing assessment, which is situated in the middle of the teaching and learning process and which serves as scaffolding following Vygotsky's idea of a zone of proximal development.

Shepard (2005) expanded on this idea of using formative assessment for scaffolding. She stressed that formative assessment was more than a mechanical data-gathering approach; instead it provides a model for learning in which the assessment step provides the insight necessary for providing the support necessary to move the child through the zone of proximal development. Chauppis and Stiggins (2002) show that in classroom assessment for learning

the student is no passive participant. Instead, in student involved assessment the students take control of their learning by determining the attributes of good performance, use scoring guides, revise work samples, create practice tests and communicate to others how they have grown.

If not explicitly, it appears that the continuous formative assessment process described in these studies builds and elaborates on the continuous assessment model first implemented in African countries. However, approaches in the developed world are much more theory-focused and process-oriented. They pay greater attention to teacher development instead of procedures and methods; although implementation remains a challenge in most countries. Some of these differences are due to greater funding and better insight into the nature of professional development for teachers. Several formal studies of systemic change using AfL are available and projects in four countries were reviewed, Britain, New Zealand, the US, Canada, and Hong Kong. The latter country was chosen because it has an examination-oriented focus much like the Caribbean.

The United Kingdom

In the UK, Shirley Clarke developed a package of formative assessment strategies for use in several English schools, including those in the Gillingham Partnership (Clarke, McCallum, & Lopez-Charles, 2001). The classroom strategies were based on the work of Black and Wiliam (1998b) and the ARG. It includes significant interactive formative assessment components and goal setting. The project was conducted in 2000 to 2001 and consisted of four components as illustrated in Table 7. Implementation was centred upon in-service staff development exercises, with various strategies outlined, discussed and practiced collaboratively in learning teams. Towards the end of each term, observations of lessons and interviews with children and teachers took place in every classroom.

The first and critical component of the formative process as described by Clarke is to share learning intentions so that students are not at the whim of the teacher. The learning intention is what teachers hope children will know, understand, or be able to do by the end of the lesson or set of lessons. The clarification of the learning intention enables the teacher to create a matched task that will fulfill the learning intention. Five steps in using leaning intentions are:

1. Identify what the students will be learning (We are learning to …).
2. Explain the reason for that learning (We are learning this because …).
3. Share (and sometimes negotiate) the learning and the reason with the students.
4. Present the information in language that they can understand.
5. Revisit the learning intention throughout the activity or lesson.

Teachers were also required to tell children the success criteria of the task. Success criteria summarize the key steps or ingredients the students need in order to fulfill the learning intention – the main things to do, include or focus upon.

Table 7: Phases in the Gillingham Formative Assessment Education Project

TIME PERIOD	ACTIVITY/GOAL
Autumn Term 2000	Share learning intentions and success criteria and begin pupil self-evaluation
Spring term 2001	Focus oral and written feedback around learning intentions
Summer term 2001	Introduce writing/social targets with optimum pupil involvement
Autumn term 2001	Analysis and final report writing by Institute of Education staff Ways forward for Gillingham

Significant findings of the project were that most children understood and could explain the success criteria in relation to the learning intention and the majority of children expressed pleasure in knowing the learning intention and the success criteria of lessons. Seventy five percent of teachers said that children understood tasks better and almost all teachers said that sharing learning intentions and success criteria had a positive effect on their teaching. Of the teachers who tried pupil self-assessment, almost all said it had a positive effect on their teaching.

New Zealand

In New Zealand the *Assess to Learn Project* (AtoL) was implemented in 2002. The project had four objectives, namely to:

a. Improve student learning and achievement
b. Shift teachers' knowledge and assessment practice
c. Develop coherence between assessment processes, practices and systems in classrooms and in schools so that they promote better learning
d. Demonstrate a culture of continuous school improvement.

In full implementation, AtoL is a large project delivered to 200 schools across New Zealand by eight service providers. The providers include five colleges of education and three private companies, each with a Director and a team of facilitators. In primary schools, the programme was delivered in the following manner:

1) The provider held an initial meeting with the principal (and possibly a school-based professional development team) to discuss the desired outcomes of the AtoL project.
2) A facilitator (from the provider organization) met with the staff and together they constructed a list of the features of effective quality assessment practice (including the purposeful use of assessment tools).
3) The facilitator and teachers then negotiated an aspect of practice (as discussed) to trial in the classroom. The facilitator and the teachers planned together the strategies they will use for implementation, and the facilitator observes the implementation in the classroom.
4) This was followed with further facilitator-led individual or group meetings to discuss feedback from the classroom observations.

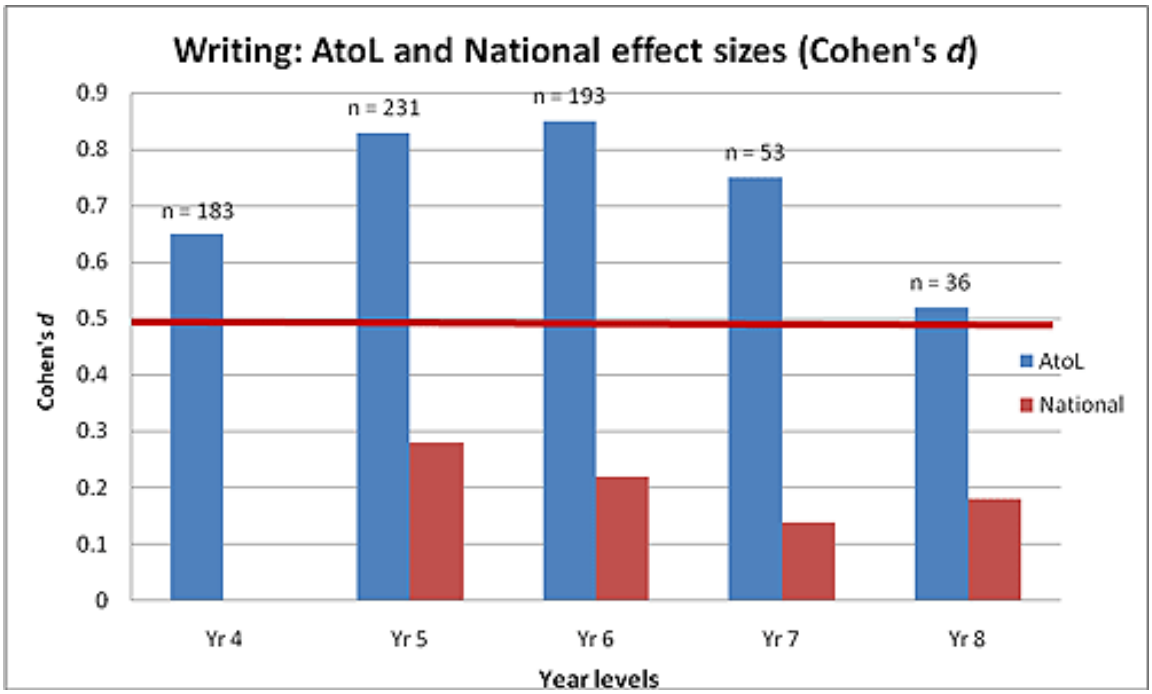
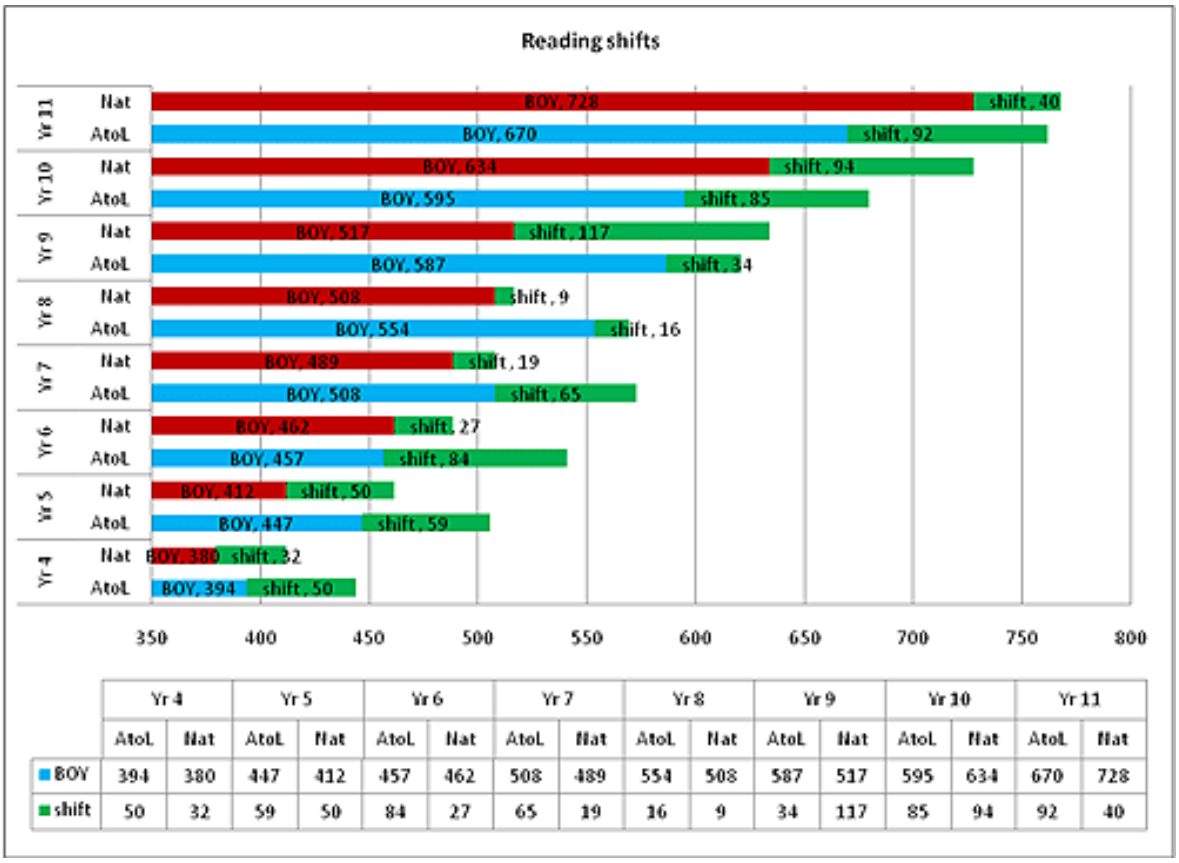


Figure 17: Comparisons of Effect Sizes for Years 4-8 for AtoL and National Schools

The project had an ongoing and summative evaluation. The data suggested impressive gains in student learning and achievement and some improvements in overall school efficiency. For example, Figure 17 shows the effect size for the difference between mean scores for the start and end of the year in Writing for AtoL and non-treatment schools. As shown, the AtoL schools all have effect sizes above 0.5, a medium sized effect, with larger gains in years 5 and 6 at the level of a large size effect. These are very impressive statistics, given that most educational interventions tend to have effect sizes between 0.25 and 0.35. The larger shifts across student years 4 to 11 are more clearly seen in Figure 13, which compare the shifts from the beginning of the year (BOY) for national and AtoL schools.

Thus, the project appears to have met most of its objectives. Perhaps most outstanding is the sustainability of the reform, with a study of 38 schools confirming that 80% continued improvement after the close of the project. Additionally, teachers were clear about their role and involvement. One of the keys to success may have been the flexible professional development employed. The strategies employed include staff meetings, team meetings, classroom observation, one on one support, and professional readings. Young's (2009) study of teachers in primary schools facing the challenge of managing assessment confirmed that professional development was one of the stronger points in the New Zealand assessment reform, with teachers able to cite both assessment texts and professional development experiences as they related their practice of formative assessment. Vercauteren's (2005) study of 7 to 10-year olds confirmed that students understood some aspects of the feedback process.



	Yr 4		Yr 5		Yr 6		Yr 7		Yr 8		Yr 9		Yr 10		Yr 11	
	AtoL	Nat	AtoL	Nat	AtoL	Nat	AtoL	Nat	AtoL	Nat	AtoL	Nat	AtoL	Nat	AtoL	Nat
BOY	394	380	447	412	457	462	508	489	554	508	587	517	595	634	670	728
shift	50	32	59	50	84	27	65	19	16	9	34	117	85	94	92	40

Figure 18: Shifts from Beginning of Year for AtoL and National Schools, Ages 4-11

The United States

STARS (Nebraska)

This Nebraska programme has some formative elements but is also used for accountability purposes. The important characteristic of the system is that it is teacher-led. The state has operated this School Based Teacher-Led Assessment and Reporting system (STARS) for several years under the No Child Left Behind (NCLB) Accountability system (Nebraska Department of Education, 2004). The process required teacher expertise in assessment and integrates the accountability requirements of the NCLB 2001.

The professional development systems included 18 education service units, which provide support materials, hands-on workshops, conferences, communication through satellite broadcasts, state-wide information sharing sessions, training sessions and interactive data bases.The belief was that STARS is high impact but not high stakes. Teacher used locally designed assessments in combination with national tests and a statewide writing assessment to determine the performance of students. To ensure quality in the locally designed assessments a peer review process was used to evaluate teacher assessments according to the following six criteria.

1) Match and measure the standards. Districts must determine that the assessment measures the standards and that students have sufficient opportunity to demonstrate their ability to meet the standard.
2) Provide opportunity for students to have learned the content. Districts must have examined their own local curriculum to determine that the opportunity to meet the

standards exists within the local district's curriculum and that instruction on the standards occurs at an appropriate time in the formative assessment cycle.

3) Be free of bias. Districts must examine the assessment to be sure that any of the items or tasks are free of bias and are not insensitive to any group or circumstance.
4) Be written at the appropriate level. Districts must examine the assessment items or tasks in order to determine that the expectations are appropriate for the assessed grade level.
5) Be reliable and consistently scored. Districts must document that they can have confidence in the results of the assessment, that assessment results have produced an appropriate level of reliability, .70 or higher.
6) Have appropriate mastery levels. Districts must describe the systematic way they have determined mastery levels for the assessment, including both professional judgment and actual student results.

The data is used for analysis and in school wide improvement initiatives. The data use questions are (1) what does the data tell us? (Factual), (2) what might this data mean? (Hypothesis) and (3) what are the implications? (Next steps). Although the authors and supporters of STARS claimed that growth is modest, the system does not meet NCLB requirements and is under pressure to conform.

BEAR Assessment System

The BEAR (Berkeley Evaluation & Assessment Research) assessment system designed by the University of Berkeley at California is also multifunctional but used primarily for formative purposes (Wilson & Sloane, 2000). The system consists of easy-to-use tools for generating solid diagnostic information and feedback, especially in large classes. It is described as a system embedded with instruction and has four major purposes, to:

1) Assess student performance on central concepts and skills in the curriculum.
2) Set standards of student performance.
3) Track student progress over the year on the central concepts.
4) Provide feedback (to themselves, students, administrators, parents, or other audiences) on student progress and on the effectiveness of the instructional materials and the classroom instruction.

Thus, the embedded activities serve three different purposes, feedback, monitoring, and reporting. The embedded assessment activities are inserted at points in the instructional program that represented critical junctures to make sure students adequately prepared for the next segment of the curriculum. The system is built on four principles and four associated building blocks shown below:

Principle 1: Assessment should be based on a developmental perspective of student learning
Building Block 1: Progress Variables
Principle 2: What is taught and what is assessed must be clearly aligned
Building Block 2: Item Design
Principle 3: Teachers are the managers and users of assessment data
Building Block 3: Outcome Space

Principle 4: Classroom assessment must uphold sound standards of validity and reliability
Building Block 4: Measurement Model

Principle 1 is formative in intent by arguing that the primary purpose of assessment is to determine how students are progressing in expertise for the domain of interest, rather than limiting the use of assessment to measuring competence after learning is completed. Principle 2 is also related to formative assessment calling for seamless integration of assessment into the teaching and curriculum. Principle 3 indicates that teachers are the managers of assessment and Principle 4 focuses upon proficiency measurement and the quality of the evidence gathered. The system was first implemented for science content in the middle school grades and the examples of assessment prompts below are from that subject area.

1. Based on everything you have learned about acid-base neutralization, do you think neutralization can be used as a solution to acid-base pollution? Explain your answer thoroughly. Describe both the advantages and disadvantages and other factors that must be considered before reaching a conclusion.

2. You are a public health official who works in the Water Department. Your supervisor has asked you to respond to the public's concern about water chlorination at the next City Council meeting. Prepare a written response explaining the issues raised in the newspaper articles. Be sure to discuss the advantages and disadvantages of chlorinating drinking water in your reponse, and then explain your recommendation about whether the water should be chlorinated.

Figure 19: Assessment Task in BEAR Assessment System

Canada

Canada has a decentralized system based on its provinces; however, all of the provinces do very well in the Progress in International Reading Literacy Study (PIRLS) and the Programme in Student Assessment (PISA) international assessments. The OECD has studied formative assessment practice in Canada, particularly in the provinces of Saskatchewan, Manitoba, Ontario, Quebec, and Newfoundland and Labrador. The education systems differ across these and the other provinces. Several models of formative assessment have been developed since 1989. Some are very similar to that proposed by Clarke in the UK whereas

others are assessment focused. Evaluations have provided some insight into assessment reform, but unlike the UK and New Zealand, there is insufficient study of these processes.

Ontario is one of the better performing districts and in recent years has attempted to develop a comprehensive system of formative assessment. Such assessments compete with the extensive large-scale testing conducted by the Education Quality and Accountability Office and results in a tension. There is no formal requirement to use classroom assessment data for accountability purposes in Ontario. Thus, school improvement plans also contain an emphasis on large-scale assessments.

The Ontario Ministry of Education recently published a document on assessment, evaluation and reporting in Ontario schools. The policy document attempts to redress the balance by refocusing teachers on formative assessment. The document states as policy:

> Teachers will obtain assessment information through a variety of means, which may include formal and informal observations, discussions, learning conversations, questioning, conferences, homework, tasks done in groups, demonstrations, projects, portfolios, developmental continua, performances, peer and self-assessments, self-reflections, essays, and tests.
>
> For Grades 1 to 12, assessment is based on evidence of student achievement of the provincial curriculum expectations. Teachers will ensure that students' demonstration of their achievement is assessed in a balanced manner with respect to the four categories of the achievement chart, and that achievement of particular expectations is considered within the appropriate categories. All specific expectations must be accounted for in instruction and assessment (p. 28).

The document then provides a four-step protocol for conducting formative assessment:

1) plan the assessment concurrently and integrate it seamlessly with instruction;
2) share learning goals and success criteria with students at the outset of learning to ensure that students and teachers have a common and shared understanding of these goals and criteria as learning progresses;
3) gather information about student learning before, during, and at or near the end of a period of instruction, using a variety of assessment strategies and tools;
4) use assessment to inform instruction, guide next steps, and help students monitor their progress towards achieving their learning goals;

One of the more useful aspects of the document is a classification of the three assessments types and clarification of nature and uses. This is illustrated in Table 8.

Volante (2009) studied the three different conceptions of assessment, assessment of learning (summative assessment, assessment for learning (formative assessment) and assessment as learning (formative assessment for the development of student' metacognition). She found that Ontario teachers' classroom practice was weak in assessment as learning. Factors which hindered use were lack of targeted professional development, time for discussion and cooperative learning with peers, density of the provincial curriculum, and student, parent, and teacher perspectives around assessment and evaluation.

Table 8: Purpose and nature of assessment with use of information in Ontario (Ontario Ministry of Education, 2010, p. 31)

Purpose of Classroom Assessment	Nature of Assessment	Use of Information
Assessment for learning "Assessment for learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go, and how best to get there." (Assessment Reform Group, 2002, p. 2)	***Diagnostic assessment:*** • occurs before instruction begins so teachers can determine students' readiness to learn new knowledge and skills, as well as obtain information about their interests and learning preferences.	**The information gathered:** • is used by teachers and students to determine what students already know and can do with respect to the knowledge and skills identified in the overall and specific expectations, so teachers can plan instruction and assessment that are differentiated and personalized and work with students to set appropriate learning goals.
	Formative assessment: • occurs frequently and in an ongoing manner during instruction, while students are still gaining knowledge and practising skills.	**The information gathered:** • is used by teachers to monitor students' progress towards achieving the overall and specific expectations, so that teachers can provide timely and specific descriptive feedback to students, scaffold next steps, and differentiate instruction and assessment in response to student needs.
Assessment* as *learning "Assessment *as* learning focuses on the explicit fostering of students' capacity over time to be their own best assessors, but teachers need to start by presenting and modelling external, structured opportunities for students to assess themselves." (Western and Northern Canadian Protocol, p. 42)	***Formative assessment:*** • occurs frequently and in an ongoing manner during instruction, with support, modelling, and guidance from the teacher.	**The information gathered:** • is used by students to provide feedback to other students (peer assessment), monitor their own progress towards achieving their learning goals (self-assessment), make adjustments in their learning approaches, reflect on their learning, and set individual goals for learning.
Assessment* of *learning "Assessment *of* learning is the assessment that becomes public and results in statements or symbols about how well students are learning. It often contributes to pivotal decisions that will affect students' futures." (Western and Northern Canadian Protocol, p. 55)	***Summative assessment:*** • occurs at or near the end of a period of learning, and may be used to inform further instruction.	**The information gathered:** • is used by the teacher to summarize learning at a given point in time. This summary is used to make judgements about the quality of student learning on the basis of established criteria, to assign a value to represent that quality, and to support the communication of information about achievement to students themselves, parents, teachers, and others.

Ontario represents a community trying to come to grips with the tension between large-scale and formative assessment and the conflict between accountability and assessment for learning. Outstanding is the recent adjustment and strong use of evidence-based policy to further accelerate the high achieving system.

Hong Kong

As measured international assessments, Hong Kong has a very efficient education system. However, the need for reform in assessment practice has been recognized as early as the

1990s. The issue of assessment reform was implied in the Education Commission report of 1990 and the Report on Review of 9-year Compulsory Education in 1997. The reforms gained impetus in the Reform Proposals for the Education System in Hong Kong published in the year 2000 (Yu, 2007).

In 2000, the high stakes selective examination at the end of primary school, the Academic Aptitude Test (AAT), was removed. The idea was to sop memorization and drillings and focus on meaningful learning. The Basic Competency Assessments (BCA) was introduced as a replacement consisting of student assessments and system assessments. The Student Assessment is designed to help teachers better understand the learning needs and problems of students and to provide timely assistance to enhance students' learning effectiveness. The System Assessment provides the Government and school management with information on students' standards at the levels of Primary 3, Primary 6 and Secondary 3 on a territory-wide and school basis.

In the main, the desire to change assessment roles has been fueled by a shift in education for selection to educative function. The Quality Education Fund (QEF) of Hong Kong funded a study of four primary schools. Yu (2007) found that several schools conducted continuous assessment using projects and other innovative formats. However, when it came to reporting, the traditional assessments were strongly favoured. Some schools considered tests as one kind of continuous assessment, often reducing the number of examinations but increased the number of tests. Figure 15 illustrates each of the cases and the types of assessments used for the subject areas, C (Chinese Language), E (English Language), M (Mathematics), and G (General Studies). As shown most schools maintained the traditional assessments despite the reform. On a more positive note, teacher did begin to value formative assessment, student self assessment and feedback. However, they found diversity, lack of knowledge, examination culture, and external examinations as forces which impeded good practice.

Forms of assessment	IPS				PPS				CPS				SBS			
	C[9]	E	M	G	C	E	M	G	C	E	M	G	C	E	M	G
Examination	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	
Test	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆	☆
Dictation	☆	☆			☆	☆			☆	☆			☆	☆		
Project		☆		☆	☆							☆	☆	☆	☆	☆
Portfolio											☆	☆				
Oral task													☆	☆		
Reading task	☆				☆				☆	☆			☆	☆		

Figure 20: Different types of assessments used in the 4 study schools in Hong Kong.

Critical Summary

The literature review on classroom and formative assessment in different parts of the world suggest several similarities and unique differences. Firstly, these reforms, in the form of continuous assessment, curriculum based assessment, formative assessment, assessment for learning, or assessment as learning are all designed to improve student learning and have

often come about as a response to the increasing influence of external assessments. In the developing world, these external assessments were in the form of external public examinations, but increasingly in the developed countries, accountability testing has been the opponent as in the case of Ontario.

Figure 21 illustrates the development of both continuous assessment schemes and the later theory of formative assessment in both the developed and developing world. As shown, much of the growth in ideas took place in the 1990s, with emergence of formative assessment and data driven decision making as separate innovations. Continuous assessment schemes in the developed countries are often predominately summative in function and are even sometimes used in combination with external assessments to make a high stakes decisions as in Zambia and Hong Kong. Some countries have also used both statistical and qualitative moderation before combining scores, as in South Africa and most parts of Asia (Singapore Examinations and Assessment Board, 2006). The summative function of continuous assessment is considered both in the design and implementation of the process in both Asia and Africa. However, programme designers have frequently insisted that continuous assessment is meant to enhance student learning.

The early work on curriculum-based assessment in African countries does suggest that the intention of continuous assessment in the developing world was to influence student learning. Nitko's (1995) framework, widely referenced in the African literature, puts forward a framework that is primarily curriculum-based and criterion-referenced, factors which might enhance the validity of the teachers' assessment for high stakes purposes. Likewise, Harris and Pasinga's 1993 work for the USAID funding agency emphasized the formative aspect of continuous assessment when they reported:

> Curriculum-based assessment (CBA) can do much to minimize, if not eliminate failures and dropouts and thereby enhance the effectiveness of the education process in the schools. This is particularly important at the basic (primary) education level in developing countries where a large percentage of the school-age population leave school after completing the first educational cycle. It is the responsibility of the primary education system to produce functionally literate graduates who are able to contribute positively to the social and economic well-being of the country. CBA is the practice of asking students to perform tasks that have been drawn directly from the curriculum and then using assessment results to adapt instruction to reflect the learners' needs. CBA provides a way of linking curriculum with learning, and in effect, adjusting instruction to fit the students (pp. 1-2)

Perhaps there is some evidence that the functions of high stakes selection and promoting learning might be integrated or even act synergistically in the case of the standards based assessments in the US. However, the weight of the current evidence from elsewhere suggests that in the vast majority of cases, the tension between the two functions might lead to unpredictable negative consequences.

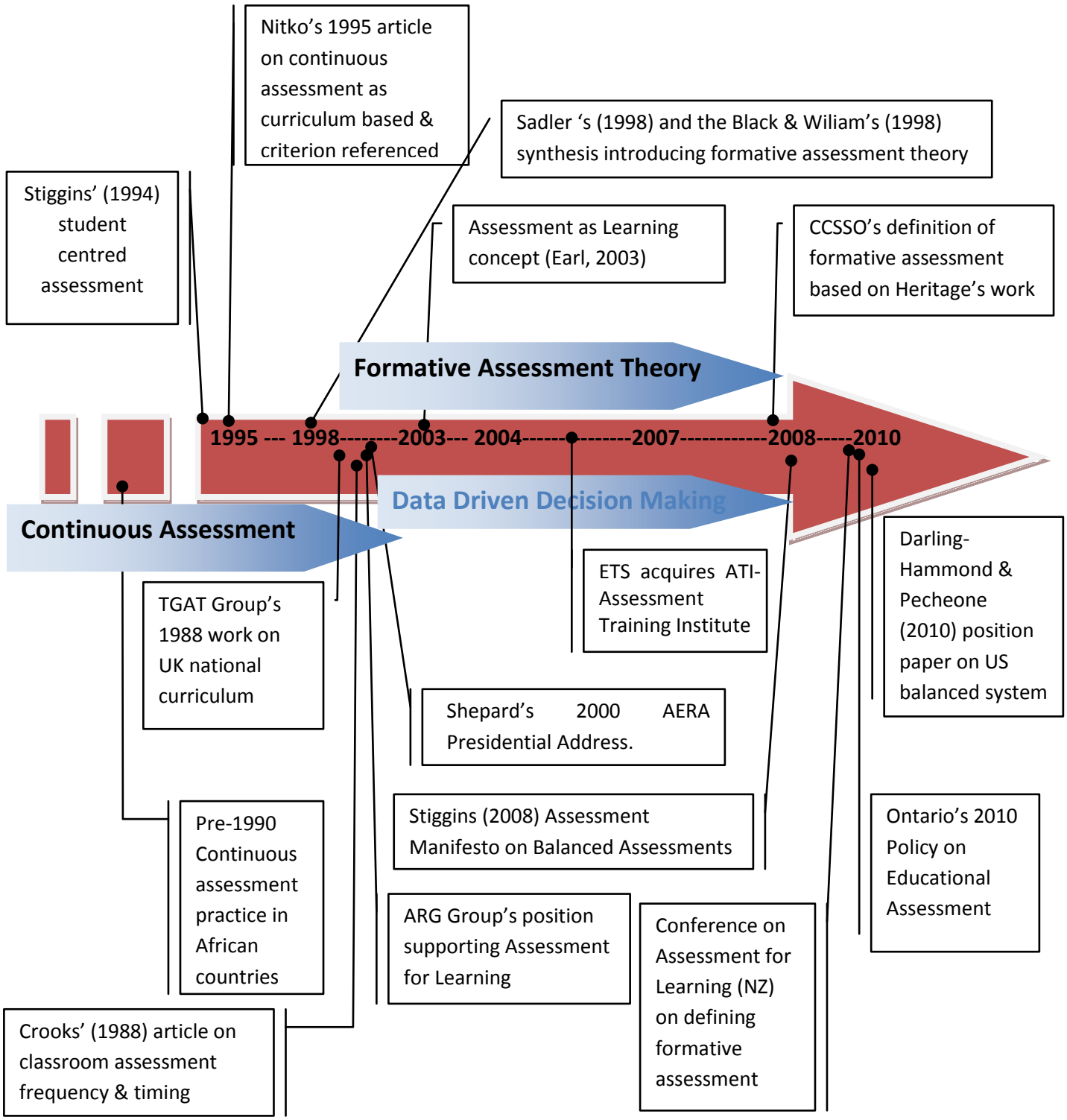


Figure 21: The development of theory on curriculum embedded assessment (formative and continuous assessment)

In several developed countries such as New Zealand and the United Kingdom, innovations centring on formative assessment practice have developed primarily based on the work of Black and Wiliam (1998b) and Sadler (1998). Such reforms position assessment for learning as the predominant function of classroom assessment and in newer nationwide balanced assessment systems, the summative function is held by other assessments, including interim

and benchmark assessments, which are used formatively. The stand alone formative assessment innovations in developed countries are often strongly supported by professional development of collaborative teams. Collaborative teams are also extensively used in data driven decision-making, although the data for analysis might come from multiple sources, including both formative and summative assessments.

Formative assessment theory and policy has grown considerably in the last five years and important concepts have been added, including assessment as learning in Canada and the UK, the worldwide data driven decision-making movement, and the recent call for balanced assessment systems by Stiggins (2008). The latter work is evident in the position paper of Darling Hammond and Pechone (2010) and in the Ontario Ministry of Education's 2010 policy on educational assessment.

Assessment Reform

Models of change

Assessment reform has often been used in the service of general education reform and therefore the nature of education change has to be considered in this review. However, there is also more specific information on strategies for facilitating assessment change including the role of assessment literacy and the influence of principal leadership. Also important to consider in explicating assessment change theory is the impact of the various forms of assessment within a system on each other. As argued already the relationship between internal and external assessment is not by nature synergistic and the resulting tensions may distort the implementation process. Education reform has as its ultimate goal increased student achievement, but this is itself measured by assessments of learning. Therein lies a significant paradox, because the change measured may be meaningless if the assessments lack validity. Assessment for learning may be used to facilitate progress towards the improving student learning, but requires significant changes in the users and the school.

As shown in figure 22, there are several different perspectives on the change process, (1) focusing upon the innovation (e.g. Rogers, 1995), (2) the user (e.g. CBAM) and (3) the organization (Systemic Change) (e.g. Senge, 1990;1996). The main bodies of ideas are illustrated in Figure 16. Schools are complex systems and so some models might be more useful than others in explaining the change process. The early linear models have limited value because an educational intervention like CAP targets the minds and beliefs of the users and must be implemented with a wide variety of organizational contexts. Programme designers for a multi-component intervention such as CAP must consider multiple change perspectives related to the nature of the intervention. For example, since teachers will be heavily involved in conducting the CAP process, a strategy for facilitating change based on the needs of users would be important. However, since the change is also expected to impact upon the schools, recognition should also be given to the elements and principles of systemic change.

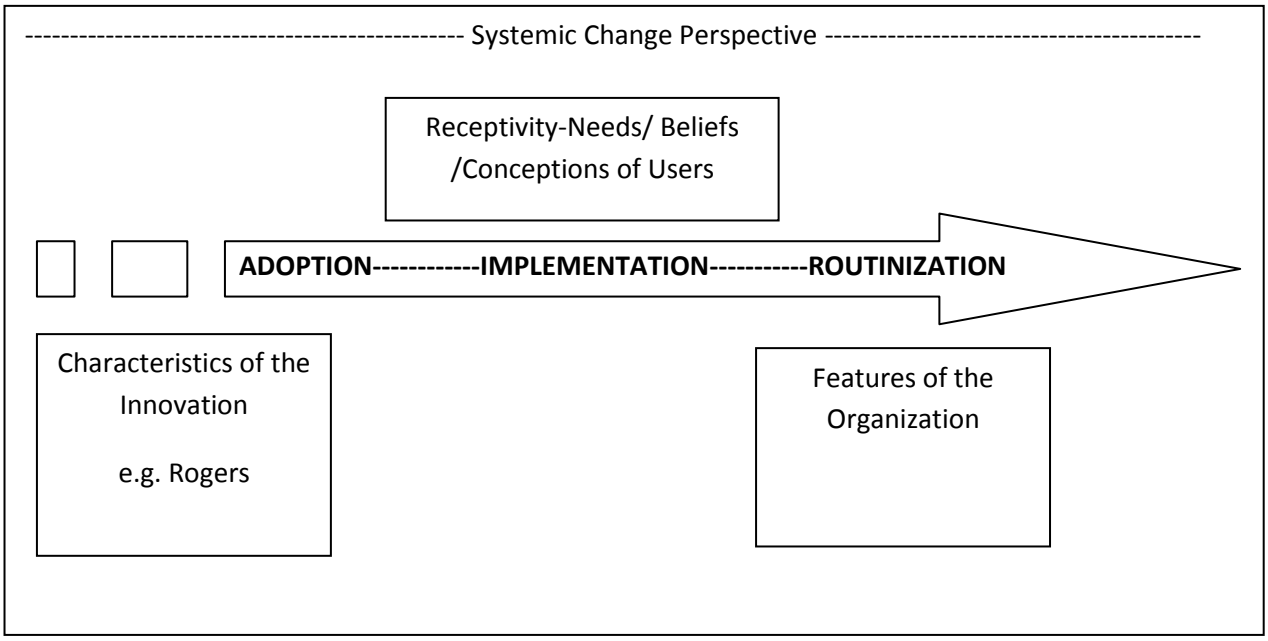


Figure 22: Multiple Perspectives on the Education Change Process

Both assessment and education reform will involve the introduction of some new practice, structure or organizational arrangement. From the standpoint of Rogers (1995), an innovation, is an "an idea, practice, or object that is perceived as new by an individual or other unit of adoption" (p. 11). Any innovation will experience varying degrees of resistance and this might be accommodated in a formal change strategy or plan. Knowing the innovation, stakeholders, and sources of resistance is important because this will prevent planners from minimizing the extent of change. Many innovations in education are subtle and are perceived as normal practice but in reality they might require dramatic changes in teachers' ideology and pedagogical practice. The process of change occurs through three stages: initiation, implementation, and institutionalization or routinization. Adoption is the formal decision to make full use of an innovation as the best course of action available. For the CAP in Trinidad and Tobago this would have occurred in 2000 with full implementation.

Roger's (1995) believed that five attributes were the key to effective adoption:

1) The innovation's relative advantage as compared with the status quo
2) The innovation's compatibility with the individual's existing values, past experiences, and needs
3) The simplicity/complexity of the innovation
4) The degree to which it may be experimented with on a limited (and safe) basis
5) The degree to which its results are visible to others

The CAP scores highly in 4 and 5, but its complexity and incompatibility with current Caribbean pedagogy is worrying. Whereas Rogers focuses upon the innovation, models of implementation in education such as the Concerns-Based Adoption Movement (CBAM) focus on the needs of the users. In CBAM, change is a very personal experience as individuals move through the six levels, (a) orientation; (b) preparation for use; (c) mechanical use or task mastery; (d) routine use and refinement; (e) integration; and (f) renewal or reinvention. Another body of literature looks at the teacher as user and their

receptivity to change. Moroz and Waugh (2000) defined receptivity in terms of overall feelings, attitudes, behavioural intentions, and behaviour. In an earlier study, Waugh and Punch (1987) elaborated on the range of variables that might be included. Their model is illustrated in Figure 23. Feelings about the education system in general and expectations about the impact of a new educational intervention would be significant factors in Trinidad and Tobago. Also important is the perception of support by the school and administration.

Thus, teachers are important agents in the change process. Datnow and Castellano (2000) in their study of teachers' receptivity to whole school reform found that at one site, some teachers may be strong supporters whereas some teachers might vehemently oppose the reform. More importantly, even when teachers supported the reform, they may adapt practices. In some innovations, such adaptation may lead to a lack of fidelity. Implementation can proceed for some time, but when the innovation becomes part of regular practice and is no longer seen as new, then institutionalization has occurred. Institutionalization is about the organization and requires both administrative and teacher action to occur. Older change models are often linear and may not always capture the complex realities of schools. For example, it might well be that innovations are scaled up depending upon local conditions. Newer models of change, such as those proposed by Fullan (2007), focus more on the capacity of the organization to change, as in systemic change. Systemic change requires resources in the form of people, money, supplies, facilities, and time to learn and experiment. The capacity to learn is an important variable as referred to in models by Schein (1996) and Senge (1990; 1996).

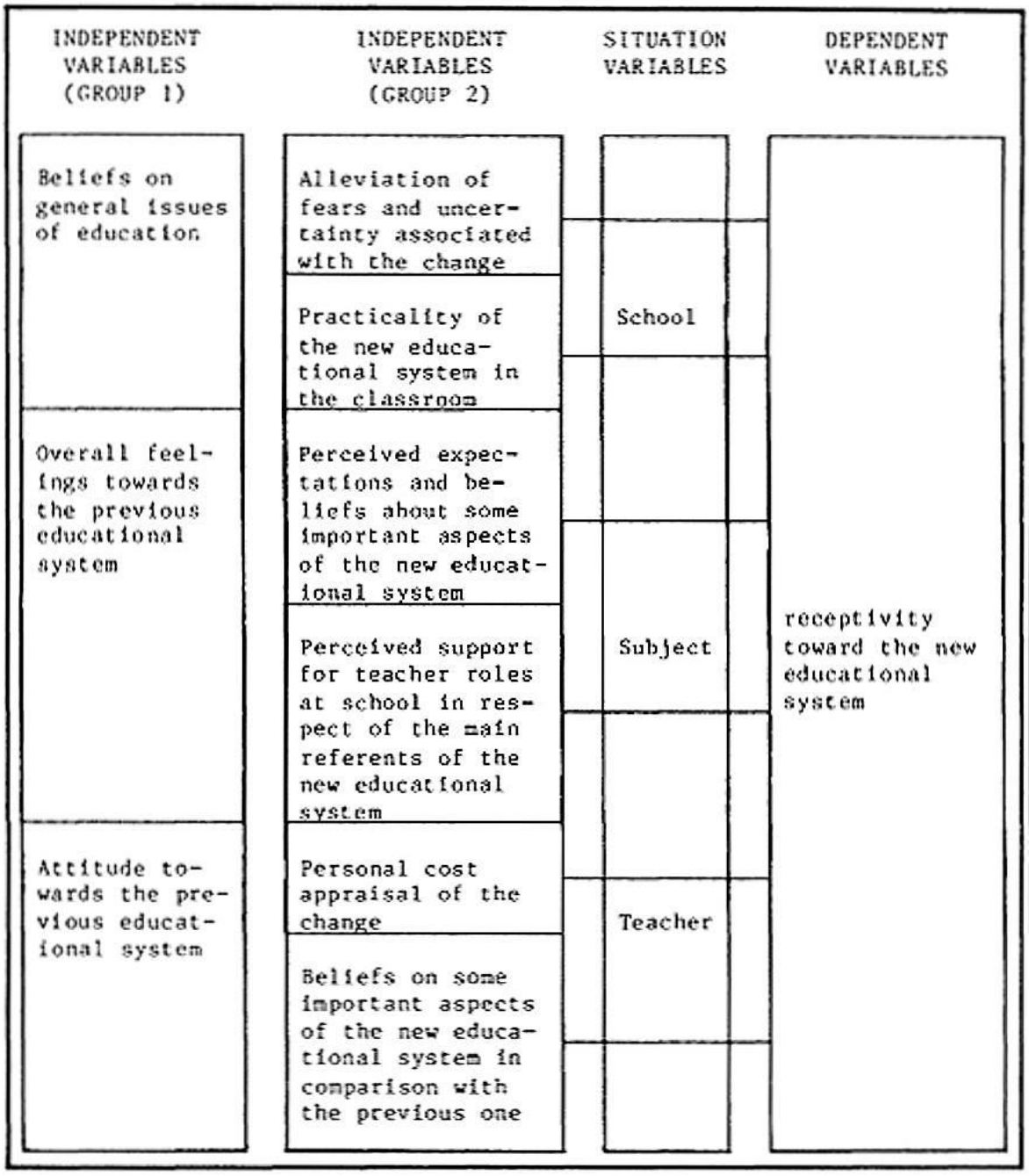


Figure 23: The Waugh & Punch Model of Receptivity to Change (Waugh & Punch, 1987, p. 249)

Although the focus of CBAM and teacher receptivity is on the user, theories about implementation adherence or fidelity turn the focus back upon the role of the innovation during implementation. The argument is that evidence-based innovations must be implemented with the key elements of the treatment that create the impact intact in order to ensure the predicted outcomes. This does not mean that teachers do not have a role in adaptation, but what is necessary is to understand how the innovation works and what elements are to be installed in order for the intervention to prove effective and sustainable. For example, as seen in the case of the African countries, if continuous assessment schemes ignore the formative element, a large part of its utility and impact will be erased.

Fidelity might be defined as the degree of fit between the developer-defined elements of the programme, and its actual implementation in a given organization. Fidelity is aided by preplanning, the clarity of goals and a lack of complexity of the innovation, the level of integration, training and technical support, and implementer and organizational characteristics. There are five facets of fidelity:

1) Adherence refers to whether the intervention is being delivered as designed or written, with all core components delivered to the appropriate population; staff trained appropriately and using the right protocols, techniques, and materials
2) Exposure (or dosage) includes the application of the intervention in the schools including the amount of time devoted to the suggested practices.
3) Quality of delivery is the skill and enthusiasm in delivering the programme
4) Participant responsiveness is the extent to which students are engaged in the program.
5) Program differentiation is the uniqueness of the programme to what was done before.

An important outcome of the change process from the standpoint of models of systemic change is sustainability. An innovation might fail to gain momentum if there is (a) fear and anxiety; (b) a gap between the change initiative and the organization's ways of measuring; and (c) an escalating dynamic of perceived threat and siege mentality. The idea, then, is to deal with these perceive threats and organizational issues in order to fuel the implementation process.

Assessment change

Supplementing the general literature on education change is a smaller body of work on assessment reform or changing assessment practice. Critical factors influencing assessment reform in the literature are variables such as assessment literacy, leadership, professional development, support and collegiality. Assessment literacy has already been highlighted as one of the key variables involved in successful assessment change. Increasing assessment literacy requires professional development that is *continuous* and *ongoing, collaborative* and *site-based*. This is not a common local model, instead occasional workshops and staff development days are the primary strategies for PD. The role of the principal in assessment change has also emerged as a variable of note. Support and collegiality is a component of team-based learning and professional learning communities. A professional learning community is a group of teachers (or entire staff) committed to learning together with the goal of stimulating student success. The core values of the group are shared vision and values, collective responsibility, reflective professional inquiry, collaboration, and promotion of both group and individual learning.

Matthews (2007) analyzed the relationships among teachers' assessment practices, instructional leadership, and student achievement in the middle level schools of Missouri. They found significant relationships between instructional leadership, collaborative assessment practices, and assessment in the classroom. They used both the partial correlation coefficients and the standardized regression coefficients to explain the relationships between variables in their constructed models. As shown, a focus on instructional improvement and curriculum development in the instructional leadership facet influenced collaboration and use

of summative assessment data. Collaboration had mutual relationships with both assessment practice in general and assessment integrated with instruction (formative assessment). Thus, the role of assessment leadership appears important in all aspects of continuous assessment practice.

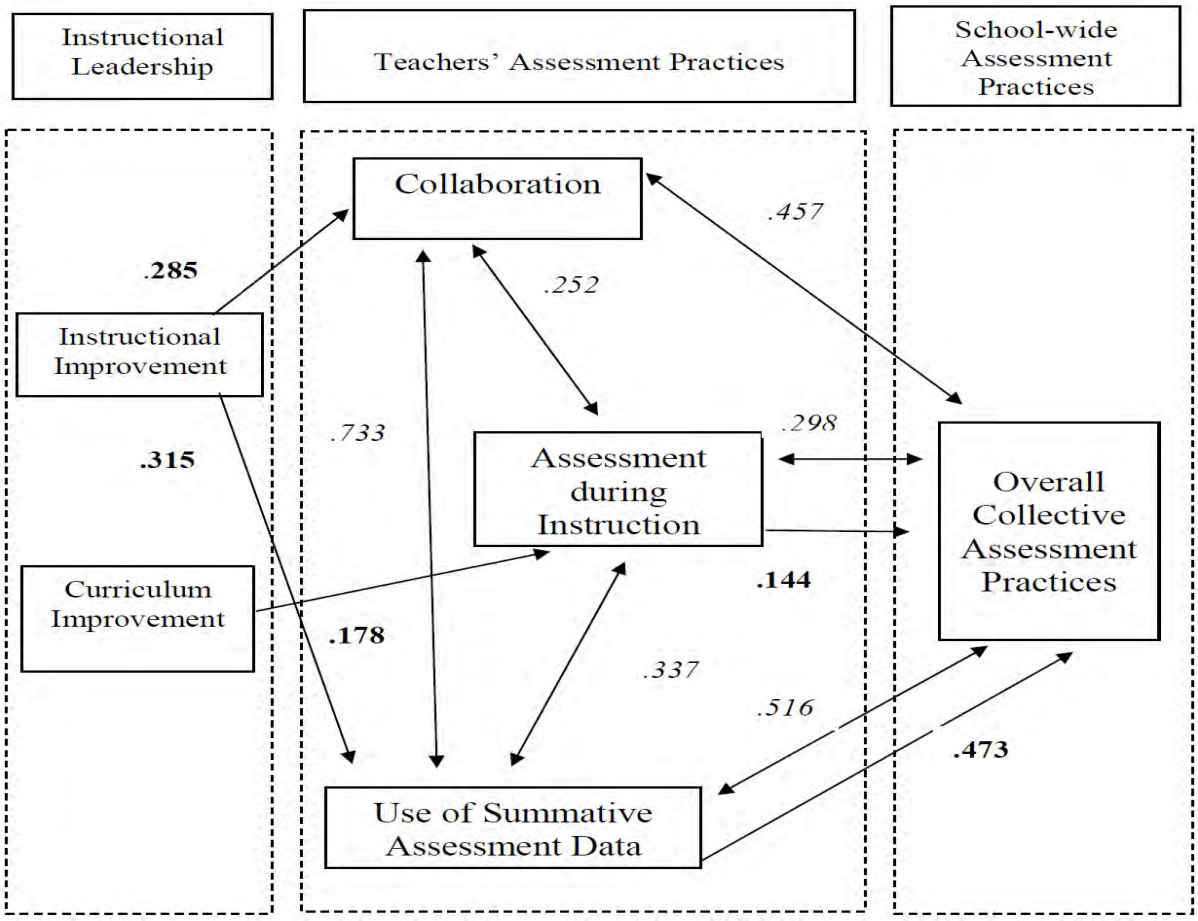


Figure 24: The role of leadership and collaboration on teachers' assessment practices (Morris, 2007, p. 148)

Professional learning communities (PLC) may be especially helpful in influencing assessment change because they help teachers learn and experiment with the many new assessment ideas and practices in the classroom in a supportive team. In the PLC, teachers are provided with a safe, collegial atmosphere where they can test new ideas in teaching and reflect on ongoing practice. Thus, PLCs will facilitate both the introduction of formative assessment and DDDM, as indicated in one US study of two elementary schools, in which a teacher noted:

> We are data driven here. In our PLC groups, we look at the data. We look at how we are comparing with the other classrooms, as far as our grade level. We ask ourselves, we have discussions, and we determine our weaknesses. If my students have fallen off in math, or if my students are dropping in reading, I ask others, "What are you doing in math or reading that I might not be doing?" We ask ourselves, "How can I be better at doing what I am doing?" . . . During PLCs, our instructional coach will give us strategies to use and take back into our classrooms that we are

> expected to use. After we try them, we come back and report on how we used them. We share what we used, we put it up on the board, and then the instructional coach shares it all throughout the day with the whole staff. We all have our different ideas of how to use the strategies, so we can take others ideas back to our classroom (p. 143).

Another benefit of a professional learning community is the sense of team that provides support motivation and feedback during experimentation (Stiggins, 2000).

The quality of implementation is an important variable to consider when considering education change. Usually, an intervention is designed to improve student learning, but it may not do so, if the level of implementation is low or if the intervention is adapted in ways that alter the way it works. This can happen if teachers do not fully understand the intervention or beliefs and ideologies conflict with expected practice. Thus, a third focus on users directly applicable to the work on assessment reform comes from the research by Gordon Brown in New Zealand. Brown (2004) focused first upon conceptions of assessment directly influencing the use of assessment by teachers. His later work extended to conceptions of curriculum and teaching and learning (Brown & Lake, 2006). Table 9 summarizes the different conceptions in each of the three areas. He identifies four conceptions in assessment and pedagogy and three in curriculum. Although teachers may score in all these areas, some conceptions may be dominant and have variable impacts upon assessment practice.

Table 9: Conceptions of Assessment, Curriculum, and Pedagogy (Brown, 2004)

Conceptions		
Assessment	Curriculum	Pedagogy
School Accountability	Academic	Apprenticeship-Developmental
Student Accountability	Social Reconstruction	Nurturing
Improvement	Technological	Social Reform
Irrelevant		Transmission

Brown's work is also grounded in wider theories about the influence of context upon new innovative practices. This work is useful because it captures an important element of beliefs, the educational ideology of teachers in the Caribbean. It might well be that some conceptions are more dominant in different cultural contexts making implementation of some forms of assessment more difficult to implement. For example, one would think that academic curriculum, transmission pedagogy, and student accountability assessment conceptions might predominate in Caribbean contexts.

Carless (2005) developed a model that examined the relationship between wider social factors and teacher beliefs in the implementation of assessment practice. His model

illustrated in Figure 25 shows that for innovations such as assessment of learning, the personal domain forms an inner layer of influence. However, this later is also influence by the immediate school context and wider societal philosophies. Thus, if assessment for learning runs counter to wider societal understandings or the values and expectations of parents, it will likely be resisted. This means that a change strategy must also consider other stakeholders in the system. To a large extent, this was admirably attempted in the CAP, although the impact was not sustained.

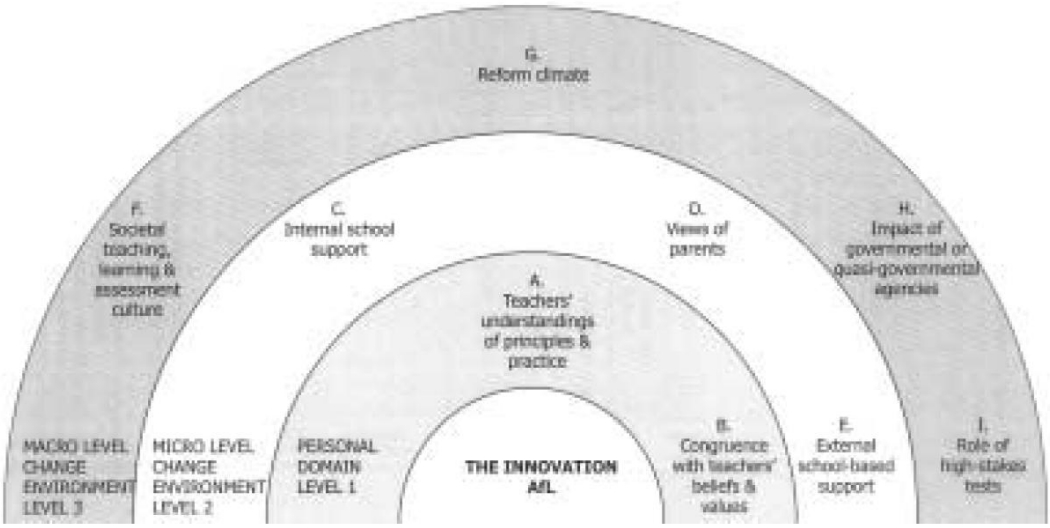


Figure 25: Layers influencing the implementation of assessment for learning in New Zealand (Carless, 2005)

The operation of continuous assessment within a nationwide assessment scheme

Studies of evaluation reform in both the developed and developing world confirm that the operation of continuous assessment systems is strongly influenced by external assessments in the system. External assessment can indeed have an unpredictable and negative impact on both instruction and formative assessment practice embedded with instruction. The effects may be magnified in societies which are very test-oriented. Less than a decade ago, Broadfoot (2002) reminded readers of the journal, Assessment in Education, of the powerful influence of assessment.

> Assessment activity now shapes the goals, the organisation, the delivery and the evaluation of education. For children starting school, every aspect of their lives is likely to be framed and shaped by the demands of assessment, whether this is the assessment activities they themselves are subject to, such as weekly class tests, national testing sweeps at regular intervals, mandatory diagnostic testing, public examinations or entrance examinations—or the results of the assessments that their teachers and their school are subject to, the effects of which are likely to be felt in terms of curriculum priorities, teaching methods, homework policies, classroom organisation and so on (pp. 285-286).

There are several theories describing this phenomenon and two will be discussed in this section, *measurement-driven instruction* and *washback*. Measurement-driven instruction is a very simple but extremely intoxicating idea. Popham (1987) argued that MDI might be more

cost-effective in improving education than replacing mediocre instructional materials with "empirically proven alternatives improving the quality and skill of teaching staff. Airasian (1988) described the strategy in this way:

> Measurement Driven Instruction (MDI) involves the use of high-stakes achievement tests to direct the instructional process. The logic of MDI is that when an important consequence or a high stake, such as obtaining a high school diploma or a teaching certificate, is tied to test performance, the content reflected in the test will be incorporated into instruction. The consequence associated with test performance will force an instructional response and the content of the test will "drive" instruction. The higher the stakes, the greater the impact on instruction (p. 6).

There are elements of measurement-driven instruction in high stakes accountability standards-based systems. Indeed, some have promoted measurement driven instruction especially in developed countries where selection examinations persist (Heyneman, 1987; Eisemon, 1990). Measurement-driven ideas appear to have been important in the context of Trinidad and Tobago, possibly influencing the use and design of the Common Entrance Examination (London, 1997; Chapman & Snyder, 2000).

MDI met much opposition from the outset. For example, Bracey (1987) in the same issue of *Phi Delta Kappan* suggested that MDI was impressive sounding but would simply enhance the level of fragmentation in US classrooms. By fragmentation, he meant the tendency to break learning up into little pieces and to treat those pieces in isolation. Several early studies pointed to unintended negative effects. Smith and her colleagues carried out a series of studied in the 1990s to examine the effects of testing programmes on teachers in elementary schools (Smith, et al. 1989; 1991; Smith & Rottenberg, 1991). Smith and Rottenberg (1991) found that the effects of high stakes testing were to reduce instructional time and to exclude topics not given in the test. Additionally, teachers tended to use materials that resemble the test. Moreover, selection and placement systems could exert pressures on primary schools to alter organizational practices. These four areas are very evident in examining the effects of the Eleven plus system in Trinidad and Tobago.

Emerging in the 1990s, especially in the field of English as a Second Language (ESL) testing, was the concept of washback. The term washback (originally backwash) has come to mean an active direction and function of intended curriculum change by means of the change in external examinations. The phenomenon is complex with both positive, negative intended, and unintended effects. Thus as Diane Wall has argued the question is, can washback effects be controlled? (Wall, 2000). Her work (Wall, 2000, 2005) and that of Cheng and his colleagues (Cheng, Watanabe, & Curtis, 2004) suggest that the answer to that question is, no.

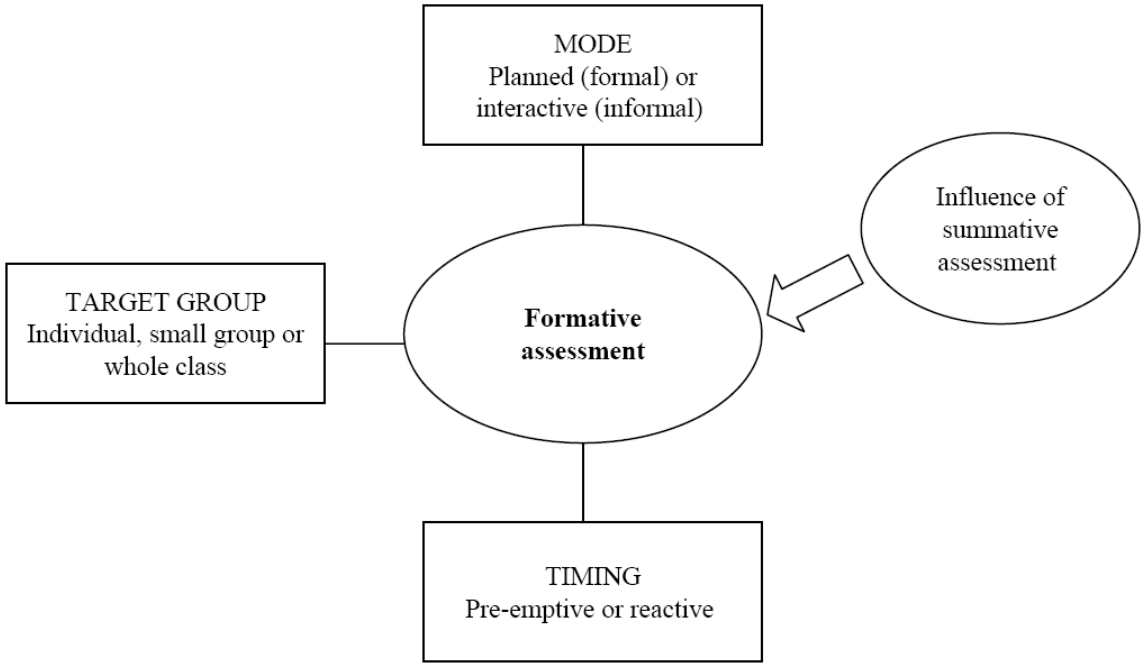


Figure 26: The relationship between summative and formative assessment (Carless, 2007, p. 177)

Thus, a system of continuous assessment will not be immune to the effects of national testing and the 11+ selection/placement examinations and the effects should be considered antagonistic and unpredictable. It is best to model the relationship between internal and external assessments as shown in Figure 26. The summative assessment is not supportive of formative assessment practice, even when that summative assessment is internal and the design of classroom assessment needs to consider this fact. Moreover, changes made in internal assessment must be accompanied by changes in external assessment. It follows that a single minded focus on the Eleven Plus and national testing systems without an accompanying concern for classroom assessment is foolhardy, because the effects on learning of the latter (assuming that there is quality formative assessment) are significantly higher.

Education Change in Developed and Developing Countries

It is important to understand that assessment change will be different in different contexts. While the literature does not necessarily support the view that continuous assessment is a Western innovation, it is clear that change models must be adapted to learning environments and school improvement processes in the South. Shandomo (2008) reminded us:

> It is important to understand real life situations in developing countries. Schooling in developing countries takes place under conditions that are very different from say, industrial countries. In many developing countries that are low income, students are likely to attend a "shelter-less' school, one that is poorly constructed and equipped. Their curriculum is likely to be poorly designed. Sometimes the teacher will only have 10 years of formal education. The learning environment typically has few resources, and classes consist of more than fifty children, some of whom may be chronically undernourished and hungry. The job of introducing

> worthwhile innovation in these countries is thus significantly more difficult than in developed countries (p. 48).

Although Caribbean countries such as Trinidad and Tobago have invested heavily in the basic education infrastructure compared to developing countries such as Africa, great variations still exist in the quality of schools and learning environments. Even problems such as student absenteeism, which can affect the impact of Continuous Assessment, might still be a problem in some contexts, as has been noted in the PIRLS 2006 dataset. Thus an evaluation study might consider theory, but must also ground that theory within the context of the social and educational environment.

The primary issue in the resource-deficit environments of the South might relate to achieving sustainability. Thus, Shondomo (2008) noted that some innovations might thrive for a while with consensus and authority even though there are limited resources. However, these changes will not prove sustainable even if the energies and commitment of teachers drive the reform in early implementation. Using I stands for infrastructure, C for consensus and A for authority, Shondomo (2008) noted:

> In some innovations it is possible to have (I- C+ A+), an educational system with (I-) and yet successfully implement a large-scale innovation. In such innovations, numerous problems encountered in the implementation, many of which have resulted from rapid planning and over ambitious solutions, are overcome simply by the energies and commitment of the people concerned. However, most of these innovations survive just a little beyond the introductory phase (p. 58)

As identified by the programme designer, the notable deficiencies of the CAP on implementation seem to relate to the lack of resources. It is notable that in such conditions even the commitment of the stakeholders and teachers would not have been enough to ensure sustainability.

The Evaluation Design

Elements of the Evaluation Design

This evaluation comes at the mature stage of implementation, as described in Chen (2005). The mature implementation stage is that which follows initial implementation activities, when the programme is first put into action. At the mature stage, one would expect that rules and procedures are put place and the initial fluidity of the programme has been significantly reduced. However, in the mature stage, clients may still be concerned with the adequacy of services but will likely progress towards concerns over improvement and accountability. Another expectation would be the availability of data from a viable working monitoring and evaluation system. However, no monitoring and evaluation data was available for the CAP to supplement the findings of this mature stage evaluation. Monitoring and evaluation data was not available from the pilot phase. From an analysis of the Operational Manual for the Pilot Stage, it appears that structures were set up in the pilot phase for monitoring and evaluation, but these were not implemented.

Figure 27 provides a road map of the adoption and implementation of the CAP over the time period, 1998 to 2010. An evaluation at the final stage might consider possible outcomes or impact, reconstruct some of the processes to indicate effectiveness and strength of implementation, map the current state of practice, and postulate on the overall effectiveness and worth of the installation.

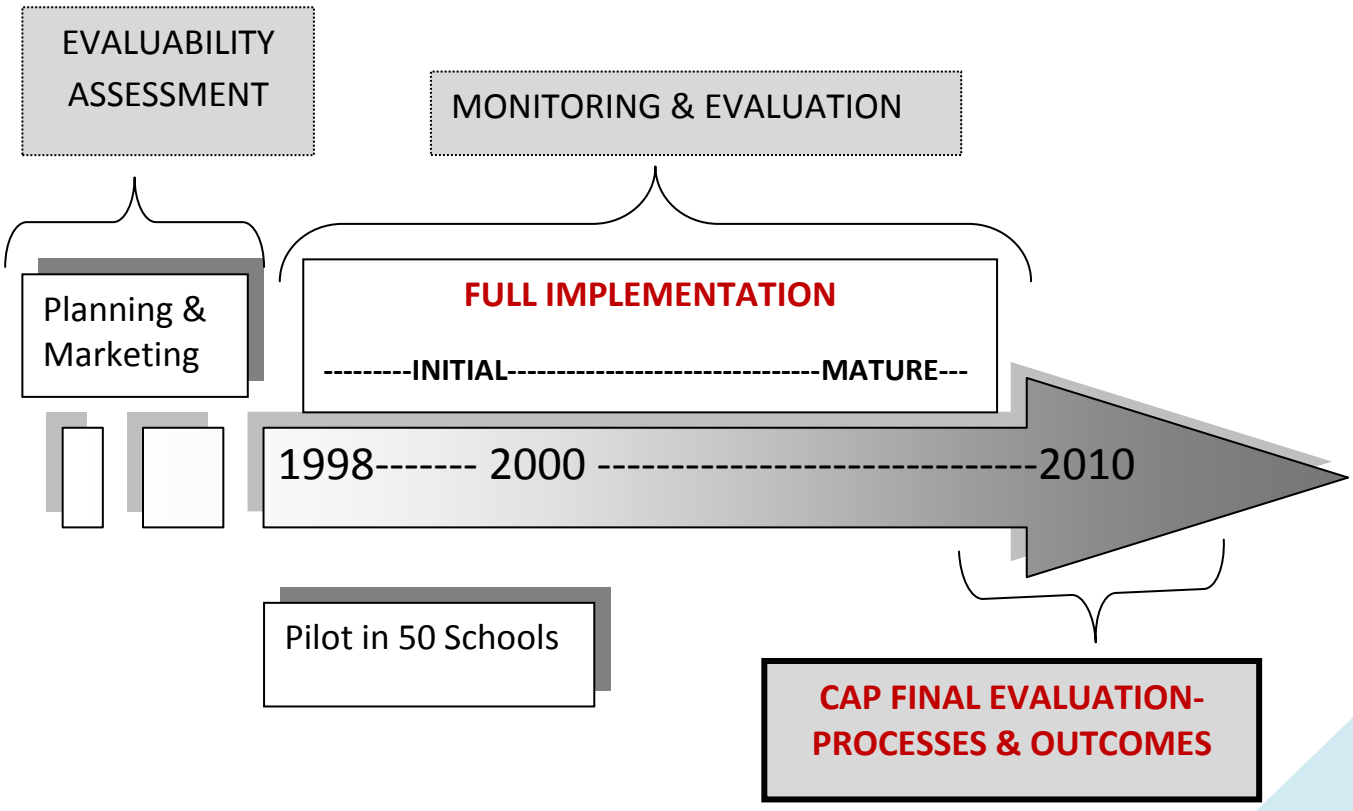


Figure 27: Implementing and evaluating the CAP in Trinidad and Tobago

Choosing from Models and Approaches

There were several approaches and models available to guide the design and conduct of the overall evaluation study. An evaluation model or approach might be considered as an overall framework for designing and conducting the study. Stufflebeam (2001) considered an evaluation approach to be an idealized or model strategy for conducting an evaluation usually guided by explicitly stated beliefs and practices. Although a single model might be chosen, it is also possible to combine multiple approaches according to Blesdoe and Graham (2005). Stufflebam (2001) classified some 22 evaluation approaches into four categories: (1) pseudoevaluations, (2) questions/methods oriented, (3) improvement/ accountability-oriented, and social agenda/advocacy approaches. Hansen (2005) considered 5 categories and 13 approaches. These categories include results, explanatory process, system, economic, actor, and programme theory models.

There were four characteristics of the evaluation approach adopted in the study. These were that the investigation (1) was programme theory driven, (2) made use of causal modelling, (3) employed a mixed methods research design for data collection and (4) included a comparison between pilot and non-pilot schools.

Programme Theory-driven Approach

Theory-Driven evaluation is a method-oriented approach that focuses upon developing and explicating programme theory associated with the intervention. The purpose is to provide scientific evidence along with formal procedures to guide the evaluation. Gascon (2006) considered theory-driven evaluation to be "a holistic assessment of a social program or intervention based upon the conceptual framework of program theory that is oriented towards scientific and stakeholder credibility and provides a foundation for program improvement" (p. 17). He stressed that "change and action models dominate the landscape of program theory" (p. 17).

Donaldson and Gooler (2003) provided a working definition and purpose for theory driven evaluation approaches:

> One of the central tasks of the theory-driven evaluation is to fully understand the nature of the program, the true purpose and context of the evaluation, in an effort to design the most rigorous and sensitive evaluation possible within practical constraints. This is accomplished by developing program theory that is used to identify and prioritize the important evaluation questions, and to tailor the evaluation methodology to answer those particular questions (p. 355)

Therefore, there is an assumption that the intervention is some sort of theory designed to create an effect. Thus, theory driven evaluation is about explicating the theory of the implemented programme. The evaluation of the CAP is especially suitable for a theory-driven approach for the following reasons:

1. The CAP intervention includes elements associated with both a complicated and complex design.
2. There is substantial theory explaining predicted outcomes for formative assessment and data use.

3. A mechanism for change is specified in the original CAP intervention, which can be incorporated in the Theory-Driven evaluation.
4. Causal modelling can produce high quality, scientific data, providing credible evidence to inform future implementation.

Rogers (2008) defined a complicated intervention as one having multiple and/or causal strands and multiple agencies involved. A complicated intervention might involve both simultaneous and alternative causal strands whereas a complex intervention includes both recursive causality and emergent outcomes. The latter occurs when some features of the programme develops on implementation. CAP has elements of both a complicated and complex intervention in that it (1) includes multiple components, (2) is implemented across varying sites, (3) is managed by several MOE departments, and (4) may require teachers to develop and crystallize elements of the innovation (An emerging aspect of the innovation).

Theory driven-evaluations provide an alternative approach to quasi-experimental and experimental approaches to gathering credible evidence for complex interventions in the context of quality improvement. This approach involves developing a common understanding of how the intervention works to create impact. Theory-driven designs are supplemented with logic models, concept modelling, and other graphic organizers that illustrate functioning. These graphic organizers illustrate the causal links between different components and stages of the programme (Donaldson & Gooler, 2003; Rogers, 2004). The study by Gascon (2006) illustrates the use of a theory driven evaluation for studying assessment programmes. He used programme theory to evaluate the school improvement thrust in the Canton City School District, Ohio, which consisted primarily of teachers using both formative and value-added assessments, based on summative assessments. Gascon generated five hypotheses and used hierarchical linear modelling to link the growth in achievement scores with both the formative and value-added assessments. He found some evidence for the influence of formative assessment on gains in student achievement.

Causal modelling is an important and useful feature of theory-driven evaluation. There are two types of models, theoretical and empirical models. Rogers (2000) explained that theoretical models must explain both how the programme is understood to work and provide empirical models of how the programme actually works. The theoretical or normative causal model is specified in the literature review while the empirical or descriptive model is the explanation and meta-inferences derived from the quantitative and qualitative data. The programme theory can also accommodate change theory explicated as part of successfully implementing a programme. The change theory will inform the programme rationale for the programme. The change theory will specify the mechanism through which the intervention acts to achieve the outcomes. The CAP included consideration of change strategy although assumptions and influential factors were not sufficiently explored.

Designing a theory-driven evaluation will involve five steps. The evaluation study must explicate and map the theory. This has already been done in the literature review. Formalizing the theories to be put to the test will occur in the following section. An explicit procedure will be adopted for data collection and analysis.

1. Eliciting and surfacing the underlying programme theories
2. Mapping and selecting the theories to put to research
3. Formalising the theories to put to test
4. Data Collection
5. Analysis

It is necessary to clarify through the programme plan, the action model. This model might include elements such as intervention and service delivery protocols, implementing organizations, program implementers, associate organizations/community partners, the ecological context of the program, and the target population. In the model constructed in Figure 28, the antecedents are understood to have an independent effect but might well operate interactively across different school sites and education districts. A logic model will illustrate some of these features, but a simple linear model cannot capture all the features of a complex intervention such as CAP. As shown, several resources (determinants) needed to be in place to ensure that all CAP activities occurred. If this were done, it might ultimately lead to the expected outcomes and final goal of improved student learning. It might also be argued from the logic model that achieving these outcomes would be dependent on ensuring that teachers were fully engaged in all the activities across the various school sites.

The literature review has already provided substantial theory for the possible operation of the programme. There is a solid body of work describing the way classroom formative assessment is supposed to look and function in improving student learning. Admittedly, some of this work is quite recent (Andrade & Cizek, 2010). The other body of theory was data driven decision-making and more specifically data use in instruction. This theory is pertinent here because the collection and use of achievement data by teachers was a central part of the CAP innovation. To develop the theory further, continuous/formative assessment programmes at both primary and secondary levels in both Western and non-Western countries were reviewed. The continuous assessment programmes in developing countries are longstanding, but they were not necessarily focused upon formative assessment nor do they explicate formalized data use strategies when compared with the programmes in more developed countries, such as the United Kingdom, the United States and New Zealand. However, the review of assessment reform in African countries highlighted important issues such as teacher resistance, large class sizes, and lack of resources. In some countries, continuous assessment scores were combined with final examinations for selection/certification processes at the end of the first cycle of schooling.

Developing the Graphic Models

Figure 29 is based on the current and evolving body of work in formative assessment and data use. It presents a graphic model for the idealized operation and predictive impact of CAP based on the theory explicated in the literature review, CAP documentation, and interview with the programme designer. The model captures how and why the evaluand works. In Figure 29, it is hypothesized that the primary impact might be through the formative assessment elements of CAP; with feedback improving student's functioning in the classroom environment. The CAP documentation pays considerable attention to the conduct of diagnostic assessment and early intervention in line with the stated focus on "readiness for learning". The system is also decidedly focused on multimodal assessment and record

keeping which is meant to provide data for use by teachers and also by the centralized authority.

The theorized model of change and impact for the Trinidad and Tobago CAP is presented in the graphic illustrated in Figure 30. This change model is considerably expanded from the original CAP documentation and incorporates modern elements of assessment change applied to the context of Trinidad and Tobago. This includes antecedents such as organizational characteristics, assessment and curriculum beliefs, leadership and support, assessment literacy, and fidelity and intensity of implementation. Several of these elements were considered by the programme designers, but not in the detail provided here. In Figure 30, there are three groups of antecedent variables: organizational, teacher related, and leadership-focused. The organizational variables include both contextual and institutional characteristics related to change capacity and organizational learning whereas the teacher variables includes knowledge, attitudes, and practices related to assessment and instruction. The teacher variables capture teachers' receptivity to the change process (Moroz & Waugh, 2000). Assessment leadership is considered to be a critical variable in the process. These variables are hypothesized to impact on both the intensity and fidelity of the programme, but with different levels of valence. Intensity and fidelity, in turn, will determine the expected positive impact.

PROBLEM STATEMENT: Assessment in the primary school is currently dominated by the Eleven Plus resulting in the absence of diagnostic systems and target setting at early levels in the primary school (Contained in 1993-2002 White Paper, p. 2, 47). This reduces the capacity to reach all students.

GOAL: To upgrade the practice of ongoing teacher-led assessment in the primary school system, thereby establishing benchmarks and targets for diagnosis, remediation and intervention and to improve "students' readiness for learning".

RATIONALES

Focus on assessment as learning by teachers and continuous recording of data on students will enhance the quality of classrooms and minimize the current focus and emphasis on Eleven Plus testing.

ASSUMPTIONS:
Existence of Assessment & Evaluation Unit. Teachers have the capacity to learn and understand the nature of the reform.

RESOURCES

Assessment & Evaluation Unit

Teacher Training

MOE Publications

Informed Leadership

Supportive physical infrastructure

ACTIVITIES

Regular staggered monthly tests

Performance Assessment

Record Keeping

Screening and Referrals

Feedback

Use of multiple assessments

OUTPUTS

Student Cumulative Record Cards

Student assessment products

Diagnostic and Remedial Systems

Teacher Journals and other records

Enhanced student learning

OUTCOMES

Changes in the nature of teaching-learning

Improvement in Student Achievement

Provide data for school and MOE

Support of remediation and diagnostic systems

Improvement in the school's effectiveness as measured by student learning

EXTERNAL FACTORS: Some linked components not fully developed, e.g., SEN. Initial opposition by Teachers' Union focusing on lack of resources in schools. Slow pace of education reform in Trinidad and Tobago

Figure 28: Logic model for the CAP

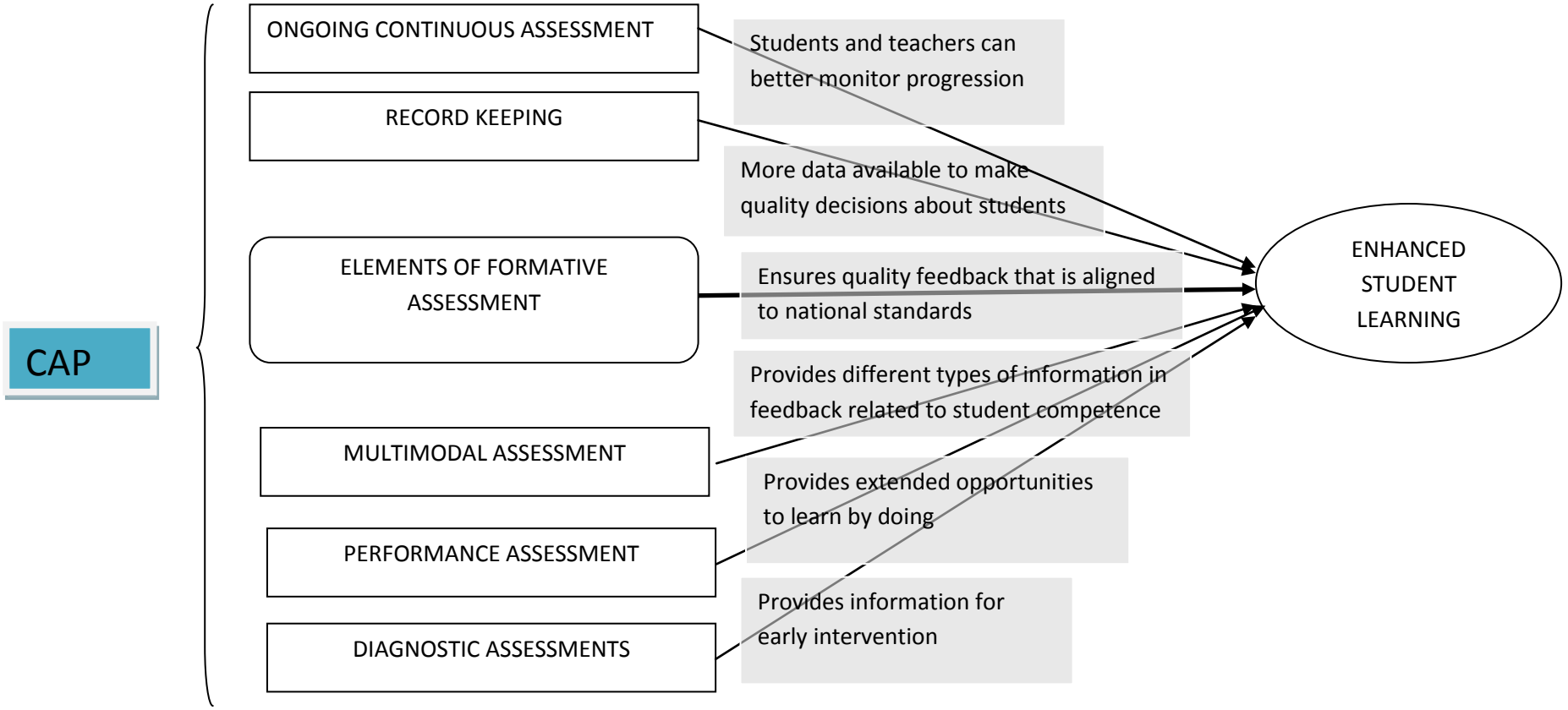


Figure 29: CAP programme theory-Hypothesized causal links between CAP and student learning outcomes

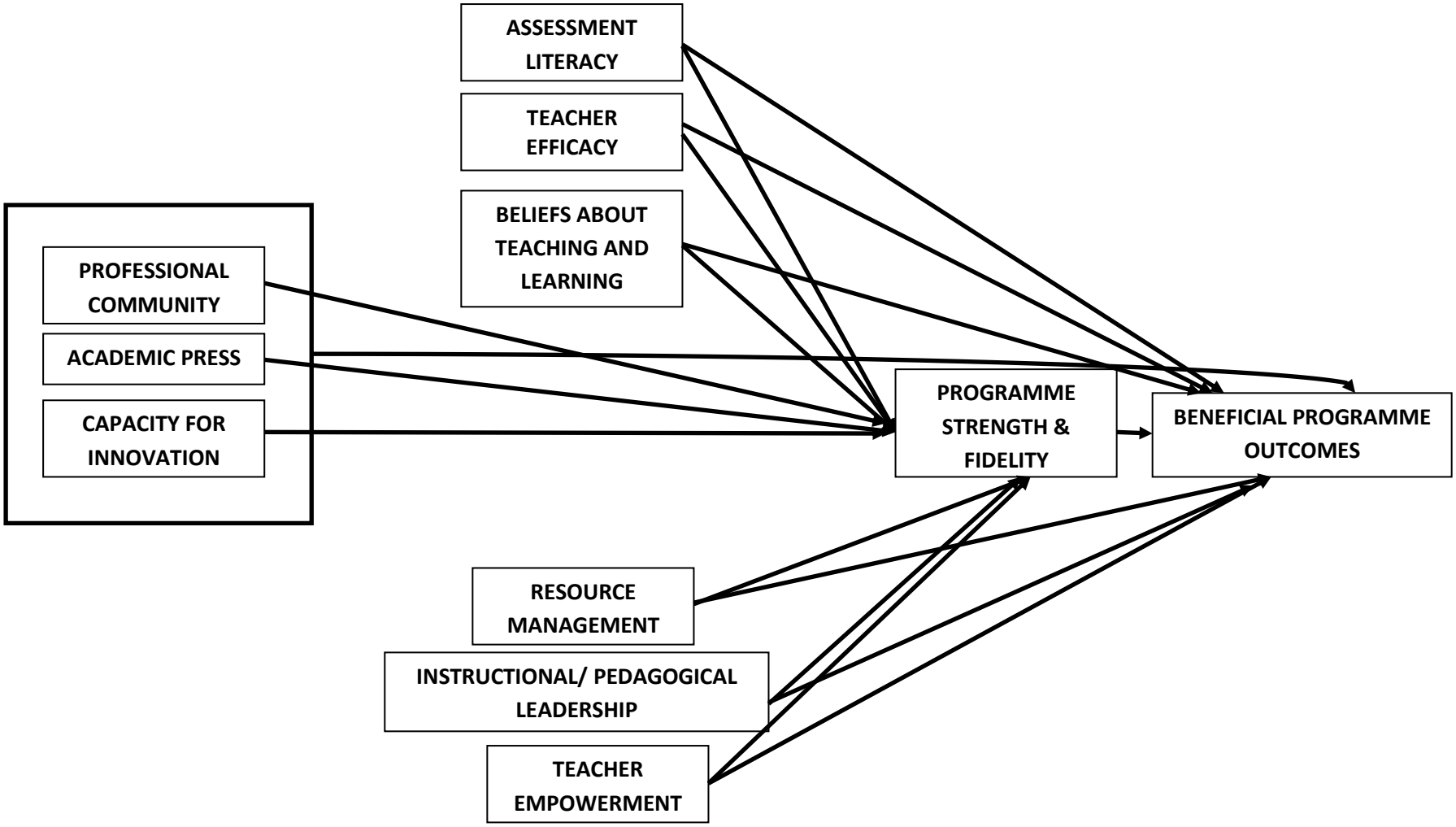


Figure 30: Critical elements of change and impact when implementing CAP in Trinidad and Tobago

Focusing the evaluation

The step, focusing the evaluation, involves clarifying and specifying focal areas for the study. According to Donaldson (2007), focusing the evaluation by creating evaluation questions should come after the programme theory is explicated. The theory behind the programme should guide the focusing, including the development of research questions. Using a broad framework of system issues, innovation characteristics, and various stakeholder concerns in the terms of reference, 19 general evaluation questions were first developed as listed in Box 7. These are focused on antecedents in individuals, organization, and context, which is in line with the multiple perspectives on change describe earlier. Twelve of these questions were then prioritized by considering the financial and labour constraints and linked to specific data collection activities. These are listed in Table 10

---------- Box 7: EVALUATION QUESTIONS ----------

System Concerns

1) **What were the leadership and organizational issues involved in managing the CAP implementation?**
2) **To what extent was support available at the Centre and Divisional Offices?**
3) **What was the extent of monitoring and evaluation in place?**

Innovation Characteristics and Implementation

4) **What was the level of clarity and detail of the documentation on the innovation?**
5) **How was training and support organized?**
6) **What were the essential design features of the assessment innovation?**
7) **What was the level of use among teachers?**
8) **What was the quality of the programme and fidelity of implementation?**
9) **What were the system and site concerns regarding adoption and/ or implementation?**

Site and Leadership Factors

10) **What were principals' perceptions of the innovation?**
11) **What leadership and management roles did the principals adopt during implementation?**
12) **How did the principal manage site-based professional development and training?**
13) **What site factors might have facilitated or hindered implementation**

Teachers

14) **What were teachers' assessment literacy/beliefs/perceptions/practices associated with this assessment innovation?**
15) **What was the relationship between teacher's assessment literacy, beliefs, perceptions and classroom assessment practice?**
16) **What were the levels of use and concerns about adoption?**
17) **How did individual and contextual factors (including those of classroom and site) influence levels of use and rates of adoption?**

Students

18) **What were student's experiences of continuous assessment in the classroom?**

Parents

19) **What were parent's knowledge and experience with continuous assessment?**

Table 10: Focusing the Evaluation: Selected questions prioritized

	Priority	Data Collection		
		Phase 1	Phase 2	Phase 3
1) What were the leadership and organizational issues involved in managing the CAP implementation?	**High**	√		
2) How was training and support organized?	**Medium**	√	√	
3) What was the level of use among teachers?	**High**	√	√	
4) What was the quality of the programme and fidelity of implementation?	**High**	√	√	
5) What were the system and site concerns regarding adoption and/ or implementation?	**High**	√	√	√
6) What leadership and management roles did the principals adopt during implementation?	**Medium**	√	√	
7) What site factors might have facilitated or hindered implementation	**High**			√
8) What was/ were teachers' assessment literacy/beliefs/perceptions/practices associated with this assessment innovation?	**High**		√	
9) What was the relationship between teachers' assessment literacy, beliefs, perceptions and classroom assessment practice?	**High**		√	
10) What were the levels of use and concerns about adoption?	**High**		√	
11) How did individual and contextual factors (including those of classroom and site) influence levels of use and rates of adoption?	**High**	√		
12) What were students' experiences of continuous assessment in the classroom?	**Low**		√	

Data collection strategy

In theory, programme theory driven evaluation is considered method-neutral. However, as proposed, the data collection approach was explicitly mixed methods in order to provide both credible in-depth information and empirical data from instruments and measures. The mixed methods research design was a multiphase study. As described by Creswell and Plano-Clark (2010), this design variant has the following characteristics:

> Multiphase combination timing occurs when the researcher implements multiple phases that include sequential and/or concurrent timing over a program of study. Examples of multiphase combination timing include studies conducted over three

> or more phases as well as those that combine both concurrent and sequential elements within one mixed methods program (p. 66).

Figure 31 illustrates the three phases employed in the design. Phase 1 was exploratory and qualitative, Phase 2 was quantitative and explanatory (with some data qualitized), and Phase 3 was qualitative and explanatory. The information from the exploratory study was used to guide instrument development and design of the survey. In the quantitative phase, the target population was 100 schools with 1604 teachers. In the mixed method design, the intention was to place greater emphasis on the quantitative phase.

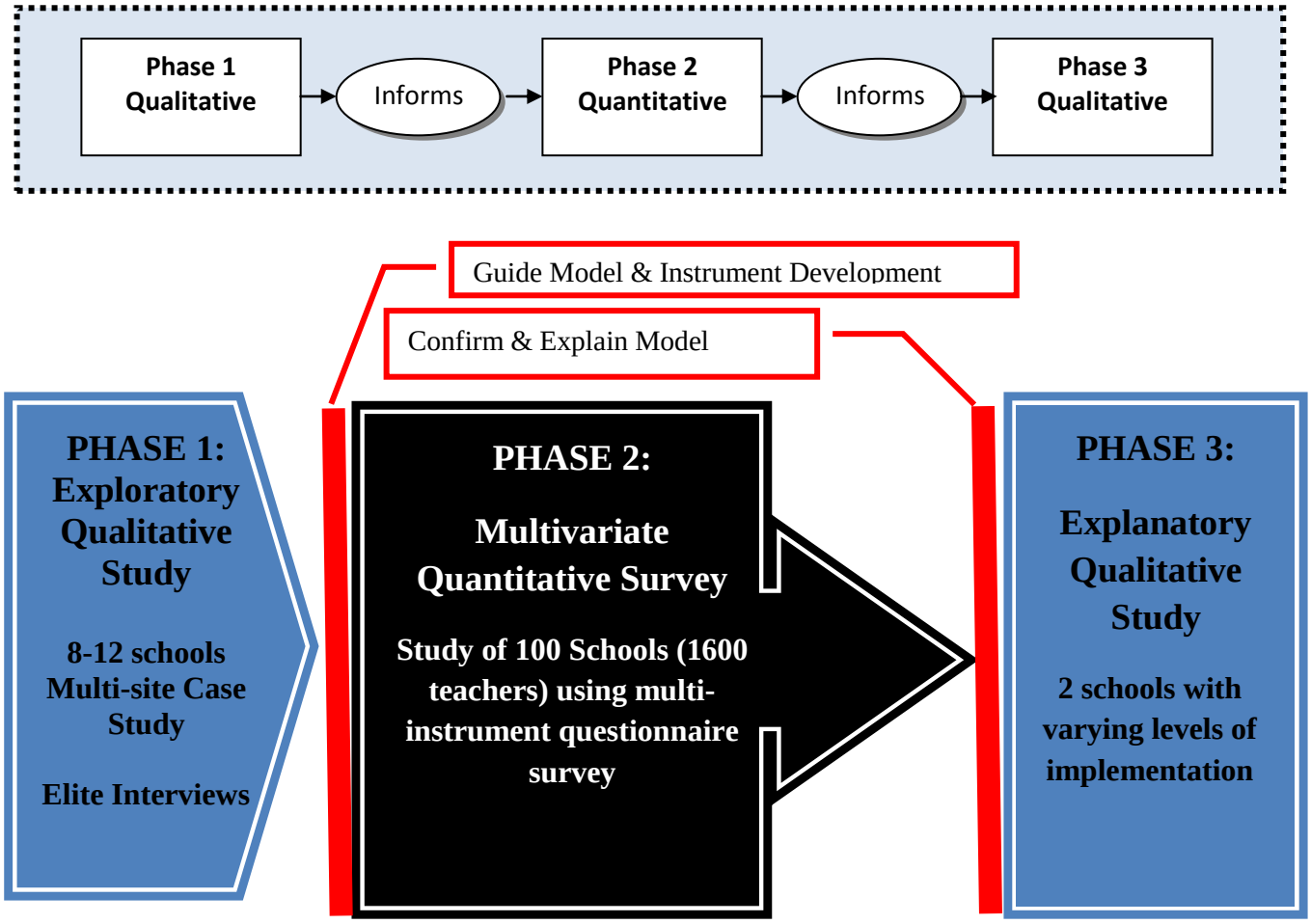


Figure 31: Representation of the *qual-QUAN-qual* multiphase mixed method design used in the evaluation study

Sampling Strategy

Separate sampling strategies were implemented for each of the phase in line with the quantitative or qualitative nature of the study. Qualitative case studies usually have smaller samples, but quantitative studies required randomization and/or proportional sampling along with an adequate sample size in order to achieve generalizabilty.

Phase 1

In Phase 1 elite interviews were conducted with the programme designers and additional personnel involved in installing and piloting the innovation. Twelve schools were then purposively chosen from the 60 sample schools in order to conduct the exploratory survey. The large sample size would allow the capture of variation in contexts across sites. Thus, the schools were selected based on location and overall achievement status (classified high, medium, low based on performance in the 2005 to 2007 national assessments of educational achievement in Language and Mathematics). It was believed that the achievement classification captures the academic challenges faced by the school as well as the nature of key organizational characteristics and teacher beliefs such as collective teacher efficacy. On implementation we omitted the Scarborough RC because of the high cost of travel to Tobago and the limited time for Phase 1.

Table 11: Proposed Sample for Phase I

1. Scarborough RC
2. Anstey Memorial Girls' AC
3. Holy Saviour Curepe Anglican
4. St Catherine Girls' Anglican
5. Rampanalgas RC
6. Carenage Boys' Government
7. St Mary's AC
8. St Ann's RC
9. Arima Presbyterian
10. St Joseph Government
11. Penal R C (St Dominic's)
12. Patna/River Estate Government

Proposed Sample for Phase 2

For the quantitative survey, the target population was 100 schools in the eight educational districts, including the 60 schools identified by the sponsor and 40 pilot schools listed in the CAP documentation. The target population was 1605 teachers and 27, 942 students. We sampled 1600 teachers and 3000 students (just over 10%) in standards 3, 4, and 4. All principals in the 100 schools were also surveyed along with selected school supervisors familiar with the work of the institutions. The return distribution system did not function as expected and the return rate of 35% was rather poor. However, the final sample size was still adequate for regression modelling and since there was additional data from the other Phases, the study could still be repaired.

Phase 3 Sample

The Phase 3 sample consisted of three schools selected based on the levels of implementation: high, medium, and low. On reviewing the data collected from Phase 2, it was decided to look at only 2 schools, one with high implementation and low achievement context and the other with low implementation and high achievement context. This decision was made after the high achieving, high implementation site refused access. Phase 3 data was supplemented by a focus group study of two principals at high implementation sites. A team

of four field assistants visited the schools and conducted interviews, observations, artefact collection and videography.

Table 12: List of 100 schools used in quantitative sample

INTERVENTION SCHOOLS		
Felicity Hindu	Holy Saviour Curepe Anglican	Rio Claro Presbyterian
Gran Couva R C	St Catherine Girls' Anglican	Fifth Company Baptist
Freeport Presbyterian	Salybia Government/Matura	Cocoyea Government
Orange Field Hindu	Jerningham Government	Macaulay Government
Chaguanas R C	Gasparillo Government	Caratal (Sacred Heart) R.C.
Milton Presbyterian	Patna/River Estate Government	Barrackpore Vedic
Tamana Hindu	Princes Town Methodist	Golden Lane Government
Sangre Grande R C	Point Fortin R C	Belle Garden A.C.
Valencia R C/Government	El Socorro South Government	Charlotteville Methodist
Rampanalgas R C	La Puerta Government	
Newtown Girls' R C	Sangre Grande Government	
Nelson Street Girls' R C	Carapichaima R C	
Diego Martin Government	La Romaine Government	
Point Cumana Government	Speyside A.C.	
St Ann's R C	PILOT SCHOOLS	
Belmont Boys' R C	Palmiste Government	
Carenage Boys' Government	Chaguanas Government	
Princes Town R C	Caparo R.C.	
Rio Claro Vedic	Waterloo Hindu	
St Mary's Government	Carapichaima A.S.J.A.	
Sixth Company A C	Caroni Presbyterian	
La Lune R C	Chandernagore Presbyterian	
St Joseph T M L	Coryal R.C.	
El Dorado South Hindu	L'Anse Noire Moravian	
Aranguez Hindu	Sangre Chiquito Presbyterian	
La Horquetta South Government	Belmont Government	
Arima Presbyterian	Diamond Vale Government	
St Joseph Government	Ascension A.C.	
St Mary's A C	Newtown Boys' R.C.	
Malabar Government	Bethlehem Girls' R.C.	
Bourg Mulatresse R C	Morvant R.C. (St. Dominics)	
Fyzabad Presbyterian	Arima Boys' Government	
Cap-de-ville Government	St. Augustine South Government	
Penal R C (St Dominic's)	D'Abadie Government	
Cedros Government	Aripo R.C.	
Siparia Road Presbyterian	Arima Girls' R.C.	
Scarborough R C	Five Rivers Hindu	
Moriah Government	Curepe Presbyterian	
Patience Hill Government	Chatham Government	
San Fernando T M L	Egypt Village Government	
Anstey Memorial Girls' A C	Brighton A.C.	
Harmony Hall Presbyterian	Erin R.C.	
Inverness Presbyterian	Rousillac Hindu	
Mon Repos R C	North Trace Government	
Bien Venue Presbyterian	New Grant A.C.	
St Margaret's Government	Guayaguayare R.C.	

Procedures

In Phases 1 and 3, field researchers were required to spend from 1 to 3 days at each site. After negotiating access into the institution, the principal and groups of teachers were first interviewed. In most schools, this required multiple focus groups. Especially in Phase I, the second and third day involved informal individual interviews and observation of classrooms. Several CAP artefacts were usually collected for further analysis. At some schools, the team returned to view student projects and interview students on their experiences in project led learning,

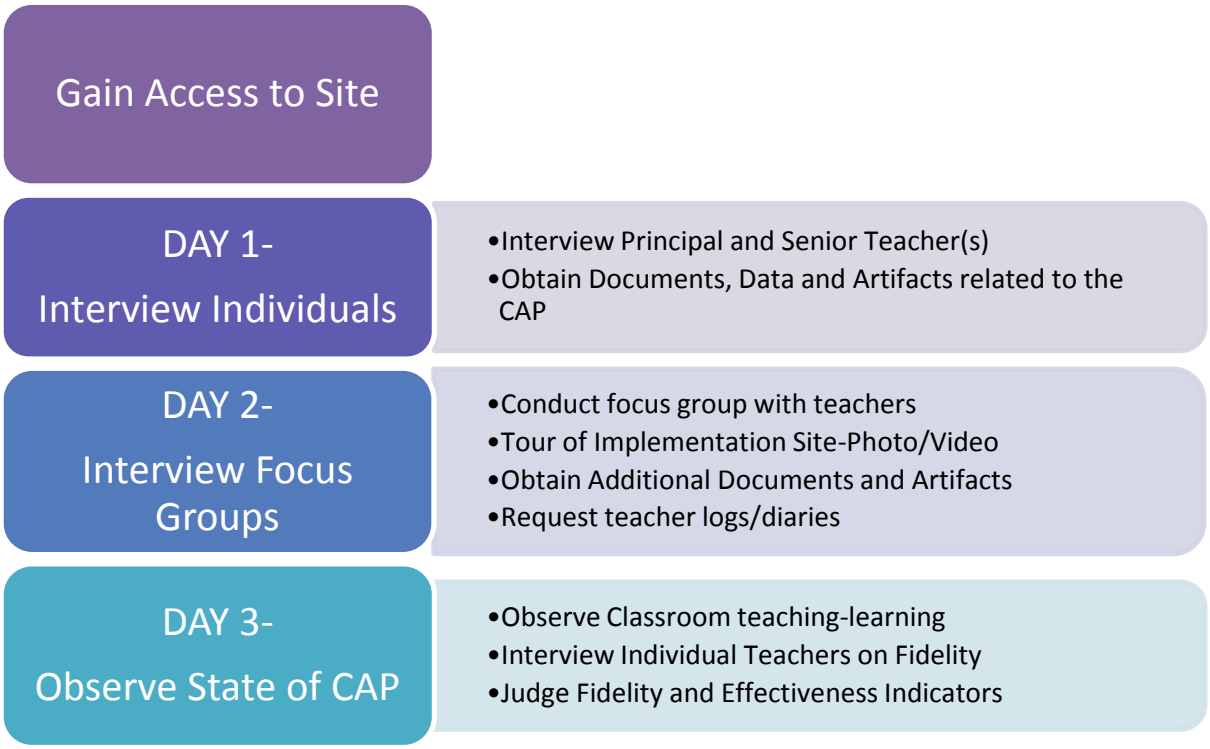


Figure 32: Procedures used in qualitative study of single institutions

Phase 2

Phase 2 involved the development of the questionnaire surveys of teachers, principals, students and supervisors. This process is more fully described in the instrumentations section. Analysis was conducted using SPSS 18, Systat, and SPSS 13 (Advanced & Regression Models).

Instrumentation

Conceptualization of the model and constructs to be measured was informed by the findings in Phase I of the study. Development of the survey questionnaires required an extensive literature review to locate critical concepts and existing instruments that might be used or adapted in this evaluation study. Preference was given to using current instruments since this would provide comparative data and benchmarks. The process of developing the survey is illustrated in Figure 31.

Measuring Predictors

Several viable items and instruments were located in unpublished theses and published texts that analyzed continuous assessment in African countries and others were obtained from the voluminous work on First World education systems. Four separate questionnaires were constructed for teachers, principals, school supervisors, and students. The primary instrument was the questionnaire developed for teachers. A copy of this instrument is included in the Appendix. Bearing in mind the limited period for the evaluation, both review and pilot processes were implemented, but these could not as extensive as they needed to be because of the limited time frame.

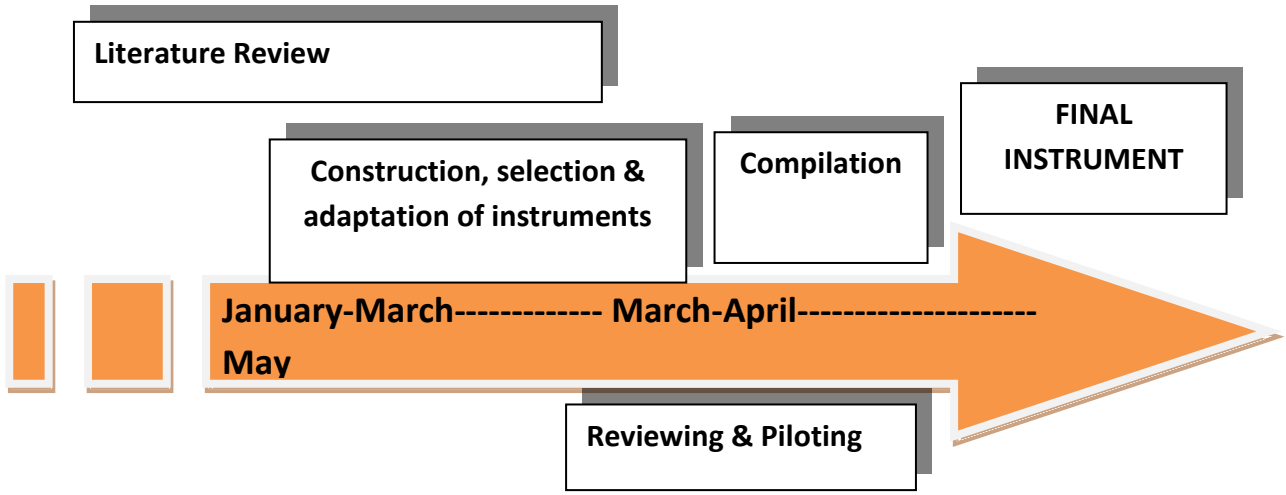


Figure 33: Processes involvement in the development of the survey questionnaires

The model that guided the teacher questionnaire identified organizational, leadership and teacher variables that might be related to valued outcomes, such as teacher use of the intervention. Table 13 provides the main themes captured in the questionnaire design. As shown, several of the instruments from the literature were multiple items scales. The use of multiple item scales was necessary to improve the reliability of measuring each variable, thereby possibly enhancing the ability of the predictive ability of the model. Unfortunately, this made the questionnaire very long, which possibly contributed to the high rate of non-returns and damaged questionnaires. Missing items were handled according to the fixed protocol, which assigned the average value on the scale to the missing item. However, in modelling, individuals with missing items were removed from the analysis.

Table 13: Themes, constructs and measures in Multi-instrument Teacher Questionnaire

Focus	Themes	Constructs	Measures	Location
Teacher	Biographic/ Individual/ Demographic Data	School type and individual characteristics	Age/Ethnic Identity/ Marital Status/Tenure (System)/ Tenure (school)/ Academic Qualifications/ Professional Qualifications/ Class taught/ Classes taught in the last five years/ Work load-	Questions 1-12
School	Organizational Capacity for Learning	Professional Learning Community	SEDL Instrument by Hord et al. (1999).	Page 20-21 (Rubric)
School	Organizational Change	Readiness for Change	Holt et al. (2007)	Page 18 (25 items)
		Organizational Innovation Scale	Adapted from Ismail et al (2002)	Page 6
Principal/ School	Leadership	Assessment Leadership	Perception of Leading CAP	Page 12 (17 items)
Teacher	Knowledge, Attitudes, and Beliefs about Assessment	Assessment Literacy	Teacher conceptions of assessment	Page 9
			Teachers' knowledge of assessment	Pages 22-24
Teacher	Beliefs, Expectations, & Behaviours related to learners	Teacher Efficacy	Teachers' Attitude towards CAP	Page 14 (25 items)
		Extra-role Behaviour	Collective Teacher Efficacy	Page 16 (21 items)
			Individual Extra-role Behaviour	Page 17 (24 items)
			Group Organizational Citizenship Behaviour	Page 19 (24 items)
Teacher	Understanding of teaching and learning in context	Beliefs about Teaching and Learning	Conception of Curriculum & Teaching	Page 15 (19 items)
Teacher	Change Process	CBAM Model of Change	Stages of Concern (SoC) Questionnaire	Pages 10-11
			Levels of Use (LOU) Questionnaire	Page 4/ Question 17
Intervention	Outcomes	Programme Intensity	CAP checklist- Key activities in CAP	Page 5, Question 18.
			Assessment Use	Page 5, Question 19
		Programme Fidelity	Classroom Assessment Practice Inventory	Pages 7 to 8 (67 items)
			Feedback to Students	Page 13 (11 items)
			CAP Definition	Page 3/Question 12

Interview schedules were constructed for both the focus group and individual elite interviews. These contained both open-ended questions and probes as illustrated in Figure 34.

FORM 3-MOE CAP EVALUATION-
Teacher's Interview Schedule

For focus group and individual interview

1. What has been your experience with the continuous assessment programme in this school?
    a. Probe and ensure timeline of perceptions
    b. Clarify awareness and adoption issues
2. What were the strengths and challenges of the CAP at this school?
    a. Probe leadership
    b. Probe support
    c. Probe staff team
    d. Probe training

Figure 34: Excerpt from teachers' interview (semi-structured schedule)

Measuring Outcomes

In line with the stated purpose of the evaluation and the evaluation terms of reference, both the quantitative and qualitative components of the study included explicit measures of programme intensity and fidelity as primary outcomes. The number of CAP tasks completed at each site was considered the most important measure of programme intensity. Of concern, however, was whether these tasks were being conducted in the manner understood or expected by the planners.

Fidelity was considered to be important in classroom assessment reform because of the perceived impact of formative assessment on student learning. However, the magnitude of that impact would depend in large part upon whether feedback and associated activities were implemented as described in theory. In reality, then, low fidelity would mean low impact. Fidelity was thus a primary target of the evaluation study, suggesting that just the presence of the artefacts or activities related to CAP will be insufficient to judge implementation quality. Instead, the intervention must be installed as proposed by the local planners and classroom assessment theory. On observation of sites, we focused especially on formative assessment and feedback in classrooms and this was obtained from the observation rating scales shown in Table 14 and 15, which quantify the key elements in the conduct of formative assessment. These included multimodal assessment, goal setting, varied instructional activities, feedback, and student engagement in the process.

Table 14: Programme Fidelity Observation Instrument

Core Question	Evidence to Look for	Judgement (High=5; Low=1)
1. To what extent does the classroom culture encourage multiple interactions and the use of assessment tools?	Assessments used by the teacher/ group work activity/ projects/ portfolios/ teacher diary/student interviews	
2. To what extent have learning goals for the class been established and is individual progress tracked?	Cumulative Record Cards/Record, Forecast and Evaluation	
3. To what extent is instruction varied to meet the needs of different learners	Record And Forecast	
4. To what extent is assessment varied to meet the needs of different learners	Record And Forecast	
5. To what extent is feedback to learners provided and to what extent is instruction adapted in response to the data?	Teacher diary/interview	
6. To what extent are students active in all phases of teaching and learning?	Record And Forecast/Observation	

Table 15: Programme Fidelity Observation Instrument 2-Am I doing CAP?

Activity	Never	Rarely	Some-times	Often	Always
1. Ongoing, continuous testing **2. Maintains Cumulative Record Cards** **3. Gives varied feedback** **4. Sets goals/standards for learners** **5. Uses multiple forms of delivery** **6. Uses data from students** **7. Uses multiple modes of assessment** **8. Uses quality performance assessment** **9. Keeps Anecdotal Records and Journals** **10. Uses assessments diagnostically**	**Instructions:** Observe the classroom at intervals during the day. You should make about 5 observations of 15 minutes each. You may ask the teacher for artefacts and you should make substantial field notes in your field researcher's diary to be submitted. You may also ask the teacher to keep a diary of his classroom activity for the week of your data collection.				

Programme strength or intensity was also considered critical based on the idea that an intervention might stall and not be scaled up or expanded. Thus, it might be that although the intervention was still practiced in a number of schools, the number of activities and the level of commitment could vary across sites in the system. Table 16 provides the instrument used to assess programme strength on the field visits. As shown, a measure of programme strength was obtained from the extent of implementation in the school and classrooms by estimating the quantity of products and rooms in which the intervention was still active.

A final outcome measure was the comments and rating of students on the assessment practice. We reasoned that students might provide a useful measure for triangulation if they were asked to provide their views on the nature of assessment in the school. We focused the data on Standards 3, 4, and 5. For the quantitative study, we developed a measure of intensity based on the conduct of all the listed CAP tasks. This outcome variables was called *CAP use*. We also developed two fidelity measures, *CAP feedback* measured by a hierarchal list of feedback activity taken from Stoute and in modelling we used *CAP multimodal assessment*, the extent to which the teacher employed multimodal assessment in the classroom.

Table 16: Programme Intensity Observation Instrument

Question	----------SCALES----------			
1) Was the School part of the CAP pilot?	YES	NO		
2) Is the CAP currently in operation at the school?	YES	NO	SOME-WHAT	
3) Overall, how successful was this school in implementing the CAP	0-25%	26-50%	51-75%	76-100%
4) Approximately, in what percentage of classrooms was the CAP practiced?	0-25%	26-50%	51-75%	76-100%
5) Approximately, in what percentage of classrooms is the CAP still practiced?	0-25%	26-50%	51-75%	76-100%
6) What percentages of Commutative Record Cards are completed regularly?	0-25%	26-50%	51-75%	76-100%
7) To what extent does the school still train teachers for the CAP?	Never	Biennially	Annually	By each term
8) Currently, are there structures and resources in the schools dedicated to the CAP (Cupboards, manuals, etc)	None	A few	Some	Many
9) How much money in the school's budget is still allocated to the CAP process?	0-5%	6-10%	10-15%	>16%
10) To what extent is the data from CAP now used by the school in making decisions about students and teaching-learning?	Never	Some-times	Often	A great deal
11) Overall, how will you describe the level of implementation of CAP in this school	Very Poor	So-So	Relatively High	Superior to most

Measuring & Analyzing Change

The CBAM model guided the measurement and assessment of change across sites and Phases (George, Hall, & Steigelbauer, 2006; Hord & Hall, 2010). The Stages of Concerns questionnaire contained 35 questions organized into 6 categories: (1) Awareness (Unconcerned), (2) Informational, (3) Personal, (4) Management, (5) Consequence, (6) Collaboration, and (7) Refocusing. Awareness concerns means that teachers are not concerned about the change. Informational means that they would like to know about it. Personal means that they want to learn the personal ramifications of the change process, management focuses upon the processes and tasks of the innovation, consequence considers the impact on students, collaboration emphasizes cooperating with teachers, and refocusing is about benefits and alternatives.

Both the quantitative and qualitative versions of the Levels of Use (LOU) questionnaire were also employed. In qualitative version of the LOU, teachers were asked to report on their current levels of use. However, this data was not coded using the LOU framework but subjected to thematic analysis.

As shown in figure 33, both user and systemic models of change were embedded within the quantitative model. Several user levels variables were considered along with key organizational variables such as professional learning community, organizational innovativeness and readiness for change. Modelling would make clear the relative impact of the variables and the variable categories.

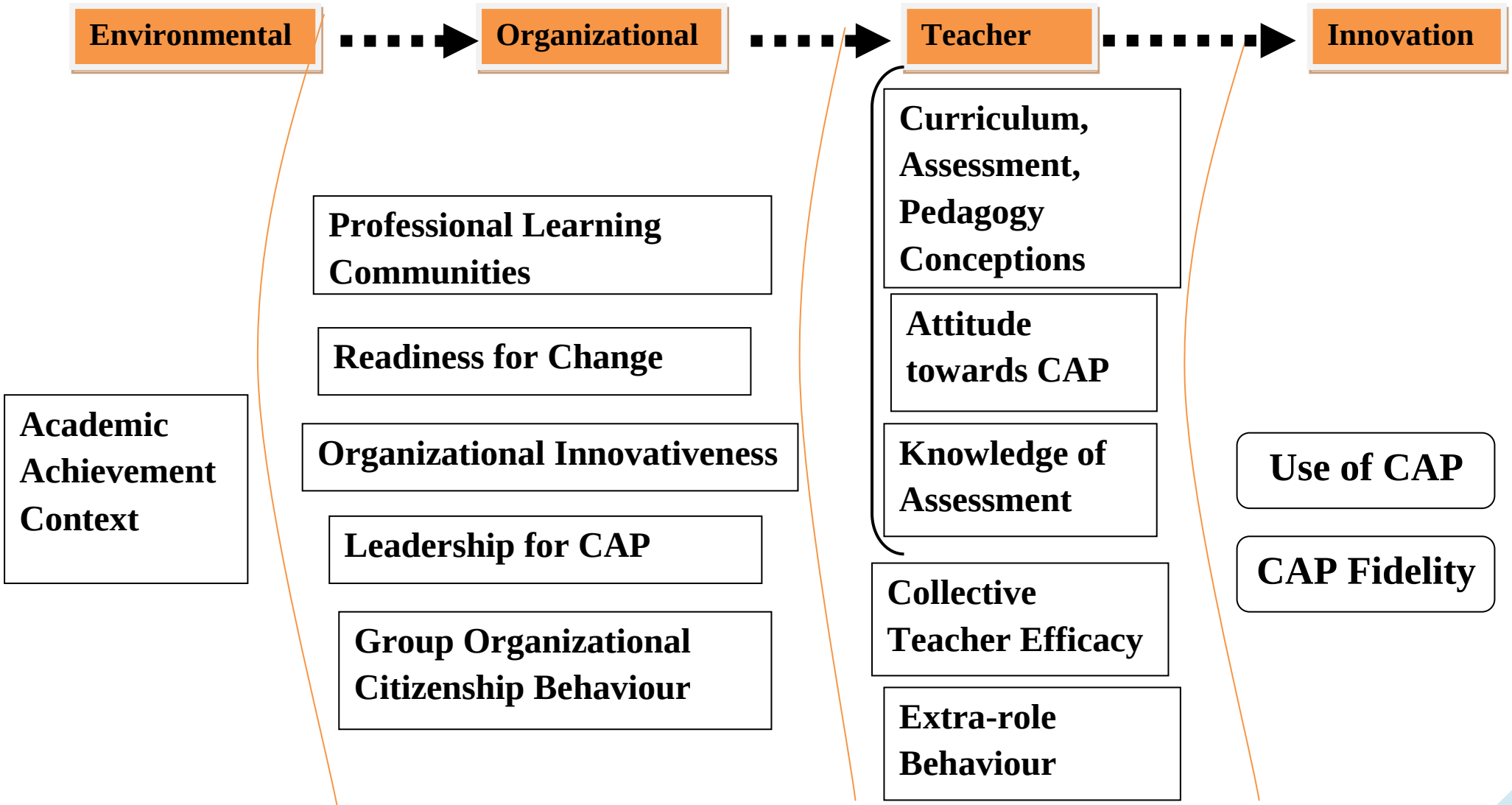


Figure 35: Modelling user and systemic models of assessment change for the Trinidad and Tobago CAP

Analyses & Presentation

In the analysis of the mixed methods research evidence, both quantitative and qualitative findings are presented along with metainferences (combination of the quantitative and qualitative findings). For Phase 2 of the study, statistical analyses were used to generate descriptive, correlation, and inferential data. Statistical analyses included multiple regression used to build the causal models and ANOVA for comparative analysis. Logistic regression and canonical correlation were also used to construct explanations of the theoretical model. Table 17 provides the different statistical analyses conducted, the more common statistics provided and the purpose of each method.

Table 17: Statistical analyses, methods and purposes used in Phase 2

Statistical Analysis	Methods & Statistics	Purpose
Descriptive	Means, Medians, and percentages. Crosstabulations with Chi-square statistic	To describe the schools and individuals in the sample in terms of demographic, antecedent, or outcome variables.
Descriptive-Graphical	Boxplots, Graphs and Barcharts	To illustrate distribution for different cases and variables
Correlational Analysis	Pearson zero-ratio, semipartial and partial correlation coefficients	To describe the relationships between variables
ANOVA	P-value of F-test; post-hoc analysis using Scheffe's multiple comparison test Effect size measure is eta-square.	To compare differences between groups
OLS Regression analysis	Beta, coefficient of determination	To compare the influence of multiple antecedents and the usefulness of different models
Logistic Regression	Odds and Odds ratio	To determine the relative impact of selected demographic variable
Canonical Regression	Beta Set correlation	To determine the relationship between the set of independent variables and the set of CAP outcomes

For Phases 1 and 3, thematic analysis was the main analytical tool. Qualitative data from the interviews and video ethnography were transcribed or reviewed and then analysed for both themes and sub-themes. The themes arose from the literature and in the case of the multi-site studies comparative analysis was also used. A qualitative judgement was made on the importance of each theme by judging both intensity and frequency; however, quantitized data was not included in the narratives constructed.

Based on the overall mixed method strategy, the meta-inferences were critical for completely understanding the nature of the evaluand and the conclusions ultimately made. Meta-

inferences were generated through multiple processes, such as linking, elaboration, and comparison (Teddlie & Takashorri, 2009). Integrating quantitative and qualitative findings is not equivalent to consolidation because some inferences are likely to be conflicting or contradictory. Thus, convergence, corroboration and consistency were not the only targets, but divergence and dissonance were considered equally important. Dissonant findings from different methodologies may point to different perspectives in the complex intervention. However, in practice and following Creswell and Plano-Clark (2010), we did try to resolve any discrepancies between the methodologies, but our philosophy remains that qualitative data provides both a more insightful and sometimes different picture of complex phenomena.

Theory Building & Causal Explanations

Data from both the quantitative and the multisite qualitative studies in Phase 3 were used to generate causal models. From the realist worldview, qualitative multisite case studies can provide information on the *recurring connections* observed or understood by participants at the particular sites. This either provides hypotheses or corroborates the causal links in the quantitative model, as suggested by Maxwell (2004a; 2004b).

The Phase 1 and Phase 3 qualitative studies were used for different purposes. The Phase I qualitative study was exploratory but Phase 3 was explanatory. However, both were designed to identify recurring patterns and processes that might further be investigated or confirmed in the quantitative modelling study, with phase 3 providing deeper insight into the causal links generated in the quantitative survey. To accommodate the use of causal inferences in the qualitative, a critical realist stance was adopted as the primary mental model. Models in Phase 2 were based significant Beta coefficients. To determine the Beta, outcomes, teacher, and organization, several multiple regressions with different dependent and independent variables were conducted following the linear sequence of the model.

Ensuring quality

Bearing in mind the limited time frame, several strategies were put in place to ensure rigour. Triangulation is a strong element in the design of both the quantitative and qualitative phases. For the quantitative study, rigour was ensured by (1) using the entire population of teachers as the sample, (2) piloting the instrument and using standard multi-item instruments, and (3) taking steps to verify the constructed databases and input data. The latter was an important step but it increased the amount of time spent on constructing the databases.

For the qualitative phases, credibility and transferability were the main focal points of the quality criteria. To ensure credible and trustworthy findings, data collection involved several team members who were required to engage in debriefing on completion of the exercise. In data analysis, wherever possible, we tried to use more than two persons to code. The Phase 1 qualitative study also involved several member checks with selected principals invited to the presentation and allowed to comment on the findings. Thick descriptions are used throughout the narrative presentation, including entire conversations between multiple members of the focus group.

Implementation Challenges

The elaborate evaluation design was necessary to provide credible evidence as part of an evaluation of CAP, an intervention that is considered both complicated and complex. However, implementing the multiphase mixed method design within the agreed six month period (an extension of two months was requested) also presented several logistical and operational difficulties. Some logistical issues were resolved by working closely with the DERE. For example, the DERE was helpful in developing a distribution and collection system for the questionnaires. DERE also contacted schools when necessary for the return of the instruments. The School of Education, University of the West Indies, St. Augustine also expedited the printing of the questionnaires.

However, the DERE remains severely understaffed, so it was not always possible to manage the return of the questionnaires in a timely manner. There were also issues with the return route and competing projects occurring at the same time. Consequently, three schools were excluded from the analysis. These issues caused a delay in the production of the final report.

Figure 36: Timeline for Evaluation Study

Activity	JANUARY		FEBRUARY		MARCH		APRIL		MAY		JUNE		JULY	
Prepare Work Plan														
Select & Train Field Assistants														
Develop Survey Instruments														
Conduct Phase 1 qual														
Data Analysis qual														
Conduct Phase 2-Quan														
Data Analysis-Quan														
Conduct Phase 3-qual														
Prepare Mid-term report														
Prepare Draft Final Report														

Some of the implementation difficulties might have been due to the decision to increase the survey sample size in response to what was believed to be the apparent heterogeneity of the system. The size of the Phase 1 sample was definitely too large and saturation was possibly met after 6-8 cases; however, it was necessary to capture schools in different locations because the context did prove important. There was also difficulty in managing the tight sequencing of the Phases, which were interfaced at the results stage and therefore required that analyses and reporting be completed before proceeding to the next Phase. Thus for Phase 2 to proceed, it was necessary to process data from Phase 1 and for Phase 3 to begin, data from Phase 2 had to be first processed. The resolution of this problem was to use sample data in order to direct the proceeding Phase.

The survey instrument was very lengthy; consequently, it would have taken considerable effort to complete. Some participants chafed under the pressure. While the final data was useful for modelling, the low sample size reduced the validity of this phase of the study. Despite the heavy demands on resources and time, the qualitative components remained essential to the process of generating insightful and contextualized evidence. However, these

components were costly and heavily resource dependent requiring field assistants and transcription, which can be a daunting task. The field assistants who volunteered their time and energy without charge were outstanding, but it was obvious that high-quality interviewing and research skills, available at only the Masters' level were required. Fortunately, the great majority of field assistants had completed Masters' level training. In the closing Phases, only Masters Level field assistants were used. An analysis of the interview skills suggested that more attention must be paid to training as questioning skills were adequate but neither outstanding nor efficient.

PHASE I: Multisite Exploratory Case Studies

Qualitative Findings

Data from eleven of twelve sites in Phase 1 were reported on, with data from the elite interviews integrated. Observation, interview and documentary data suggested that there was great variation in the intensity or strength of CAP across the sites. Fidelity in formative assessment and data use were universally low. Salient issues related to both the intensity and fidelity of the CAP were indentified in several themes. These themes were used to construct the narrative, which made use of rich, thick descriptions that captured the voice of participants. The themes were (1) the ambiguity of the intervention, (2) the variability of implementation, (3) changing reactions over time, (4) record keeping, (5) projects and rubrics, (6) tensions, pressures and fears. The intensity and frequency of these themes in the Phase I study are provided in Table.

Table 18: Qualitative Themes-Frequency and Intensity

Theme	Frequency	Intensity
1) Ambiguity of the Intervention	Medium and Low implementation Schools⊙	High ●
2) Variability in Classrooms and across Schools	Medium and Low Implementation Schools⊙	High●
3) Changing reactions over time	Some schools◘	Low⊙
4) Record keeping as a duty	Most Schools⊙	High●
5) The poor quality of projects	All Schools●	High●
6) Assessment pressures, tensions, and fears	All Schools●	High ●

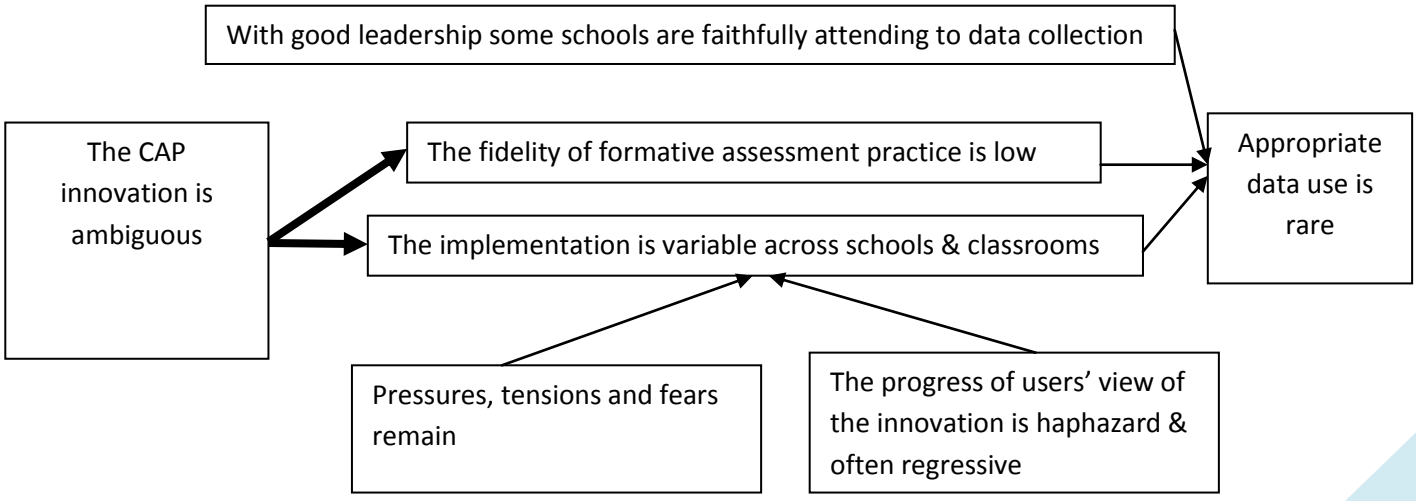


Figure 37: Relationship between main themes in interviews

Ambiguity and variability

The most notable feature of the field visits was the high variability in programme strength across sites, ranging from situations where the innovation had not been attempted at all to instances in which all the elements of CAP were still being practiced every day. The field notes of one investigator captured this variability in programme intensity even within a single site

> **To get a clearer picture of the experiences of the school [on] the CAP, the former Acting Principal, now Senior Teacher had to be called [in]. He remembered that at some time with a former principal the school had received some information related to CAP, [but] he did not know where to find the documents. He was not sure about whether the school [had been] a pilot school or not [although] he knew that teachers were given booklets of some sort but had no idea of exactly what would have been done with the booklets that they received.**

The problem at this particular site centred on the limited impact of early training and the high degree of staff turnover. This would have meant that even though some teachers had been trained during the early years of the intervention, lack of leadership and rapid changes in staff would have nullified the effect of that initial training. Consequently, the innovation had stalled, so to speak. The new principal of the same school was perceptive enough to predict the response of her staff to the questions about CAP:

> **Some on one hand would say, yes, they are sure that the CAP thing was in progress here, and some [are] completely oblivious to it. Some say they have no idea at all that anything like this was happening and some said [that] they vaguely remembered. And one or two who put their hands on the books said, yes, it had to have something for them to be able to get the book so it meant that the principal never really kept a meeting. I don't know who would have attended the workshop one or two teachers remember attending a workshop but not bringing it to the entire staff**

These issues were also present at all other sites reporting low programme strength. In some schools, several elements of CAP might be present, although the full benefits were usually still not received, often because of low fidelity. This situation contrasted with three sites in which the use of CAP was widespread with extensive data collection still in place. As illustrated in Figure 11, at these sites, each teacher was still in possession of CAP record books and they worked diligently at completing it. Of the three sites, this particular school was one of the largest in size, and yet most teachers were involved in the data collection elements of the process. Therefore, it did not appear that programme intensity was necessarily related to staff size.

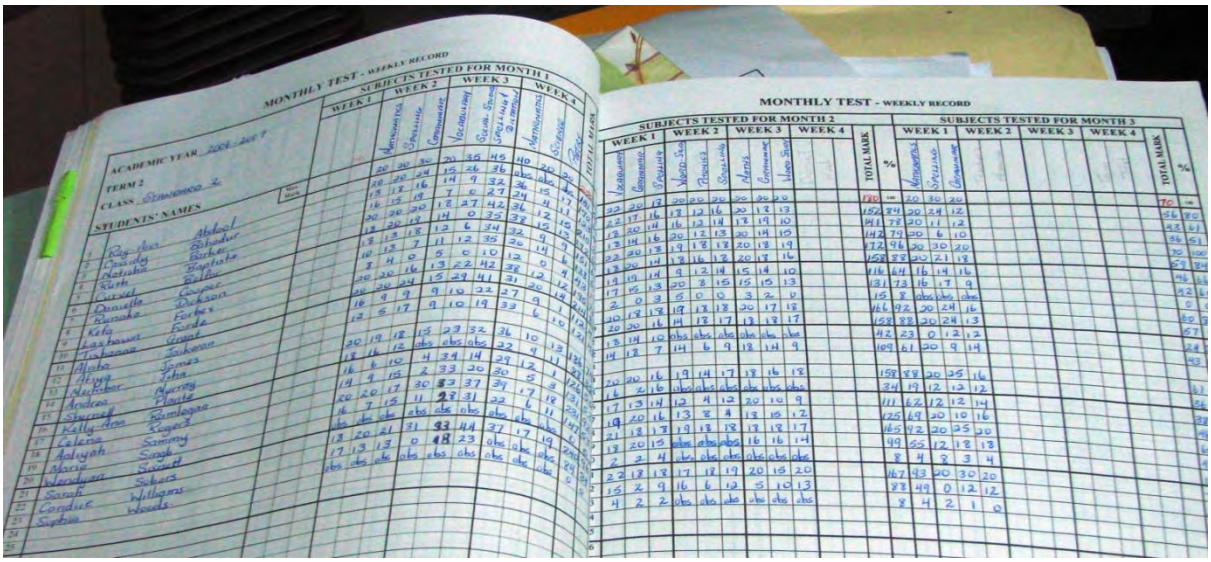


Figure 38: Teachers' CAP book still in use in Anstey Memorial Anglican

Other factors might have contributed to the variation across classrooms within a school. For example, if teachers did not fully understand the CAP and saw it only as a formal testing programme, then they resisted implementing the process in the early grades. The paradox is most apparent, because CAP was intended, according to the programme designer, to improve students' readiness for learning and therefore much of the impact would come from the use of data and application of early intervention.

> **I think it is a little strenuous for the standard 1 in particular, sitting down in a testing or classroom with testing situations. It is a little tough for them as they just came out of the infants. For the older children in Standard 2, it is not so bad. The time frame for testing itself, but the whole process, I think it's hard on the standard 1. And then when you look at the national tests, because we supervise for the national test, you can see the uneasiness in them because everything is geared towards examinations and tests. Yes, there might be benefits because you have your syllabus to complete, your scheme to complete. The workload is too much and then when you have to assess, it is too much.**

Here the teacher did not see assessment as embedded within instruction. Assessment was not used for formative purposes; rather, it was simply meant for "testing", a process which was separated from instruction. Even though teachers understood that multimodal assessment was intended, the process was regarded as a chore, unattached to learning. In such a situation some teachers naturally resisted the innovation.

Progression and Regression

At several sites, we explored in great detail the progress of the innovation in the minds, hearts and hands of teachers during the periods of adoption and implementation. Prior to adoption, there appeared to be much information about the intervention in various forums. However, this communication was not sustained over time, and so, an initial attraction or concern about the innovation might eventually turn into disinterest or even worse, disdain. This is evident

in the discussion between three teachers and the moderator in a focus group at a high achieving urban primary school.

> **[Teacher 1] It was 1998 . . . Teachers had a meeting at D'Abadie where I started to work. It was about CAP and there was a lot of confusion about it, but I had no interaction with it at that time because I had just come to the school. [Moderator] At that time, did you think it was a good thing or a bad thing? [Teacher 1] I wasn't sure what it was because at that time, I just started teaching. I was like a student teacher attached to [a regular classroom teacher]. . . . [Teacher 2] I really don't know anything about CAP to be honest with you. In the previous school, it was talked about in terms of what we had to do, but after that . . . [Moderator] What was your reaction to all the things you had to do? [Teacher 2] Well, it was now starting so you would be taken to any new ideas, so I think it was okay [Moderator] and what about now? [Teacher 2] Well, I have not been following up on it, so there is nothing now.**

Informal peer learning was common among teachers, so negative attitudes might be quickly learnt and adopted through emulation. Thus the lack of information about the innovation in non-pilot schools might have facilitated the development of negative attitudes, as expressed by the third teacher in the focus group:

> **[Teacher 3] Well I remember when the CAP was introduced; it was a very sore point among teachers. We were being told there is so much work involved, so much writing. It wasn't pleasing to us because of the amount of paperwork we had to do. At that time, I was in Tobago, a junior teacher working with the senior teachers and they didn't have a liking to it, so it was not practiced. When I came to Trinidad, I went to Morvant/ Laventille and they were very rigid at it. Once I got into it, I found the CAP is too much work and it doesn't make any sense.**

It appears, then, that the progress of the innovation at some sites could be haphazard. This pattern would be aided by the lack of support and inconsistent communication and training from the Ministry of Education. In these instances, it did not seem that the training arranged by the Ministry of Education had filtered down into and across the various sites.

The lack of leadership, professional development, and organizational support at most sites meant that the CAP innovation lacked sustainability, with new teachers coming into the system often lacking the knowledge, skills, or guidance to take up the challenge as one teacher from an urban boy's school noted

> **I am a second year teacher and I have come in the school hearing about it, but I have not been much involved in the running or planning. So my idea of the CAP is that it is an alternative assessment with the students doing portfolios and stuff like that. But to be in the class and see how things are done, no.**

The diligent record keepers

Surprisingly, even without the apparent lack of system-wide support, at some sites, several features of CAP had been sustained well into the present. This included project work and several aspects of record keeping, although the latter tended to be sustained only in schools with high quality supportive leadership. Schools such as Anstey Memorial and Arima Presbyterian were still diligently involved in data collection and record keeping and we were able to collect numerous artefacts to attest to this fact. Teachers at Curepe Anglican were also still diligently copying and using pages from the original student performance record books for student assessment data. To a large extent, this diligence defined the teacher commitment and high quality leadership within these schools. As we shall see, however, all of this might still not translate into good practice in data use and therefore much of the effect of CAP (which depends upon data use) was not achieved.

Several schools were engaged in conducting performance assessments, primarily by using projects (See Figure 39). This feature of CAP could be present even when data recording was absent. Indeed, apart from the very poorly performing sites, most institutions reported some use of projects. The principal of one school commented favourably on the impact of the projects:

> **[Interviewer/s]: So [are] you saying that they do use the assessment data from classroom teaching? [Interviewee]: Yes [Interviewer/s]: to inform their teaching? [Interviewee]: Yes [Interviewer/s]: do you see evidence of this do you see that it is really working? Or what sort of assessments they do? What do you normally look at it? [Interviewee]: I looked [and] I walked around and I saw the projects and I saw results from the projects-children, like they did it on their own and I was really amazed at the kinds of things that they did** [Principal of rural school]

Projects rule but . . .

It appeared that once meaningful and authentic tasks were set, projects might make a significant impact on the levels of engagement among students, even in schools with many students from economically disadvantaged communities. One teacher from a medium achieving school with whole school project work thus observed:

> **[Teacher 1] But all is not lost. For the summit the project participation in all the classes were tremendous. In the standards 4 and 5, we e really had good participation and I believe that when the parents don't help, its either they don't know or can't afford it. . . . [Teacher 2] I think that had to do with interest because things were so hyped up about the summit in the country, but if we gave them something that they were not interested in you would find that. [Teacher 3] culture, global warming . . . [teacher 1] But if we do the rich history (of the area the students were from) and we ask about the rich mixtures in their (multicultural populations), we still don't get anything.**

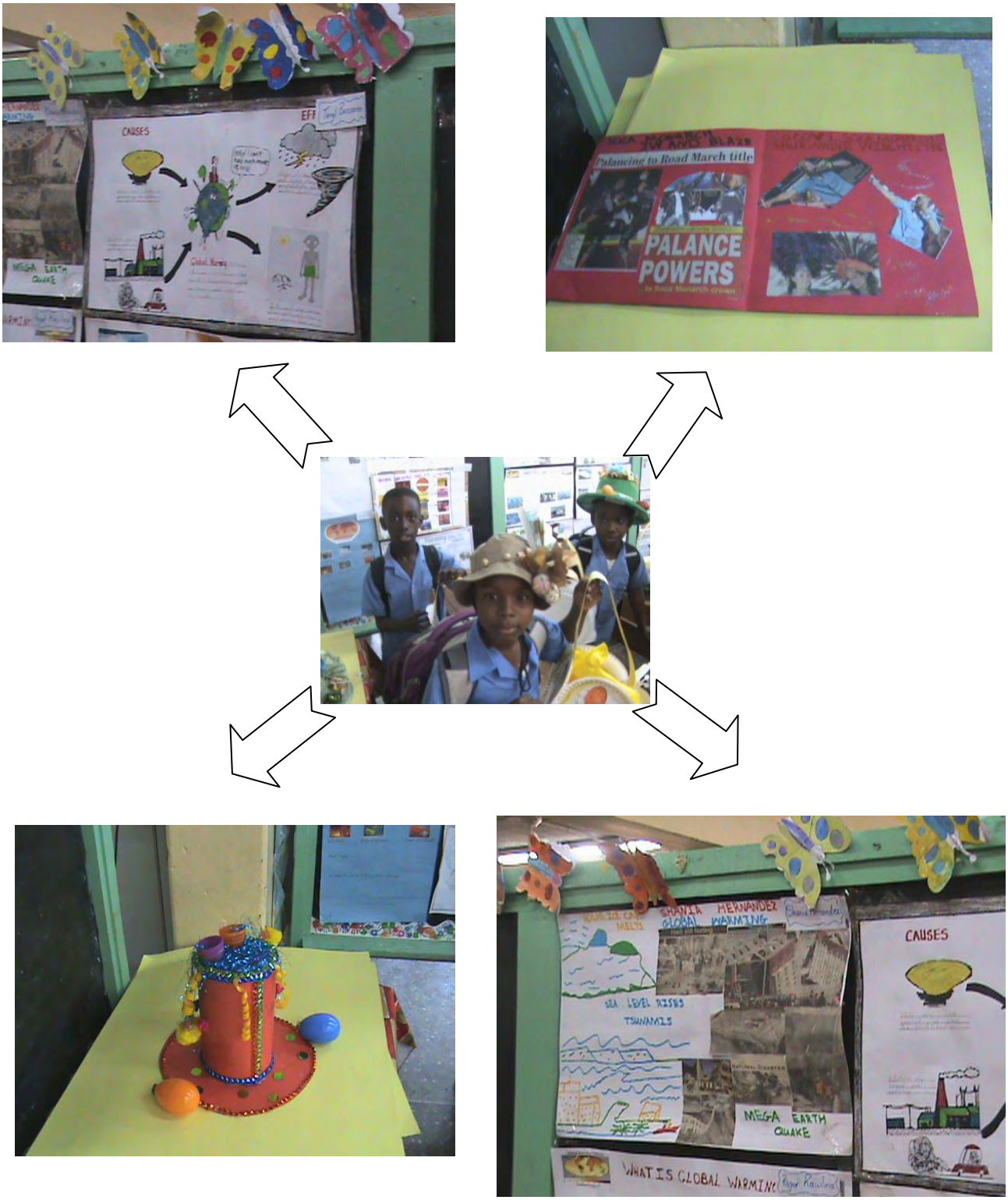


Figure 39: Term 2 projects completed at Curepe Anglican primary school (Term 2, 2009 - 2010)

Nevertheless, in most schools, the majority of projects were simply "take home" assignments. This meant that teachers were sometimes unsure of the level of involvement of the student and opportunities for feedback were very much reduced. The absence of teacher, peer, and self feedback would likely impact strongly on the quality of the outcomes.

In schools serving disadvantaged populations, teachers were very much concerned about the lack of viability in projects because they believed that they had insufficient support from parents. Some teachers also complained bitterly about the lack of school resources and the need for teachers to go beyond the boundaries of their role in financing and supporting student project work. Thus, one teacher reported on her own method of doing the projects:

> **With my second year infants, I give them guidelines to go home and do their project and, I mean, I continually guide, talk to the children and discuss it. I ask them to go home and sit with their parents and do it and at the end of the week on the time I have given to bring it in, half of them would do it. So there we have a lack of parental support.**

Some teachers were very innovative and had developed several strategies for countering the lack of support at home and the lack of infrastructure in the school.

The other side of this problem of parental involvement in the projects was present at high achieving schools, but now with middle class parents being overly keen to actually do the project for the child. Thus, one teacher in a high achieving school observed:

> **What I also found in agreement with Miss was that the parents played a lot into the project, helping the children. But you find that most of them would go on the computer, get the information, print it, and bring it out so that it became a competition for these projects.**

We arranged for the observation of several projects at different sites and interviewed several students at one school site. These interviews confirmed that much of the value of the projects was lost by insisting that it be done as a homework assignment. Where parental involvement was extensive or inappropriate, there appeared to be little learning or enhancement of student skills. The focus on take home projects also limited the use of peer learning as the following conversation with a male and female student indicated:

> **[Moderator] What about your other friends in class? Did you discuss your project with your other friends or was it a project that you did by yourself? [Female Student] Well, when we got back the projects, miss told us we could discuss the projects . . . [Moderator] When you got it back or when you started . . . [Female student] When you got it back. [Moderator] When you were in the process of doing the project . . . how long did you take to do the project? [Female Student] Well, I think it was since February or January, but I started mine early in March or February, then we got to complete it. [Male student] It was about March 12th . . . . [Moderator] But when you started the project, did you talk to your friends or did you just do it on your own? [Male student] We did it on our own because you [are] not supposed to**

> **tell anybody because they might choose the same person [do the same project]. You are not supposed to tell any other person.**

The projects were therefore divorced from regular teaching and learning in the school and were relegated to take home assignments in which more value was given to the product rather than to the process. Thus, although engaging in the project should have allowed the students to be fully engaged in doing and learning significant tasks, the process as enacted here tended to de-emphasise the critical features that would help all students learn from the performance assessments. As a result, the value of collaborative learning and peer feedback was often minimized and instead a competitive, individualistic environment was imposed.

A further obstacle to a project designed for formative use in the classroom was the limited resources in the schools to support students engaged in an authentic project. As such, although it was true that students from economically deprived situations could not find the information at their homes, it was equally true that they could not find information in the school, many of which had under resourced libraries and non-functional multimedia centres. Schools, then, tended to rely heavily on teachers to supply both media and resources, which was unacceptable.

Is this really formative assessment?

None of the interviews or classrooms observations provided evidence that teachers were very clear about the use and role of formative assessment and feedback. In practice, some teachers tended to narrowly separate the conduct of weekly and monthly tests from the projects and the record keeping. For example, one female teacher at a high achieving school with high levels of CAP implementation indicated that CAP activity interfered with her efforts at checking for understanding and improving learning.

> **[The CAP] places you under a certain amount of pressure because you have to conduct either monthly or weekly tests. You have to get these corrected, you have to put in marks and calculate the percentages, all of this when you are teaching the syllabus. So it puts you under a certain amount of pressure to get things done, even if the children don't understand a topic properly and you have to spend some more time on it before you test it.**

Thus, in this teacher's mind, the assessment activities in CAP were not to be integrated with instruction, contrary to the programme designer's intentions and current definitions of formative assessment. In the minds of many teachers, classroom assessment meant only testing at the end of the teaching-learning cycle to determine if the child had learnt (assessment of learning), rather than a system of "checking for understanding" as would be expected of true formative assessment. Perhaps, these elements were not completely clear either in the CAP documentation and the design of the reporting materials, including the student performance record book, and so the misunderstanding and misapplication might have became more increased because of this.

The research team did discover instances of very good practice in terms of formative assessment or remediation, although these were not necessarily related to the CAP process. In one school, for instance, the teachers in the early years had decided to split a struggling class

into two and one of the younger teachers took the smaller group of students with many experiencing difficulties and worked intensely with them in reading using the phonics approach. This was for a term before joining them back to the main group. The principal was very supportive of the idea and her good leadership would have facilitated the experimentation in a school where many clients were economically disadvantaged.

Rubrics, no, but where does the data go?

The use of rubrics as scoring devices was rarely mentioned or observed in use by respondents, even when the team arranged specifically to return for focused visits and therefore, it might be concluded that the overall quality and value of scoring was generally poor. Teachers appeared to be closely following the outdated scoring methods suggested in the CAP performance record books, which involved focusing upon discrete scoring criteria rather than full rubrics. Likewise, there was little evidence of the use student-involved assessment, such as student-generated rubrics. There was also was little understanding of how rubrics could help crystallize the teachers' standards and expectations for the assessment task.

It appeared too that most teachers did not truly understand the critical elements in classroom formative assessment connected to improving learning. These would include not only the use of rubrics to provide guidance on standards but also high quality feedback from teachers, self and peers. Thus, in a sense, despite the great efforts of some teachers, the projects were done mechanically at most sites. One student described the restrained role of her teacher and parents in developing her project.

> **[Moderator] What about Miss, how did Miss help you with the project? [Male student] Well, she just gave us a list of what we had to get. We went home, type it up, and get all the information. [Moderator] How did you remember miss helping you? [Female student] Well, I remember she said do it in a booklet form, but not many of the students had it in a booklet form, only a few. . . . [Moderator] What about while you were doing the project, did Miss ask you how far you had reached? Did you have to bring in what you had? Did you do things in school related to the project and did you put in at home? Or was it a project you did at home all the time? [Both students in unison] Home all the time.**

As highlighted before, several schools had efficient record keeping for the Cumulative Record cards and student performance records; however, there was little evidence of appropriate use of this data, collaborative or otherwise. The record books were often returned to and kept in the Principal's office and so the data was rarely employed in making critical decisions in the school. Thus, implementation of the CAP seemed heavy on data collection but extremely light on data use. Consequently, there was evidence that much of the data remained unused at the different sites. As one teacher at a high achieving school asked rhetorically:

> **How many of us go back into the CAP results and use it to monitor what our students are doing? I think that sometimes the CAP [results] remain in the office, placed in some draws to get dusty. We don't really go back into the results and see how the child is progressing . . .**

Poor data use practices were coupled with inattention to systems of action and protocols that could lead to the correction of identified problems. Thus, one male teacher in a boys' school expressed the view that CAP had been underutilized as a tool for reforming the system:

> **But I think the CAP has been there, [though] I think it has been underutilized in most cases because the follow-up is not there and there are other systems to put in place for CAP to work properly. Because when you deal with children and you find out their weaknesses, then how are you going to deal with remediation? Because some of the problems we encounter here, ordinarily, the teacher cannot deal with such problems. Because with CAP, must come a series of other programmes to complement the working of CAP, and I don't think that part has been put in place as yet. Because when you continue to diagnose a child as being weak, you continuously assessing this child in a whole variety of ways . . . things have to be put in place for those children after you use CAP to discern who is strong and who is weak.**

It is noticeable that the teacher from this low achieving school did not appear to own the problem of remediation, which is the basis of all remediation and further referrals to the diagnostic centre. To the teacher, remediation was something done to the child by some external agency so that although data was to be collected by the teacher, action was to be taken by some external agents. This lack of ownership is worrying and debilitating for the system.

Many teachers, however, did see the value of the CAP activity, once such appropriate data use was emphasized. Therefore, the problem of low implementation strength and teacher resistance might be connected to the lack of clarity in purpose and direction associated with the intervention at the start, as one male teacher from a high achieving primary school indicated:

> **It was really like extra work. Now, at present, I still feel it is extra work. If I had a secretary, they could help me record and these things, but there is some use if you go back and look at the marks, where you could show parents the child's performance, that's the only thing I could see about it.**

In another mixed suburban school, which had a significant economically disadvantaged population, and in which good practices of formative remediation were discovered, a young female teacher noted the futility of CAP as data collection tool only.

> **We have the continuous assessment book, but what is going to be done? Because I think if we have this book, you should have people coming in and testing if child is functioning at a particular level for two or three terms or for a particular school year. And they are to intervene, but you don't see that. So you recording, but what is happening? These children are going to remain in the system and they are going to pass through and nothing is being done. I personally would like to see them use this. If you are going to introduce something, introduce it and be consistent with it, and let something come out of it. Let the children benefit from it rather than simply jotting**

> **things down because to me that is wasting my time because what evaluation are you going to do after.**

Thus although teachers in this school were faithfully collecting data, the elements of data-driven instruction and the structures for referral and intervention remained absent. Perhaps, then, teachers understood that in the end not even the Ministry of Education would use the marks collected and the Cumulative Record Cards would not be passed on to the Secondary School. Moreover, the flawed design of the pupil performance record even made using marks across classes to record progress within the primary school very difficult. Thus, for teachers, it might have been hard to rationalize all of the extra work involved in CAP record keeping.

Assessment as burden

At sites with low levels of record keeping, teachers often complained about the utility and excessive demands of the process. The demands of record keeping by the external authority (The Ministry of Education) might interfere with the formative aspect of assessment, as one teacher noted:

> **Most of the time you cannot give a term mark without giving a project and most of the time you give the project, the parents would do it so that the children would not get much out of it. And the amount of writing you have to do, it sort of takes away from the programme because to me it was more writing than [helping the students].**

The programme designer explained that one of the obstacles to the early implementation of the innovation was the request for personnel at school sites to help with data recording. The initial delay in filling this request may have contributed to the increasing level of resistance among key stakeholders and a politicizing of the process by the Teachers' Union.

In this scenario, technology might have been a valuable tool but in reality, it was rarely available at the different school sites to ease the burden of this assessment. Indeed, the lack of technology proved to be a problem even at central administration, as the programme lead planner had observed.

> **And one of the things that was a hindrance that was resident in the DERE was getting the data from the reports because we started off with a pilot but we weren't giving back the reports from the data. We were collecting the data but within the DERE, we didn't have the technology to spin it off briskly for us and we were depending upon doing all of this manually and that was a colossal challenge. But still, even when we compromised that and we came up with a little, we still weren't writing reports and part of the thing was that we hadn't the competence within the DERE to do it.**

Another important contributory factor to the burden of assessment was the pervasive lack of training. One teacher described her introduction to CAP in this way:

> **In 1998, I first heard about the CAP from the principal and then to even now it has been a "snicker" to teachers because as {my colleague} said, it is a long**

> **and tedious process. We were never sensitized to CAP. We were just given a booklet to follow things through and work it out.**

Clearly, then, despite the extensive marketing of the programme, sufficient information did not filter through at all levels and this was an important factor in CAP remaining as a burden to teachers.

Tensions and fears

When CAP was introduced, there was some ambiguity about the role of the intervention as well as the extent of demands on teachers. At some sites, however, with use, many teachers became more comfortable with the intervention; however, the ambiguity might have persisted. The tension between classroom assessment and high stakes examinations was an impediment that loomed large in the minds of teachers even though there had been little confirmation of any official position on the use of CAP scores for high stakes selection purposes. Some teachers may have begun to value school-based continuous assessment only for its possible role in the high stakes selection process, as one female teacher from an Urban Boy's school observed:

> **I thought it was going to be used to assist the children later on in the SEA. There are some children who can't perform well in the SEA. So I thought that there was going to be a continuous recording of their results and when they reached that level [SEA], we could see what the problem is and why they are not performing and then we would use the CAP marks to assist them with the results they attained in the SEA.**

The distorting effects of the high stakes assessment is apparent even here because the teacher values the CAP not for providing corrective data that can inform teaching and learning, but because it might assist in high stakes selection. Thus, she speaks of remediation only in the future, at the stage when the student would write the Eleven Plus. Such thinking might reflect a de-emphasis on early remediation, the key for ensuring students' readiness to learn.

Some teachers, however, expressed concerned about the possible use of CAP to supplement scores in the high stakes SEA used for selection. They saw it as an issue of fairness, as one teacher from a medium achievement level school observed:

> **[Teacher] I have a question, how soon will CAP be used as part of the examination in the end of standard 5 . . . You know they said, they will use part of it? . . . You know CAP is good you know, but I have a concern with assessment that is coming from within a school and going towards a final examination. I think we will have a bit too much subjectivity. Since we are so examination oriented, I a bit concerned . . . how would the CAP work when they bring it around in the final examinations? . . . [Moderator] So you have some fears and anxieties. Is this shared by others? [Teacher 2] And not only that -- Children grow and different rates, intellectually as well. And sometimes you might have a child who has not been performing at the lower level and for some reason or the other as soon as they reach Standards 4 and 5, they start to perform and they start to do well - how do you rate that? . . .**

> **[Teacher 1} What I am looking at here is that our recording of the marks for this school might differ to Curepe Presbyterian or whatever and our marks, we may test at a lower level and our students may have the same ninety-something as another child there [Curepe Presbyterian]. Something is not meshing there because their assessment will be different from ours [Group] it's not standardized. . . . [Teacher 3] even standardization in the school is difficult because they telling you that you have to differentiate.**

The programme designer in her interview had suggested that it was the intention for CAP to have a role in the high stakes selection and transition between primary and secondary school:

> **[Programme Designer] Because remember too that part of this thing was supposed to lead to another approach to facilitate the transition from the primary to the secondary school. [Interviewer] So in your mind, you saw it too as a possible route leading to the high stakes examination? [Programme Designer] Yes, yes . . . It was more of the social issues that would get in the way of implementation. That people were so concerned about having their school, even at the primary level, of being a prestige school, and having their school seen as the best school, so they thought that people would cheat, so that we weren't trusting the teachers to submit fair marks.**

Such a dual purpose for CAP would further accentuate the tensions and ambiguities in the mind of teachers. Both high stakes selection and low stakes formative practice functions are not independent, but are intimately intertwined.

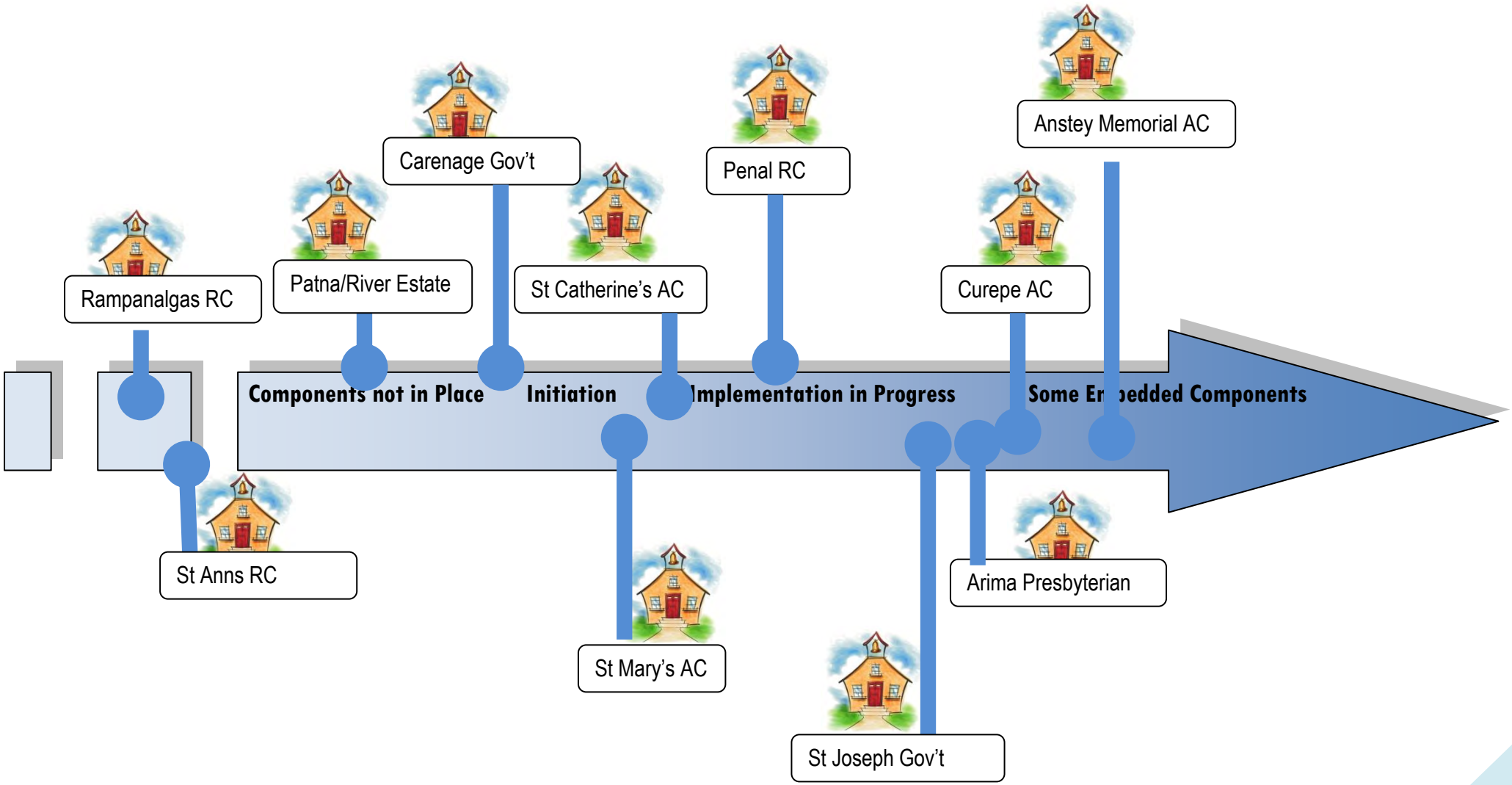
Quantifying implementation status

Field assistants observed classrooms in addition to conducting interviews. The completed checklists, field notes, and video were used to create this quantitative-type data to compare the schools in terms of the intensity and fidelity of implementation as judged by the trained raters. A sample of a completed checklist is found in the Appendix 3. As shown in Table 19, the intensity of the programme varied from moderate to almost nil on the observation scale. The classroom checklist suggested that not all elements of CAP were practiced in all schools although some schools were regarded as practicing moderately high levels of authentic formative assessment practice. Figure 40 supports this analysis by locating the schools on a map of implementation space. Only two schools may be regarded as having partially embedded components (routinization). Variability in activity among different teachers was also noticeable and the depth of the line within the progress arrow represents this.

Table 19: Programme strength and fidelity judged by field researchers

School	Observed level of implementation (7-46)	Classroom Checklist (0-40)	Formative Assessment Practice (6-30)	Observer's Comments
1. Anstey Memorial Girls' A C	28	NA	NA	Poor understanding of CAP. Limited monitoring and evaluation
2. Arima Presbyterian	28	22	20	Relatively high level of integration of innovation into work life. Good attitude towards CAP
3. Holy Saviour Curepe Anglican	23	18	25	High quality monitoring but variable understanding and application. High quality leadership.
4. St Joseph Government	23	11	21	High quality leadership with significant monitoring, but variable application in the classroom. Instances of good practice in formative assessment
5. Penal R C (St Dominic's)	23	16	14	Process not fully understood. No linkages with teaching and learning. Absence of support systems.
6. St Mary's Anglican	19	12	15	Present to some extent, but little documentation or formalization
7. St Catherine Girls' Anglican	17	NA	NA	Poor collaborative climate and adhoc record keeping limits sustainability
8. Carenage Boys' Government	15			Support systems deficient. High staff turnover limits sustainability
9. Patna/River Estate Government	14	17	16	Resistance to organizational changes. Division among staff.
10. Rampanalgas R C	10	11	9	CAP never adopted or practiced. Absent in the majority of classrooms.
11. St Ann's R C	10	7	15	CAP never really adopted or practiced. Artifacts stored and not in use.

Figure 40: Position of schools in implementation space



Hypothesis Developed from the Phase 1 Exploratory study

Several hypotheses were developed based upon the analysis of the data collected in the exploratory Phase 1 study. These included hypotheses already stated implicitly in the model presented and new hypotheses based on the interview data. Some of these hypotheses are listed in Box 8. It appeared on the surface that contextual and institutional factors such as pilot and academic achievement status might be powerful determinants of CAP use. We also expected fidelity and intensity measures to be different, explained by a different set of factors. Training was considered important and a measure of CAP training was developed for the survey study

---------Box 8: Sample Hypotheses generated from Phase 1 ----------

Hypothesis 1: There is a no relationship between pilot school status and the outcomes and quality of implementation of the CAP

Hypothesis 2: There is a no relationship between leadership by the principal and CAP implementation intensity and fidelity

Hypothesis 3: There is no relationship between school achievement status and CAP implementation intensity and fidelity

Hypothesis 4: There is no relationship between the quality of training and CAP implementation intensity and fidelity.

Quantitative Findings

Descriptive Analysis

Data for the quantitative analysis was based on 378 teachers in 36 schools (Appendix 1). 13 schools were categorized as original pilot schools and 22 as non-pilot. One school was not categorized and data from 3 schools came in too late for inclusion. The sample included 255 teachers in the non-pilot schools and 120 in the pilot schools. Table 17 provides the important descriptive statistics for each group of schools.

Table 20: Demographic data for teachers in pilot and non-pilot schools (final sample)

	---------- % ----------	
Variable	Pilot (120)	Non-Pilot (255)
Age		
• <30	8.6	18.3
• 31-50	71.5	65.3
• >51	19.8	16.4
Tenure		
• <5	3.4	12.6
• 5-14	33.0	37.2
• 15-24	38.9	27.7
• >25	24.6	22.5
Tenure in school		
• >5	18.6	30.7
• 15-24	70.3	63.4
• >25	11.0	5.9
Highest qualifications		
• A-Levels	7.7	11.7
• Academic Undergraduate Degree	21.4	21.0
• Training College	79.6	72.1
• B.Ed	18.3	21.6
• M.Ed.	0.9	0.4
Number in Class		
• <20	63.8	36.7
• 21-30	30.2	57.0
• >31	6.1	6.4
CAP first experienced in current school	90.0	79.6
Received Formal Training	32.4	21.4
Major Service Provider		
• Ministry of Education	23.5	14.5
• School PD	4.7	6.6
• University	1.2	.8
Quality of Training		
• Very Effective	8.0	16.4
• Mostly Effective	21.1	29.5
• Somewhat Effective	52.6	37.7
• Minimally Effective	18.4	13.1
• Not Effective At All		3.3

As shown in Table 20, the population of teachers from the pilot and non-pilot schools appeared to be very different. Teachers in the non-pilot schools were younger (8.6 % under age 30 compared with 18.3% in the pilot schools), 12.6% of the teachers in the non-pilot schools were in the profession for less than five years and 30.7% of the teachers in the non-pilot schools had spent less than five years at the school. This compared with 18.6% of the teachers in the pilot schools have tenure of less than five years at the school they were currently in. Thus, it appeared that the pilot schools had older staff that had been retained in the institution for a longer time. This might prove a positive facilitator for the retention of the CAP innovation. Indeed 90% of the teachers in the pilot schools had their first experience with CAP at their current school.

Teachers in the non-pilot schools, however, were more qualified. For example, they were more likely to have A-Levels (11.7 compared with 7.7) and a B.Ed (21.6 compared with 18.3). Classes in the non-pilot schools were usually larger (57.0% with numbers of 21-30 compared with 30.2%). As expected, larger numbers had received training in CAP in the pilot schools. The predominant provider was the Ministry of Education. However, the majority of respondents at both the pilot and non-pilot schools found the training minimally or somewhat effective.

CAP use by Demographic Data

The Level of Use (LoU) scale is especially suitable for crosstabulation analysis. I used graphical figures to illustrate differences in the LoU for key demographic variables, namely age, tenure and professional qualifications. In the Modelling sections, the set of demographic variables as a predictor is considered in logistic regression.

Age and Levels of Use

Figure 41 shows the parentages in each age group for the different levels of use. The Chi-square for this distribution was 84.473 with a p-value of 0.37. As shown, levels 6 and 7, which involved advanced use, were rarely achieved, but were more likely to be achieved by older teachers. At these levels, teachers were either integrating CAP or re-evaluating the processes in terms of efficiency. The highest number of Level 7 users (Re-evaluating) was in the age group 56 to 60, however, the highest number of level 6 teachers were in the 26-30 year old age group. Indeed, 58% of these teachers were either in Levels 5 or 6. However, in the 20-25 age-group and younger, the majority of teachers were simply seeking to acquire more information about CAP. In each of the age groups, except those over 60, there were teachers at either Level 0 or Level 1 of usage.

Tenure and Levels of Use

Two measures of tenure was used, tenure in the profession and tenure in the school. Only the distribution for tenure in the profession was statistically significant (Chi-square=68.702; P= .006). The data is graphically shown in Figure 42. The data suggest that teachers with many years in the profession were more likely to be at Level 7, with 50% of the teachers with 5 years experience being at Level 1. The majority of users for teachers 25-29 (67%) and above 30 (58%) were at levels 5 to 7.

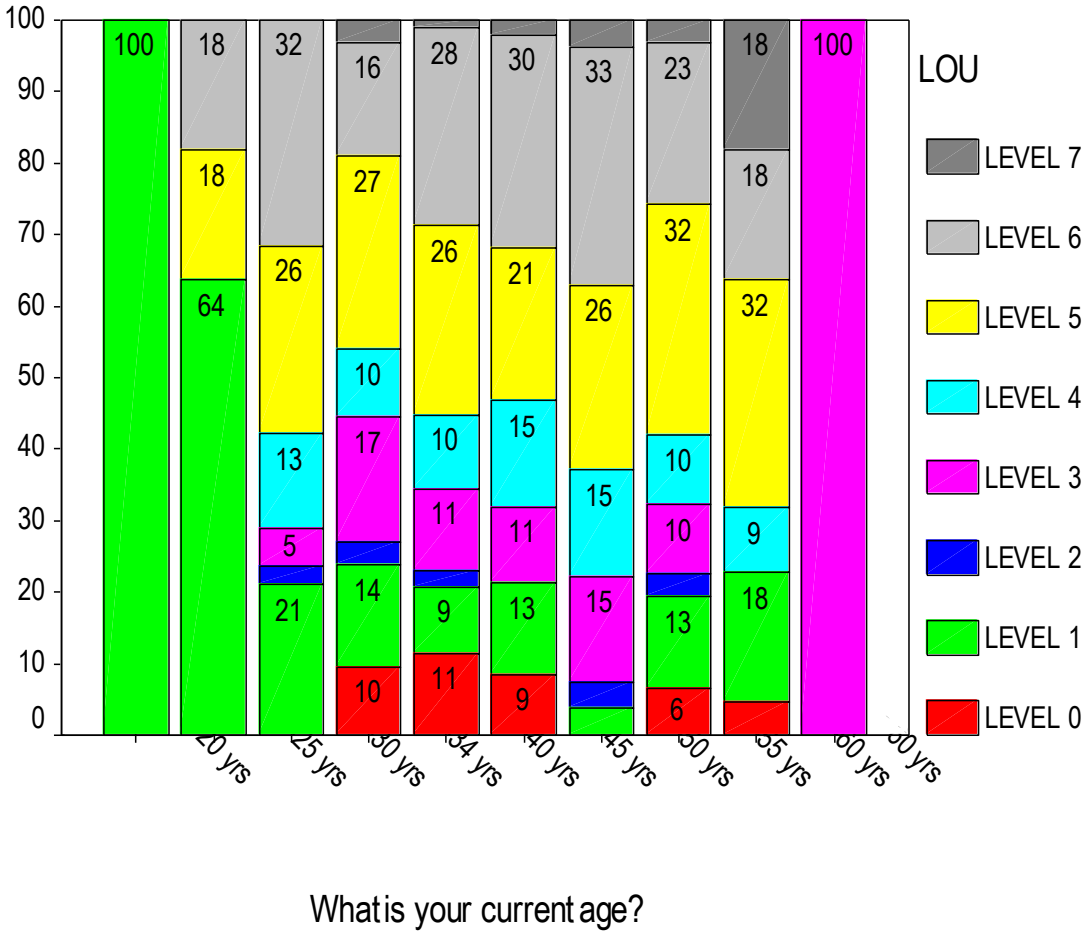
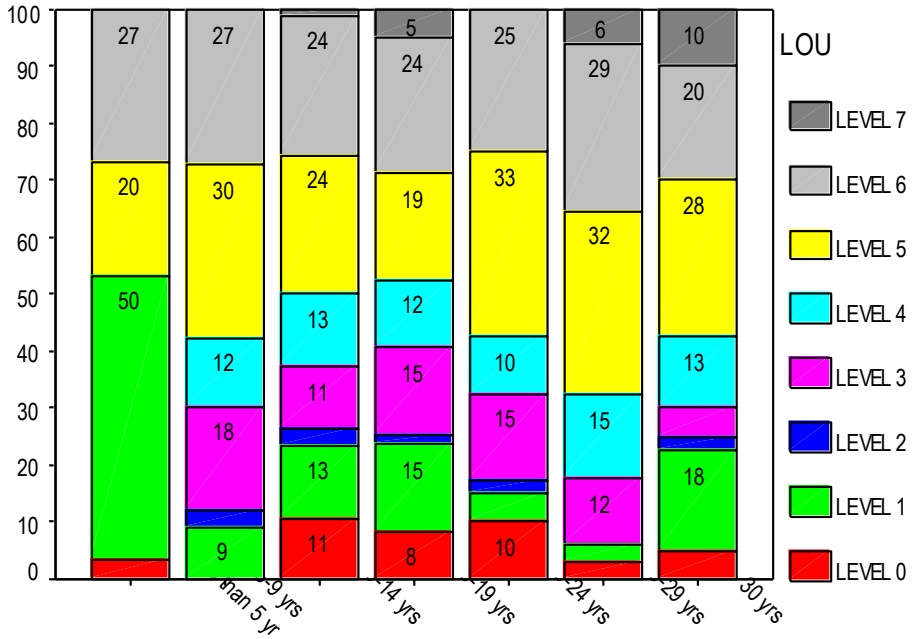


Figure 41: Graphical representation of crosstabulation between age and Levels of Use

Highest professional Qualifications and Levels of Use

This Chi-square statistic for this distribution was not statistically significant. As is evident from the 100% bar graph in Figure 43, teachers with training college certification were similarly distributed to teachers with Bachelors of Education Degrees. Indeed, the two teachers with Masters of Education Degrees in the sample were both at Levels 5 compared with 80% of the teachers with Certificates of Education at Levels 5 and 6.



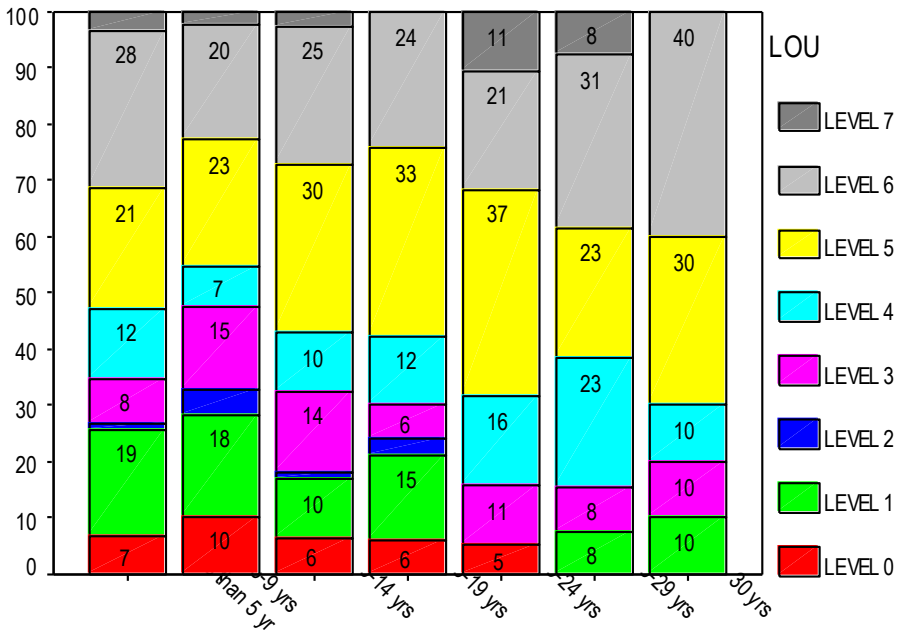


Figure 42: Graphical representation of crosstabulation between tenure in profession and school and Levels of Use

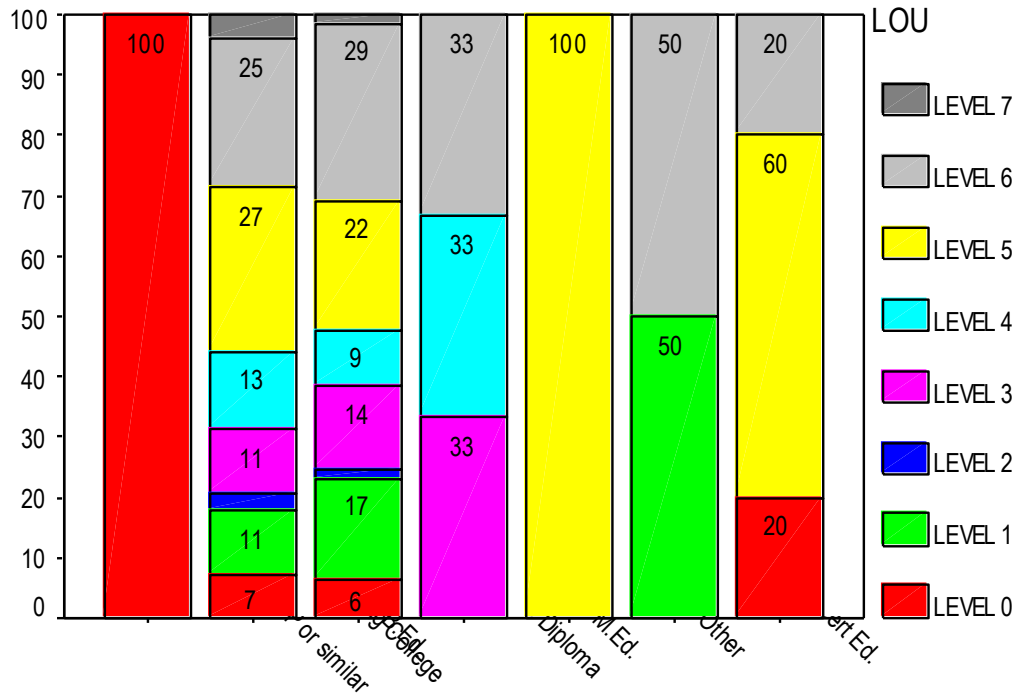


Figure 43: Graphical representation of crosstabulation between highest professional qualifications and Levels of Use

Measures of Fidelity & Intensity by School

Tables 21 and 22 list the organizational characteristics for the pilot and nonpilot schools, along with data for programme intensity and fidelity. The tables also provide triangulated data on the intensity or outcome of CAP by providing data on the students' perception of assessment and the principals' overall assessment of implementation intensity. For all scores, the means are provided along with the standard deviation in brackets. The higher the mean scores the greater the intensity of fidelity and the lower the standard deviations, the less the variability in scores.

For both pilot and nonpilot schools, there was no consistent pattern between the principal's assessments, students' perception of assessment, and the intensity of CAP. Neither was there a consistent pattern between high scores on intensity as measured by the levels of use and the sum of all CAP tasks and the single measure of fidelity, the amount of feedback provided by teachers. For the nonpilot schools, the majority of institutions with higher CAP scores were in either high or low achieving contexts. Schools scoring highly in the CAP use measure tended to report smaller standard deviations for feedback scores suggesting that there was low levels of variation in this task. For the nonpilot schools, student assessment perceptions were usually positive for these schools supporting the teachers' assessment of CAP use.

Table 21: School characteristics and implementation intensity in pilot schools of quantitative sample

School Characteristics			Measures of Strength & Fidelity			Triangulated data	
School	**Achievement Status**	**Staff size**	**Levels of Use**	**CAP Tasks**	**Student Feedback**	**Student Assessment**	**Administrator's Assessment**
Carapichaima ASJA	High	20	4.40 (1.68)	57.75 (7.30)	39.33 (3.51)	46.06 (2.68)	NA
Cocoyea Government	Medium	14	5.83 (0.40)	57.33 (9.35)	41.80 (1.09)	39.31 (6.93)	37.00
Rio Claro Presbyterian	High	18	4.66 (1.61)	55.61 (7.85)	35.66 (6.28)	46.04 (5.55)	24.00
Palmiste Government	High	12	3.20 (1.48)	54.71 (8.76)	39.28 (2.56)	38.41 (6.48)	21.00
Macaulay Government	Low	15	5.50 (1.00)	52.85 (12.64)	38.42 (2.93)	50.20 (8.16)	37.00
Fifth Company Baptist	Medium	21	3.81 (2.08)	52.58 (8.80)	35.50 (6.45)	44.86 (6.19)	30.00
Barrackpore Vedic	Low	6	4.33 (1.15)	51.66 (7.02)	37.00 (1.73)	47.00 (9.28)	34.00
Brighton A.C.	Low	17	1.00 ()	50.50 (7.77)	27.00 ()	49.70 (3.68)	30.00
Waterloo SDMS Hindu	Medium	7	3.28 (1.79)	50.14 (5.45)	36.57 (5.47)	47.66 (5.31)	28.00
North Trace Government	Medium	9	2.66 (2.25)	50.12 (4.05)	36.83 (6.01)	47.25 (9.03)	26.00
Guayaguayare R.C.	Medium	16	3.85 (1.83)	48.81 (8.51)	38.00 (5.26)	53.70 (5.60)	20.00
Ascension A.C.	Low	17	3.42 (1.81)	48.75 (8.90)	38.00 (3.02)	45.61 (5.28)	9.00
D'Abadie Government	Low	24	4.00 (2.58)	46.69 (12.27)	38.71 (8.03)	43.11 (5.92)	NA

Student Assessment Max-72; Administrator's Assessment Max-44

Table 22: School characteristics and implementation intensity in non-pilot schools of quantitative sample

School Characteristics			Measures of Strength & Fidelity			Triangulated data	
School	**Achievement Status**	**Staff size**	**Levels of Use**	**CAP Tasks**	**Student Feedback**	**Student Assessment**	**Administrator's Assessment**
Freeport Presbyterian	Medium	21	4.71 (1.89)	59.52 (7.34)	38.94 (3.56)	NA	17.00
El Dorado South Hindu	High	17	5.31 (0.87)	58.47 (5.20)	39.70 (3.36)	46.82 (2.46)	33.00
Felicity SDMS Hindu	High	22	4.50 (2.02)	58.33 (9.66)	38.80 (5.40)	39.30 (8.41)	29.00
San Fernando TML *	High	17	5.20 (0.77)	58.26 (6.21)	39.60 (2.77)	46.24 (6.00)	38.00
Harmony Hall Presbyterian	Medium	16	5.00 (1.15)	57.84 (8.91)	38.30 (4.13)	43.46 (4.91)	NA
Inverness Presbyterian	Medium	15	4.50 (1.58)	57.10 (9.85)	38.10 (5.38)	44.43 (5.89)	30.00
Sixth Company A.C.	Low	6	5.50 (0.57)	57.00 (18.01)	34.00 (14.73)	46.71 (6.18)	29.00
Rio Claro Vedic	Medium	8	4.00 (2.00)	56.83 (12.59)	39.16 (3.71)	43.54 (10.72)	35.00
Diego Martin Government	Medium	24	0.80 (1.30)	55.80 (6.37)	34.80 (3.49)	29.30 (6.46)	9.00
Bien Venue Presbyterian *	Low	19	4.14 (1.77)	53.57 (6.02)	39.28 (4.11)	44.50 (6.94)	32.00
Orange Field Hindu	Medium	20	3.22 (2.55)	53.11 (8.95)	38.12 (4.34)	40.57 (5.00)	7.00
Holy Saviour Curepe Anglican	High	24	4.23 (1.30)	52.64 (6.90)	36.29 (4.56)	44.65 (4.97)	9.00
La Puerta Government	High	17	1.20 (1.81)	52.00 (5.07)	34.09 (8.85)	40.20 (7.53)	9.00
Carenage Boys' Government	Low	11	4.00 (2.21)	50.90 (8.08)	40.90 (2.18)	NA	NA
Anstey Memorial Girls' A.C.	High	18	4.22 (2.04)	50.38 (9.69)	37.92 (6.06)	NA	
St. Joseph Government	Medium	21	3.25 (2.05)	50.15 (7.08)	43.36 (3.93)	44.57 (7.65)	NA
St. Mary's A.C.	Low	17	4.40 (2.06)	49.53 (12.79)	38.46 (3.13)	47.58 (3.71)	33.00
Unknown			5.66 (0.57)	48.66 (4.04)	38.66 (3.05)	45.02 (5.30)	9.00
Aranguez Hindu	High	9	3.50 (3.14)	46.66 (9.86)	33.66 (8.10)	45.16 (6.81)	23.00
Nelson St. Girls' R.C *.	High	19	1.87 (2.03)	46.33 (11.37)	39.33 (2.50)	43.18 (5.47)	34.00
Patna River Estate Government	Low	25	4.25 (1.25)	45.75 (2.62)	39.00 (2.16)	43.27 (8.18)	NA
Moriah Government	Medium	9	3.16 (2.31)	45.55 (10.50)	40.66 (3.32)	43.50 (7.11)	34.00
La Lune R.C.	Low	6	3.00 (2.09)	45.500 (7.63)	39.83 (3.97)	35.20 (2.98)	20.00
Mon Repos R.C.	Medium	22	1.50 (2.07)	43.28 (7.60)	31.33 (12.17)	42.05 (8.21)	9.00

Students' Assessment Max-72; Administrators' Assessment Max-44 * Schools selected for Phase 3.

Comparative Analysis

Table 23 provides scores on the different antecedents and measures of CAP for the pilot and non-pilot schools. Statistically significant differences were found for leadership of CAP, organizational citizenship behaviour, readiness for change and in the professional learning communities of the school. The size of the differences, however, suggested a small effect. These differences were not in the direction that was predicted. The non-pilot schools proved to have slightly better leadership, somewhat stronger group organizational citizenship behaviour, greater readiness to change and greater professional learning community. However, there were no statistically significant differences in Levels of Use, CAP use, and feedback.

Table 23: Differences in pilot and nonpilot schools for antecedents and outcomes

Variable	----------Means (Standard Deviations)----------			P-Value
	Pilot	Non-Pilot	Total	(Effect Size)*
LoU	4.04 (1.87)	3.91 (2.11)	3.95 (2.04)	0.618 (0.001)
CAP Use	52.21 (9.12)	53.09 (9.73)	52.81 (9.54)	0.409 (0.002)
FEEDBACK	37.86 (4.98)	38.25 (5.40)	38.14 (5.28)	0.539 (0.001)
KNOWLEDGE	6.63 (2.64)	6.54 (2.43)	6.57 (2.49)	0.781 (0.000)
CAP SKILL	6.26 (7.68)	7.84 (8.50)	7.36 (8.28)	0.099 (0.008)
CAP ATTITUDE	112.16 (22.29)	113.26 (21.82)	112.95 (21.93)	0.679 (0.001)
ASSESS TO IMPROVE	50.14 (9.41)	51.32 (9.05)	50.98 (9.16)	0.277 (0.003)
CTE	81.62 (10.64)	81.18 (10.54)	81.31 (10.55)	0.729 (0.000)
EXTRA ROLE	101.98 (19.74)	105.54 (18.90)	104.56 (19.18)	0.128 (0.007)
LEAD CAP	68.19 (19.52)	75.49 (19.47)	73.43 (19.73)	0.003 (0.028)
OCB	102.32 (21.26)	111.39 (20.25)	108.87 (20.90)	0.000 (0.038)
READINESS CHANGE	112.83 (21.64)	118.48 (22.97)	116.94 (22.72)	0.044 (0.012)
INNOVATE	58.54 (11.34)	59.54 (11.27)	59.23 (11.29)	0.443 (0.002)
PLC	19.55 (4.59)	21.12 (5.02)	20.67 (4.95)	0.014 (0.020)

*(0.01=small effect, 0.06=moderate effect, 0.14=large effect)

Table 24 provides differences between antecedents and outcomes for the schools classified according to achievement status as measured by performance in the national assessments of educational achievement in standards 1 and 3. As shown, there were significant differences for CAP use, CAP Attitude, Collective teacher efficacy, organizational citizenship behaviour, and organizational innovativeness. Surprisingly, the lower achieving schools reported more innovative behaviour and also reported higher collective teacher efficacy. Perhaps, the later finding may be explained by ego-defensive or compensatory behaviour. Low achieving schools also had a better attitude towards CAP. High achieving schools had greater organizational citizenship behaviour and greater CAP use. This complex pattern suggests that although achievement might be important, that influence was neither linear nor independent.

Table 24: Differences in antecedents and outcomes for schools at different levels of achievement

Variable	Low	Medium	High	Total	P-Value
LoU	4.11 (1.94)	3.72 (2.14)	4.11 (1.98)	3.95 (2.04)	.236 (.009)
CAP Use	49.88 (10.64)	52.95 (9.29)	54.28 (8.82)	52.81 (9.54)	.004 (.030)
FEEDBACK	38.59 (4.77)	38.32 (5.42)	37.75 (5.41)	38.14 (5.28)	.497 (.004)
KNOWLEDGE	6.60 (2.58)	6.54 (2.33)	6.58 (2.59)	6.57 (2.49)	.984 (.000)
CAP SKILL	7.77 (8.89)	6.16 (7.58)	8.34 (8.50)	7.36 (8.28)	.082 (.014)
CAP ATTITUDE	116.48 (23.25)	108.82 (20.97)	114.94 (21.67)	112.95 (21.93)	.024 (.022)
ASSESS TO IMPROVE	50.84 (8.09)	49.88 (9.69)	52.09 (9.10)	50.98 (9.16)	.136 (.012)
CTE	84.00 (10.60)	79.21 (10.58)	81.88 (10.18)	81.31 (10.55)	.006 (.030)
EXTRA ROLE	102.64 (21.38)	103.52 (18.44)	106.49 (18.60)	104.56 (19.18)	.291 (.007)
LEAD CAP	73.86 (19.10)	72.51 (18.30)	74.05 (21.32)	73.43 (19.73)	.805 (.001)
OCB	109.43 (20.11)	102.29 (20.17)	114.61 (20.35)	108.87 (20.90)	.000 (.069)
READINESS CHANGE	120.15 (18.61)	113.90 (22.42)	118.03 (24.64)	116.94 (22.72)	.141 (.012)
INNOVATE	62.16 (9.93)	59.29 (10.97)	57.65 (12.00)	59.23 (11.29)	.019 (.022)
PLC	19.79 (5.02)	20.71 (4.53)	21.14 (5.20)	20.67 (4.95)	.193 (.011)

Correlation Analysis

The magnitude and statistical significance of the Pearson correlation coefficients captured the strength and direction of the relationships between the variables. Table 25 provides the inter-correlation matrix for the variables used in the teachers' survey instrument. As shown, the three CAP measures were significantly correlated with teacher both belief variables and organizational characteristics, but different patterns were observed for the three core measures. The CAP attitude and readiness to change had the largest effect size for reported levels of use (.373; .317). The correlation coefficient between CAP attitude and levels of use was large according to Cohen's (1988) guidelines. Assessment use and professional learning community were the largest coefficients for CAP use (.338; .285) and assessment use and assessment to improve conceptions for providing feedback to students (.234; .209). Both coefficients were medium sized.

Attitudes towards CAP, assessment practice in the classroom, and teachers' conceptions of assessments as tools to improve learning appear to be central variables correlated with the fidelity and intensity of CAP. Attitude towards CAP was moderately correlated with assessment knowledge (.310) and strongly assessment use (.401). Thus teachers with a better attitude towards the Cap were usually more involved in assessment and had greater knowledge.

The organizational variables seem strongly correlated with professional learning community, organizational citizenship behaviour (.397), readiness to change (.302), and organizational citizenship behaviour (.483). Professional learning community and organizational citizenship behaviour might be part of a complex of characteristics that primed an organization for successfully implementing and sustaining CAP. CAP leadership was strongly correlated with several of these variables including organisational citizenship behaviour (.449); readiness to change (.286); organizational innovativeness (.255) and professional learning community (.397). Thus, the principal had a key but indirect role in influencing CAP use.

Table 25: Intercorrelation matrix for variables in included in survey instrument

Variables	-----CAP MEASURES------			----------TEACHER FACTORS----------								----ORGANIZATIONAL FACTORS---		
	LoU	CAP Use	FEED-BACK	KNOW-LEDGE	ASSESS Use	ASSESS Skill	CAP-Attitude	ASSESS IMPROVE	CTE	EXTRA ROLE	LEAD CAP	OCB	READ-INESS CHANGE	INNO-VATE	PLC
LoU	1.00														
CAP Use	.272**	1.00													
FEEDBACK	.174**	.229**	1.00												
KNOWLEDGE	.128*	.076	.039	1.00											
ASSESS Use	.274**	.338**	.234**	.191**	1.00										
ASSESS Skill	.200**	.214**	.142*	.235**	.698**	1.00									
CAP ATTITUDE	.373**	.229**	.125*	.310**	.401**	.271**	1.00								
ASSESS IMPROVE	.210**	.271**	.209**	.116*	.225**	.207**	.211**	1.00							
CTE	-.034	.052	.083	.125*	.004	-.028	-.006	.127*	1.00						
EXTRA ROLE	.152**	.293**	.198**	.176**	.308**	.157**	.362**	.140*	.238**	1.00					
LEAD CAP	.145*	.246**	.184**	.025	.213**	.061	.256**	.166**	.125*	.316**	1.00				
OCB	.262**	.286**	.223**	.052	.206**	.068	.247**	.293**	.250**	.355**	.449**	1.00			
READINESS CHANGE	.317**	.213**	.174**	.161**	.298**	.153**	.505**	.245**	.209**	.413**	.286**	.355**	1.00		
INNOVATE	.080	.064	.095	.144**	-.069	-.084	.160**	-.079	-.059	.028	.255**	.148**	.007	1.00	
PLC	.212**	.285**	.178**	-.007	.183**	.085	.193**	.101	.156**	.321**	.397**	.483**	.302**	.127*	1.00

*<.05 ** <.001; small effect size, *r* = 0.1 − 0.23; medium, *r* = 0.24 − 0.36; large, *r* = 0.37 or larger

Students' Perceptions of Assessment

Table 26 provide students' responses on 12 of the items in the student questionnaire. The scale is positively weighted. As shown, most students believed that assessments involved assigning a grade to their work (89.1%) and used it to check progress against expectations (85.7%) and objectives (86.2%). Fewer students considered assessments as something useful for measuring the quality of schools (81.1%) and ensuring that schools were held accountable (80.6%). Students were also less convinced about the role of assessment in promoting student collegiality (80.5%). Overall, students strongly agreed with the view that assessment assigned them to a grade and least convinced of the belief in assessment as a tool to measure the quality of schools. Thus, the dominant view was a traditional one, of assessment as measurement.

Table 26: Students' response to questionnaire items in survey from 35 sample schools

Statement	Strongly Disagree	Mostly Disagree	Slightly Agree	Moderately Agree	Mostly Agree	Strongly Agree	Mean	Std Dev.
1. Assign grade to work	7.0	3.9	5.3	7.9	20.3	55.6	4.97	1.52
2. Check progress against objectives	8.7	5.1	10.2	8.7	23.9	43.4	4.64	1.62
3. Compare work to what I am supposed to know	6.7	5.9	6.0	11.9	23.1	46.4	4.77	1.54
4. Keeps schools honest and accountable	9.5	9.9	12.1	15.3	18.1	35.0	4.27	1.69
5. Measures the quality of schools	11.5	7.4	16.2	8.4	27.0	29.5	4.20	1.70
6. Provides information on how the school is doing	4.4	3.6	8.0	8.3	23.0	52.6	4.99	1.40
7. Positive force for improving how students get along	12.0	7.5	9.1	10.4	21.1	39.8	4.40	1.76
8. Engaging and enjoyable	6.0	5.9	9.1	13.4	17.4	48.2	4.74	1.55
9. Ignore or throw away	76.3	9.9	4.0	2.3	2.7	4.8	1.59	1.32
10. Don't make use of the results	64.8	12.2	5.2	6.1	3.5	8.2	1.95	1.60
11. Ignore information	80.6	10.7	2.3	2.0	2.0	2.4	1.41	1.06
12. Makes me scared or sad	46.8	15.4	15.9	8.2	5.5	8.3	2.35	1.62

N=748

Figure 44 shows the boxplots with the means and sample distributions for all schools in the sample. Since the scale is positively weighted, benchmark of 32 (50% of the sum total) was created as a reference point. As shown in Figure 44, 19 of the 35 schools had mean scores above that benchmark. Generally, it appeared that most students viewed the assessment experience in school positively.

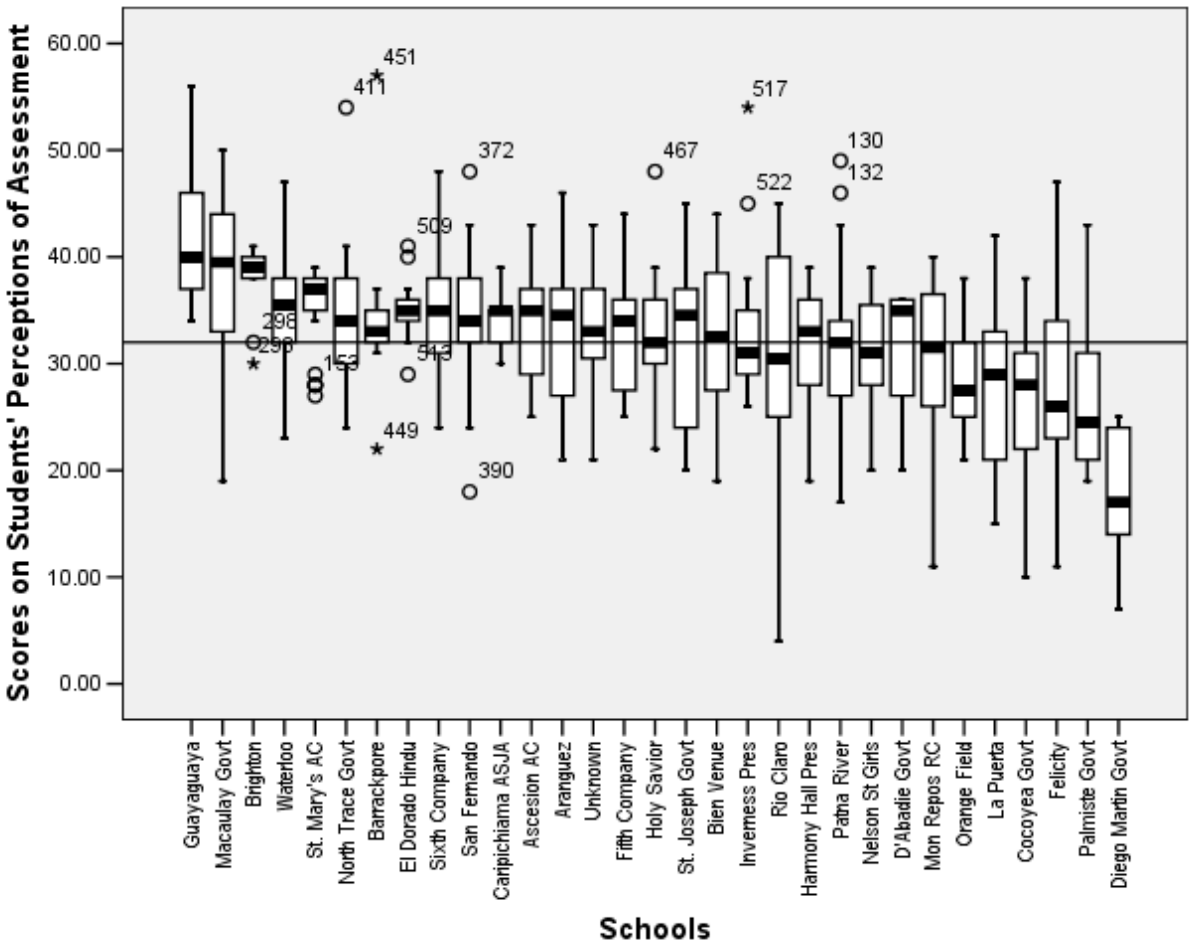


Figure 44: Boxplot of scores on 12 items in questionnaire for all schools

Content Analysis of Students' Definitions of Continuous Assessment

Students provided a definition of continuous assessment in response to a prompt. These statements were analyzed and coded. Three themes were emerged from the comments the students wrote. These themes were labelled as (1) assessment as endless testing, (2) assessment as learning and individual student progress, and (3) assessments as student and school responsibility. Two themes focused upon summative assessment and data use from summative assessment whereas themes two focuses upon formative assessment. Theme 1 was the most frequently mentioned:

----------Box 9----------

Assessment as endless testing

1. Means getting tested every week.
2. Means getting tested all the time.
3. Means getting test very often.
4. Means that in each term and class many assessments are given.
5. Means different test, projects and assignments which go on without stopping
6. Means going on and on with test, projects and assignments in class

As summarized in Box 9, students in the first category saw continuous assessments as simply repeated testing, weekly, monthly, and every term. Indeed such practices were common in several schools, in addition to the projects. The second theme shown in Box 10 was less frequently mentioned and viewed assessment as something that students could benefit from in terms of feedback, remediation and progress. As is evident in the sample statements, these positive expressions reflected assessments as a tool in education and learning. Even when the statements referred to external assessments such as national tests or the SEA, the student understood that classroom assessment was linked to their progress and preparation.

----------Box 10----------

Assessment as learning and individual student progress

1) Means being tested every week. Continuous assessment means getting information from the teacher on my performance.
2) Means getting tested regularly to know if I am improving or dropping.
3) Means being tested often to see how well I can do my work and how I can improve.
4) Allows me to learn my work and get all correct
5) Helps me do better in school it identifies my performance level and the areas I need to improve in.
6) Means thorough explanation and help on assignments which is misunderstood
7) Helps you to do well in class and in projects and big exams like SEA, National Test and CXC.
8) Continuous assessment helps in doing school work and gives an understanding of my progress. Continuous assessment test and projects helps me keep an open mind.

A third theme shown in Box 11 was that assessment was something used to keep students and even schools accountable. For students, the focus was on performance especially in external assessments. Thus, these statements often referred to the role of external agencies such as the Ministry of Education or to external assessments such as the SEA and CXC. Unlike the second theme on assessment as improvement, the primary focus of the expression was on the final external assessment.

----------Box 11----------

Assessment as student and school accountability

1) Means that if you follow a child from Standard one to Standard five in each National Test percentage they get add it up with all the other percentages to see which mark they get out of a hundred they will be judged to see which secondary school they go to.
2) Means records are kept to be given to the Ministry of Education
3) Is done to help us prepare for SEA.
4) Is a test or objective that tests your focus and academic ability repeatedly.
5) Is different set of work that comes together and teaches me valuable lessons that will help me pass test like SEA, CXC and monthly tests.
6) Is everyday in every school, and every term the children gets tested and at the end of the term they have to do a test called national test then they move on to the biggest test of their lives the SEA examination.

Evaluating the Change Process

Following the procedure outlined in George, Hall and Steigelbauer (2006), aggregate data from the seven categories of the stages of concern questionnaire were used to plot a graph showing the relative intensity of seven areas of concern. To obtain the relative intensity, mean group scores were converted into percentiles and then plotted in a graph. The profile thus represents the average of all the users or all the users within a category for schools. George, Hall and Steigelbauer hypothesized on the patterns for different types of users, three of which are shown in figure 45.

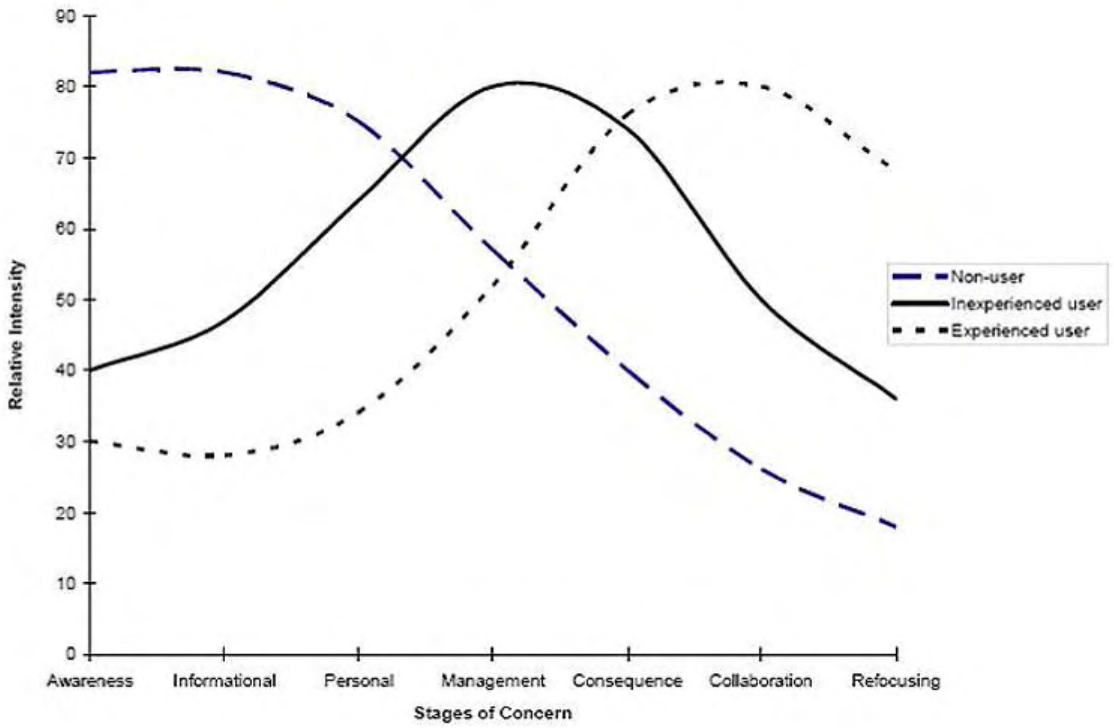


Figure 45: Hypothesized Development of Stage of Concern (Renewing User not shown)

As shown in the stages of concern profile in Figure 43, the average profile for this innovation was that of *nonuser*, representing a lack of concern for the innovation in the teacher's work. As shown, the awareness, informational, personal, and management concerns were very high, which would not be expected in a ten year old intervention. Although some teachers had begun refocusing, consequence and collaboration, these concerns remained low. The high score on stage 1-awareness concerns points to the lack of penetration and viability of the innovation. Doing CAP was not something that teachers were concerned about in their work. In other words, it was not a high priority for the great majority of teachers. At the same time, the high score in informational concerns suggested that teachers still wanted to know more about the structure and function of CAP. This should not be considered a contradiction but simply suggests that teachers had other concerns in the system but when it came to CAP, they wanted to know more about its nature and function.

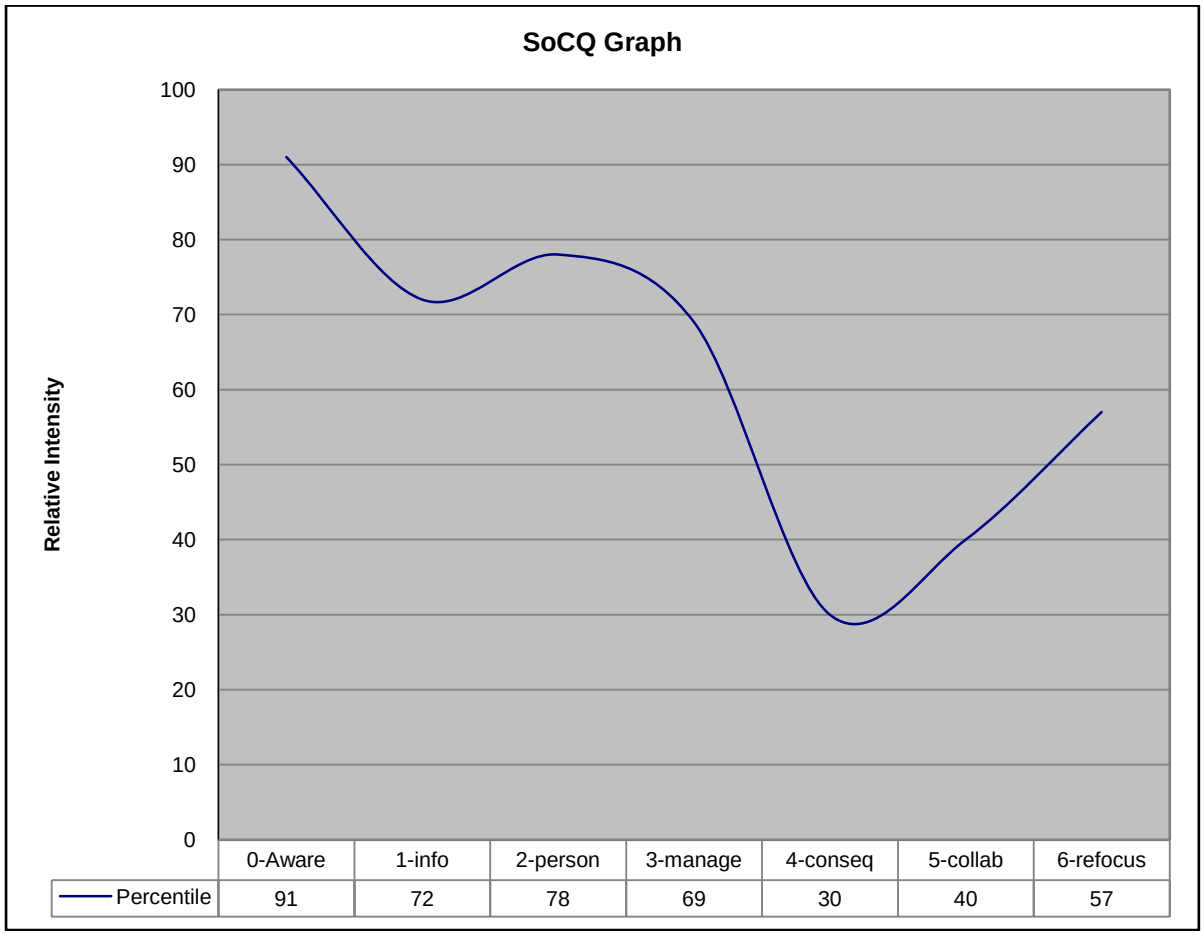


Figure 46: Sources of concern graph for all teachers in sample

Scores for stages 2 and 3 were relatively higher than stage 4, 5, and 6, indicating that teachers' personal concerns and concerns for the management, time and logistical aspects of CAP were relatively more intense than their informational concerns. The overall profile suggests that the generally users for this innovation at this point had doubts about the intervention with varying degrees of potential resistance. If the CAP was to be reintroduced now, this profile suggests that personal concerns may be further intensified.

This is not the kind of pattern we would expect in a mature successful innovation. The goal according to Hall and Hord (2010) should be to have consequences, collaboration and refocusing as the peaks. Although rare in practice, a well managed innovation would have achieved these goals, rather than an intense lack of concern for the intervention. The sharp dip in consequences and collaboration suggests that a student focus and collaborative teams was rare.

Comparing Pilot and Non-Pilot Schools

It was hypothesized that there should be differences between pilot and non-pilot schools for both the CAP outcomes and the CAP change process. However, the data suggested that differences between the means for the different stages of concern from the pilot and non-pilot schools were not statistically significant. Table 27 provides the mean and median scores for each of the levels. There were very few notable differences.

Table 27: Mean and median scores on Stages of Concern for pilot and non-pilot schools

School	Mean Scores						
	Awareness	Informational	Personal	Management	Consequence	Collaboration	Refocusing
Pilot	15.10	19.82	21.76	18.71	19.38	18.73	18.28
Non-Pilot	13.79	20.18	21.42	18.21	19.95	18.06	17.42
Total	14.16	20.08	21.52	18.35	19.79	18.24	17.66
	Median Scores						
Pilot	15.00	20.00	22.00	20.00	20.50	19.00	19.00
Non-Pilot	13.50	21.00	22.00	19.00	20.00	18.00	17.50
Total	14.00	21.00	22.00	19.00	20.00	18.00	18.00

Figure 47 shows the similarity between the graphs of the profiles for pilot and non-pilot schools. As shown, these graphs are very similar pointing to a common pattern of non-use in both pilot and nonpilot schools.

High, Medium, & Low Achieving Schools

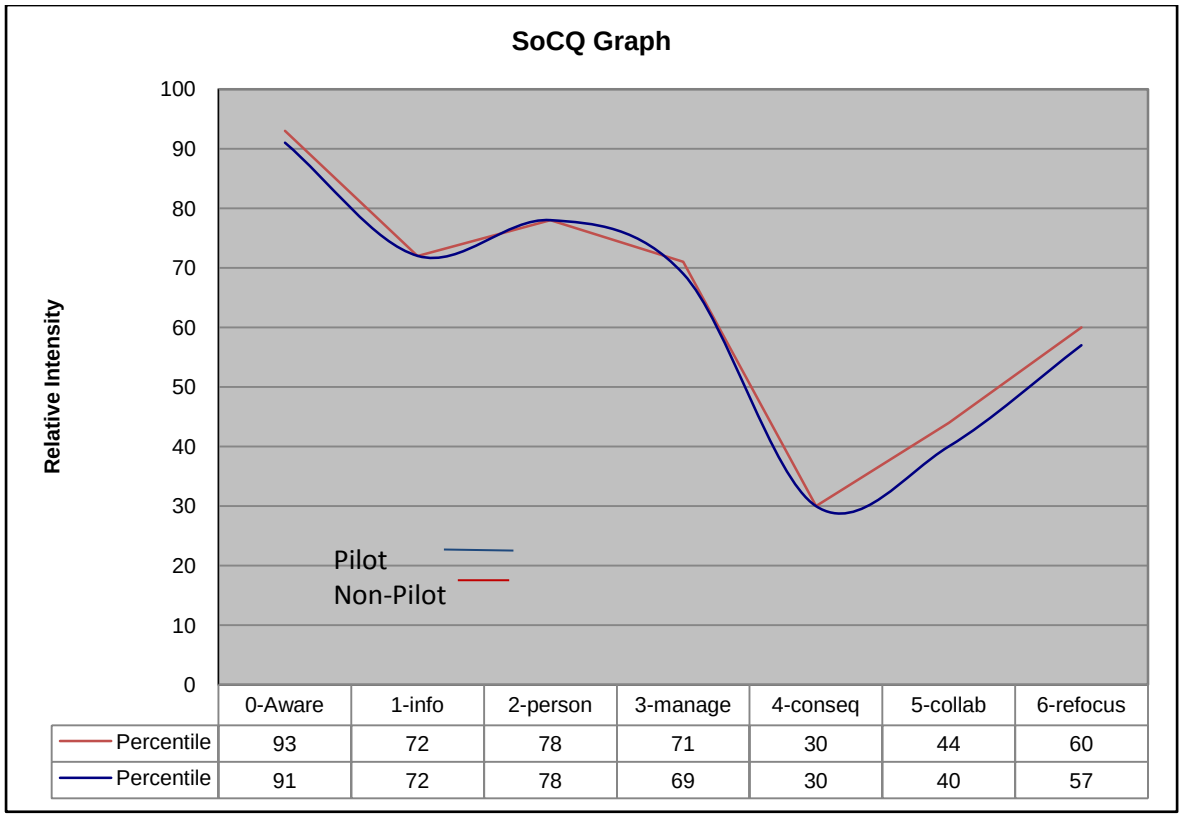
Figure 48 shows the differences in the profile for high, medium, and low achieving schools, although the percentiles for personal concerns (stage 2) for low achieving schools is slightly higher, the profiles are again basically the same. Low achieving schools may be slightly more concerned with managing CAP as an innovation.

Comparing the Phase 3 Schools

Although there was little difference in the stages of concern profiles between the pilot and nonpilot schools, there were some differences in the stages of concern profiles for the three schools purposively selected for the Phase 3 explanatory study. As shown in Figure 48, there were greater differences compared with the pilot schools and schools with different achievement levels. These differences, then, were related to the pattern of development of the innovation and the current levels of use in the institution. Notably, all schools reported intense concerns at stage 0-awareness, highlighting the failure of the innovation to take root

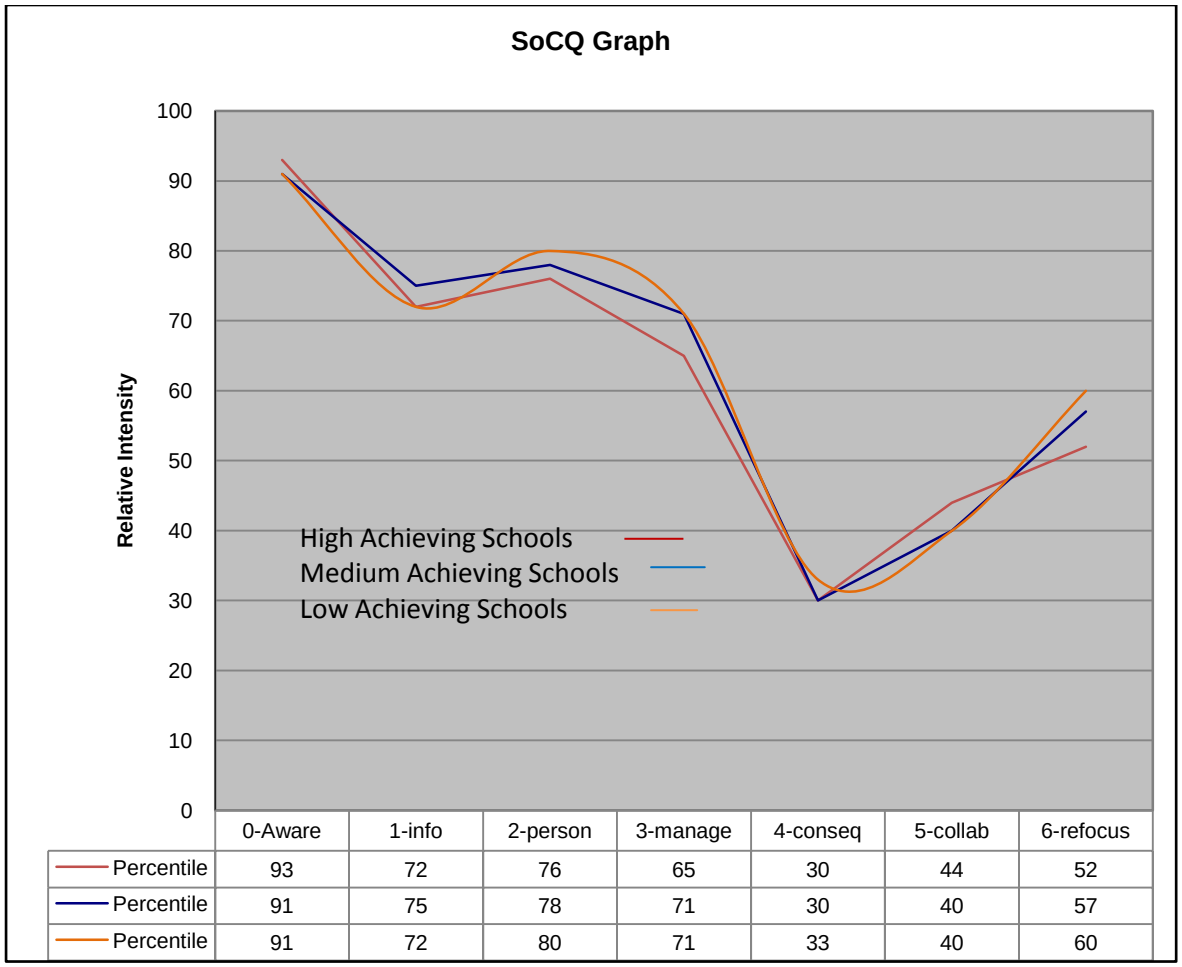
in the work lives of teachers. Although this level was much lower for San Fernando TML, the percentile of 81 still did not meet the broad benchmarks hypothesized in Hall and Hord (2006). Informational and personal concerns were about the same for all schools, but there were sharp differences at the management stage, with Nelson Street Girls having increased concerns and San Fernando TML lowered concerns. Nelson Street also reported higher consequences, collaboration and refocusing concerns. Nelson Street Girls SoC profile approaches the big W pattern described by Hord and Hall (2010). The high management, refocusing concerns, and awareness concerns suggest that teachers, from their own perspective, had strong views about what should be done differently and it would be quite difficult now to provide the information about CAP that they said they needed.

The teachers at San Fernando TML were collectively less concerned about self (0,1) and more concerned with task (3,4), whereas teachers at Nelson Street Girl's were more concerned with self (0,1) and impact (5,6). Based on the benchmark profiles provided in Hall and Hord (2006), San Fernando TML may be described collectively as positive nonusers and Bien Venue and Nelson Street Girls as distrustful nonusers. The goal of implementation is to have the highest concerns with task and impact, but these have not been achieved in any school for the CAP, even several years after adoption. These patterns point to the mismanagement of implementation and the failure to consider the needs of the users as paramount. There are no clear strategies for change in this regard, and the role of change agents and the function of monitoring resistance have been mostly neglected.



	0-Aware	1-info	2-person	3-manage	4-conseq	5-collab	6-refocus
Percentile	93	72	78	71	30	44	60
Percentile	91	72	78	69	30	40	57

Figure 47: Sources of concern graph for all teachers in pilot and non-pilot schools



	0-Aware	1-info	2-person	3-manage	4-conseq	5-collab	6-refocus
Percentile	93	72	76	65	30	44	52
Percentile	91	75	78	71	30	40	57
Percentile	91	72	80	71	33	40	60

Figure 48: Sources of concern graph for teachers in high, medium and low achieving schools

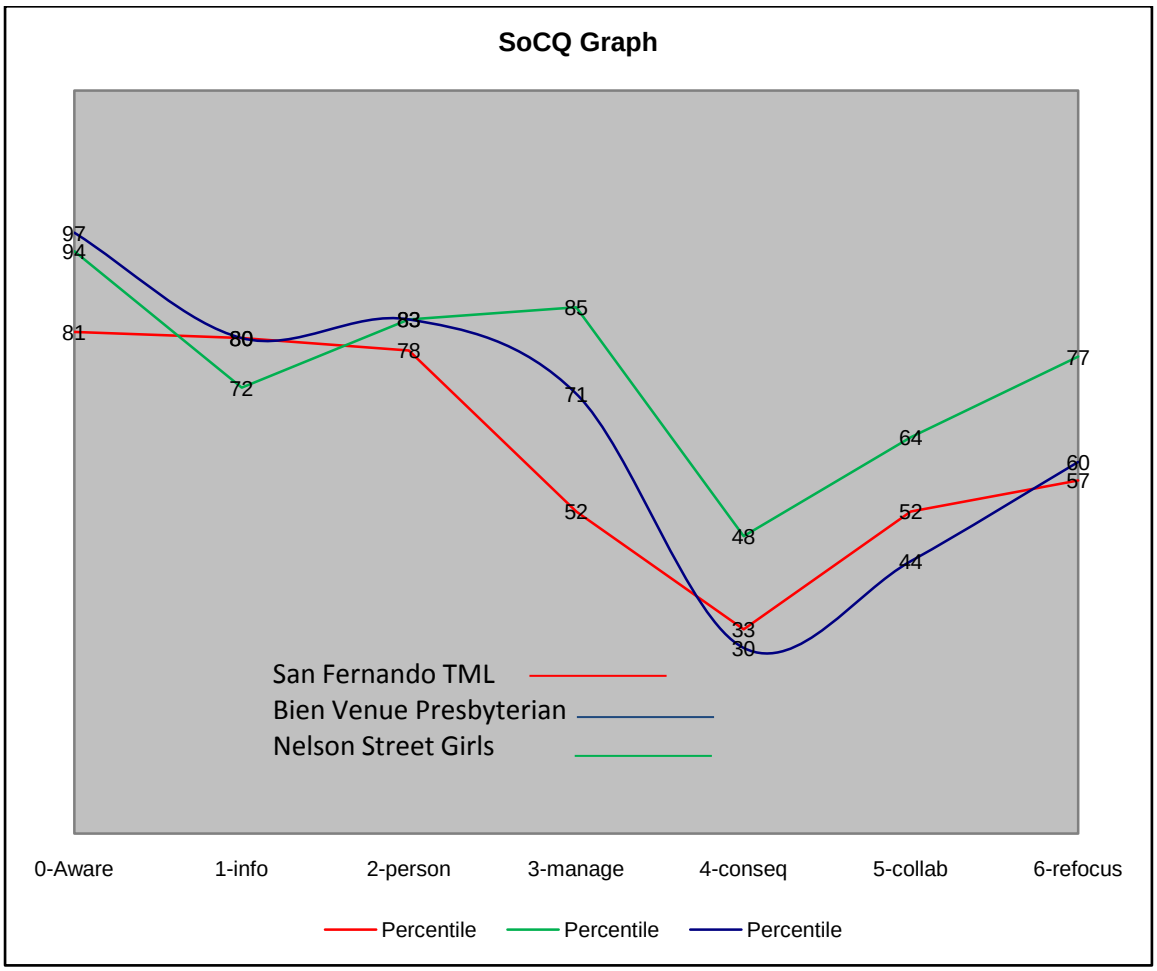


Figure 49: Sources of concern graph for teachers in schools selected for Phase 3

Table 28: Regression and Correlation coefficients for conceptions of teaching, curriculum and assessment and various measures of CAP use

Teacher Belief Systems	Measures of CAP Intensity & Fidelity								
	---------CAP Use----------			----------LOU----------			----------FEEDBACK-------		
	----------Beta & Correlation Coefficients ----------								
	Beta	**Partial**	**Part**	**Beta**	**Partial**	**Part**	**Beta**	**Partial**	**Part**
Assessment for School Accountability	-.049	-.043	-.040	.118	.099	.095	.029	.025	.023
Assessment for Student Accountability	.000	.000	.000	.031	.026	.025	.061	.053	.051
Assessment to Improve Education	.**223***	.169	.158	.085	.064	.061	.104	.078	.074
Assessment as Irrelevant	-.094	-.097	-.090	-.036	-.036	-.034	-.055	-.056	-.053
Academic Curriculum Conception	.068	.057	.053	-.004	-.003	-.003	.061	.050	.047
Social Reconstruction Curriculum Conception	.003	.002	.002	-.036	-.027	-.026	.082	.062	.059
Technological Curriculum Conception	-.009	-.008	-.007	-.001	-.001	-.001	-.089	-.073	-.070
Apprenticeship-Developmental Teaching Conception	.103	.074	.068	.147	.099	.095	-.005	-.003	-.003
Nurturing Teaching Conception	.042	.030	.028	.051	.035	.033	**.171***	.118	.113
Social Reform Teaching Conceptions	-.036	-.028	-.026	.001	.001	.001	.001	.000	.000
Transmission Teaching Conception	.**158***	.158	.148	-.111	-.108	-.104	-.053	-.053	-.050

One of the models of change considered in the programme theory was teachers' receptivity as defined by a range of attitudes and beliefs associated with teaching, curriculum and assessment. Table 28 considers this model (with all the determinants) separate from later analyses which will compare the different sets of variables which act as antecedents. Both regression and correlation coefficients are shown. For CAP use, the primary measure of intensity, perceiving assessment as a way to improve education and having a transmission conception of teaching were both significant factors positively influencing the full use of the innovation.

It is easier to explain the view of assessment as a tool influencing CAP use than it is to explain the possible influence of transmission conceptions. It must be remembered, however, that the CAP process includes weekly and monthly monitoring through traditional tests that are very characteristic of traditional good teaching and this might explain the positive impact of transmission teaching conceptions on the use of CAP. Teachers with transmission conceptions may be motivated to complete the CAP tasks but in a routine and mechanical way. This is supported by the fact that a nurturing conception of teaching positively influenced the use of feedback. Thus, teachers with high scores on nurturing conceptions were more likely to provide feedback. These results reinforce the point that there are tensions and a duality in teachers using classroom assessment for summative or formative purposes. The different purposes are not easily integrated without an explicit framework such as in standards-based assessment.

Quantitative Modelling

The statistical routines used in the modelling exercise considered all of the variables together, providing an indication of comparative impact or influence. The statistical routines included multiple ordinary least squares (OLS) linear and logistic regressions along with canonical analysis. These routines were used to both reduce the number of variables and construct models of antecedents against CAP implementation and outcomes. A path analytic model is based on the standardized regression coefficient and shows the direct and indirect pathways of each of the antecedents.

Modelling LoU from Teacher Demographics

Evaluation Model Question: What categories of demographic variable best explain levels of use? (Logistic Regression)

The levels of use outcome was a single dependent variable with 8 categories or levels, as already described and shown below in Table 29. This model considered the relative impact of selected demographic variables expressed on levels of use. Table 29 summarizes each category and shows the total number in the sample for the model after forms with missing data were deleted. The overall model included age, ethnicity, gender, tenure in profession, tenure in school, academic qualifications, professional qualifications, academic achievement status of school, and pilot status of school. Several statistics may be used to judge the utility of the model including the p-value of the chi-square statistic for the model likelihood ratio, the percent predicted statistic (Shown in Table), and several Pseudo R^2 statistics generated in the SPSS output.

Table 29: Numbers and classification prediction in logistic regression model

Category Choices	No.	Percentage Predicted by Demographic Model
Level 0 - Non-use-Having little or no knowledge of innovation.	21	42.9%
Level I – Orientation-Acquiring some information of innovation	34	45.5%
Level II – Preparation-Preparing for first use of innovation..	7	83.3%
Level III - Mechanical Use - Focusing on immediate needs of user to master tasks of innovation.	33	35.5%
Level IVA- Routine-Making few changes in ongoing use of innovation	32	12.5%
Level IVB – Refinement-Varying use of the innovation to make impact on students	73	45.1%
Level V – Integration-Combining efforts of self and colleagues to achieve collective impact on students	75	46.6%
Level VI – Renewal-Reevaluating own use, seeking major modifications, and exploring new developments.	10	60.0%
Total	285	41.9%

The p-value for the Chi-square statistic of 197.84 at 287 degrees of freedom suggests that the model was not statistically significant. Reducing the model by deleting the different variables would not enhance the output as shown in Table 30 which summarizes the Likelihood Ratio tests.

Table 30: Likelihood Ratio Tests on variables in final and reduced models for logistic regression on teachers' self reported Levels of Use

Effect	Model Fitting Criteria	----------Likelihood Ratio Tests----------		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	819.702(a)	.000	0	.
Age	848.804(b)	29.102	56	.999
Ethnicity	736.570(b)	.	28	.
Professional Tenure	851.724(b)	32.022	42	.868
Gender	827.512(b)	7.810	7	.350
School Tenure	850.539(b)	30.837	42	.898
Highest Professional Qualifications	845.246(b)	25.544	42	.979
Highest Academic Qualifications	849.172(b)	29.470	49	.988
School Academic Status	823.883(b)	4.181	14	.994
School status as a pilot school	820.662(b)	.960	7	.995

Modelling sets of independent and dependent variables

Evaluation Model Question: What sets of antecedents explain which set of outcomes? (Canonical Analysis)

Rather than conduct separate multivariate regressions, canonical analysis allows a set of independent variables to be related to a set of dependent variables. The dependent variables used were CAP use, CAP feedback, and CAP multimodal assessment. The latter is a measure of the variety of assessments used in the classrooms following the CAP. All three variables are correlated, the largest being CAP use and CAP multimodal assessment (R=0.335), followed by CAP use and CAP feedback (R=0.238), and CAP feedback and CAP multimodal use (R-0.195). The independents included the full set of demographic, user related, organizational, and contextual variables. The overall model had a statistically significant RAO F of 2.102 at 72 degrees of freedom. The percentage variance explained in the R-square was 48.3%.

The set of variables predicted CAP multimodal assessment (F-ratio=2.100; P-value=0.003) and CAP use (F-ratio=3.885; P-value=0.000), but not CAP feedback (F-ratio=1.371; P-value=0.124). Antecedents with significant betas are shown in Figure 1. Notably, none of the antecedents were predictive of feedback in the canonical analysis. Feedback was the primary measure of fidelity in the study and an indicator of the quality of formative assessment in action. The canonical analysis suggests that for CAP to succeed, teachers were required to go beyond the boundaries of their job as understood. Hence, the value of professional learning community in this context might be to provide the support necessary. Teachers who used CAP were more likely to see assessment in a positive light as having a role in improving education. However, teachers who have a social reformist view of the curriculum, seeing societal change as more important than learning, tended not to emphasize multimodal assessment in the classroom.

Table 31: Significant betas from canonical analysis

CAP USE		CAP Multimodal Assessment		CAP Feedback	
Antecedent	*Beta*	*Antecedent*	*Beta*	*Antecedent*	*Beta*
Assessment to improve Education	.213**	Extra-role Behaviour	.240**		
Extra-role Behaviour	.232**	Social Reform Conception of Curriculum	-.178*		
Professional Learning community	.142*				
Transmission Conception	.238**				

What sets of antecedents are predictive of the different outcomes as modelled in a path analysis? (Multiple Linear Regressions & Path Analysis)

Full Model compared with user and contextual models

Statistics for the regression models are shown in Table 32. The full model includes teacher receptivity, organization, and environment variables. This full model explained 31.2% of the variance in CAP use. The full model was superior to using teacher receptivity (24.7%) and organizational and environmental variables (15.5%). However, teacher receptivity variables had greater explanatory power for feedback (15.6%) compared with the full model (14.7%). Teacher receptivity variables were also a good predictor of the use of multimodal assessment (16.4%), although less than the full model (19%). This suggests that fidelity and intensity were quite different outcomes explained by a different set of factors in the user and the organization.

Table 32: Statistics explaining different models for three outcome variables related to CAP

Models	---------- Statistics ----------				
Dependent = CAP Use	**R**	**R^2**	**Adjusted R^2**	**F-value (ANOVA)**	**Sig**
Full Model	0.559	0.312	0.247	4.784	0.000
Teacher Receptivity	0.497	0.247	0.205	5.796	0.000
Organization & Environment	0.394	0.155	0.136	8.256	0.000
Dependent = Feedback					
Model					
Full	0.383	0.147	0.066	1.816	0.016
Teacher Receptivity	0.394	0.156	0.108	3.237	0.000
Organization & Environment	0.312	0.098	0.078	4.867	0.000
Dependent = CAP Multi Modal					
Models					
Full	0.436	0.190	0.113	2.468	0.000
Teacher Receptivity	0.404	0.164	0.116	3.446	0.000
Organizational & Environmental	0.212	0.045	0.023	2.090	0.055

Path Model for CAP use

In order to obtain the standardized regression coefficients for each independent variable in the path analysis, several regression analyses were conducted. Path analytic diagrams were constructed for the three continuous outcome variables, CAP use, CAP multimodal assessment, and CAP feedback, showing only the significant Betas. The path diagrams are shown in Figures 50 to 52. Figure 50 provides the direct and indirect paths to CAP use. The multiple direct paths include teachers' transmission conceptions of teaching (Beta=.194), a perspective on assessment as improving education (Beta=.226), extra-role behaviour (Beta=.133), professional learning community in the school (Beta=.149), readiness for change (Beta= -.163), and the academic achievement context of the school (Beta=.133). The largest impact was on the assessment for improving education conception showing the critical role of assessment conceptions in CAP use.

These direct paths may be explained by the multiple component nature of CAP, which is both an accounting (summative assessment) and learning system (formative assessment). As such, teachers' conceptions and extra-role behaviour were important in facilitating high usage levels. It must be remembered, however, that high CAP usage does not necessarily mean high levels of formative assessment since continuous summative assessment is a critical part of CAP. The heavy demands of CAP on the individual teacher meant that supportive aspects of the organization such as professional learning community and readiness for change were important facilitators of change. The direct influence of professional learning communities and readiness for change on CAP use also means that to be successful, schools must be prepared for change.

The indirect paths were also notable and emphasized the key role of systemic change factors in explaining high CAP use. Group organizational behaviour, for example, influenced extra-role behaviour (Beta=.201) and collective teacher efficacy (Beta=.224). CAP leadership and readiness for change in the organization also influenced extra-role behaviour. Group organizational citizenship behaviour also influenced positive conceptions of assessment as a tool for improving education. Thus, acting through these two variables, the ability of the organization to facilitate change was a key factor in high CAP use. We build on these findings in our recommendations to propose a plan for the "priming" of school organizations for the uptake and scale-up of new organizations.

Teacher perceptions and belief systems were critical factors but not always directly related to the intensity of the activity. An important example is attitude towards CAP which influenced CAP use through assessment improving education perceptions (Beta=.185) and extra-role behaviour (.314). This variable was itself strongly influenced by organizational readiness of change (Beta=.471). Thus one possible role of an organizational priming effect might be to improve user's ability to receive the innovation.

Path Model for CAP Feedback

The only variable directly influencing feedback, a critical element of formative assessment, was CAP training (Beta=.158). Attitude towards the CAP might also have influenced feedback indirectly through CAP training (Beta=.167). Readiness for change had a strong

impact on attitude towards CAP (Beta=.471), possibly indirectly influencing CAP training and the provision of feedback. Organizational innovativeness had a direct influence on CAP training (Beta=.150) and indirectly on feedback.

Feedback was a more difficult variable for this set of variables to predict because it might be rarely practiced to an adequate degree in schools or is predicted by a different set of variables. The model suggests that the change process is certainly amenable to organizational priming and training. The role of organizational innovativeness on CAP training might suggest flexible organizations willing to experiment might be able to develop PD supportive of formative assessment at the sites. This, however, is likely a rare phenomenon.

Path Model for CAP Multimodal

Multimodal assessment is an important aspect of CAP related both to intensity and fidelity of outcomes. Several variables were involved in this model. The path diagram shows four direct influences from the teacher receptivity framework, all from the "teacher as user" variable set. These include social reform curriculum conceptions (Beta=-.175), attitude towards CAP (Beta=.185), extra-role behaviour (Beta=.234), and CAP training (Beta=.186). Group organizational citizenship behaviour had both an indirect effect through extra-role behaviour (Beta=.201) as well as a direct effect on multimodal assessment (Beta=.158). Group organizational citizenship behaviour also influenced multimodal assessment through extra-role behaviour. Leadership also has a indirect effect through extra-role behaviour (Beta=.172)

Several organizational priming variables such as readiness for change and organizational innovativeness also had indirect effects. The academic achievement context had a direct positive effect on group organizational citizenship behaviour (Beta=.143) and a negative direct effect on organizational innovativeness. This meant that schools with high achievement status tended to report low organizational innovativeness scores and were possibly less prone to using CAP multimodal assessment, bearing in mind that the effect of organizational innovativeness on through attitude towards CAP and extra-role behaviour. Social reform teaching conceptions were also negatively related to CAP multimodal assessment, suggesting that teachers with this perception of teaching often tended to use less multimodal assessments.

Figure 50: Path Analysis Model for CAP Use

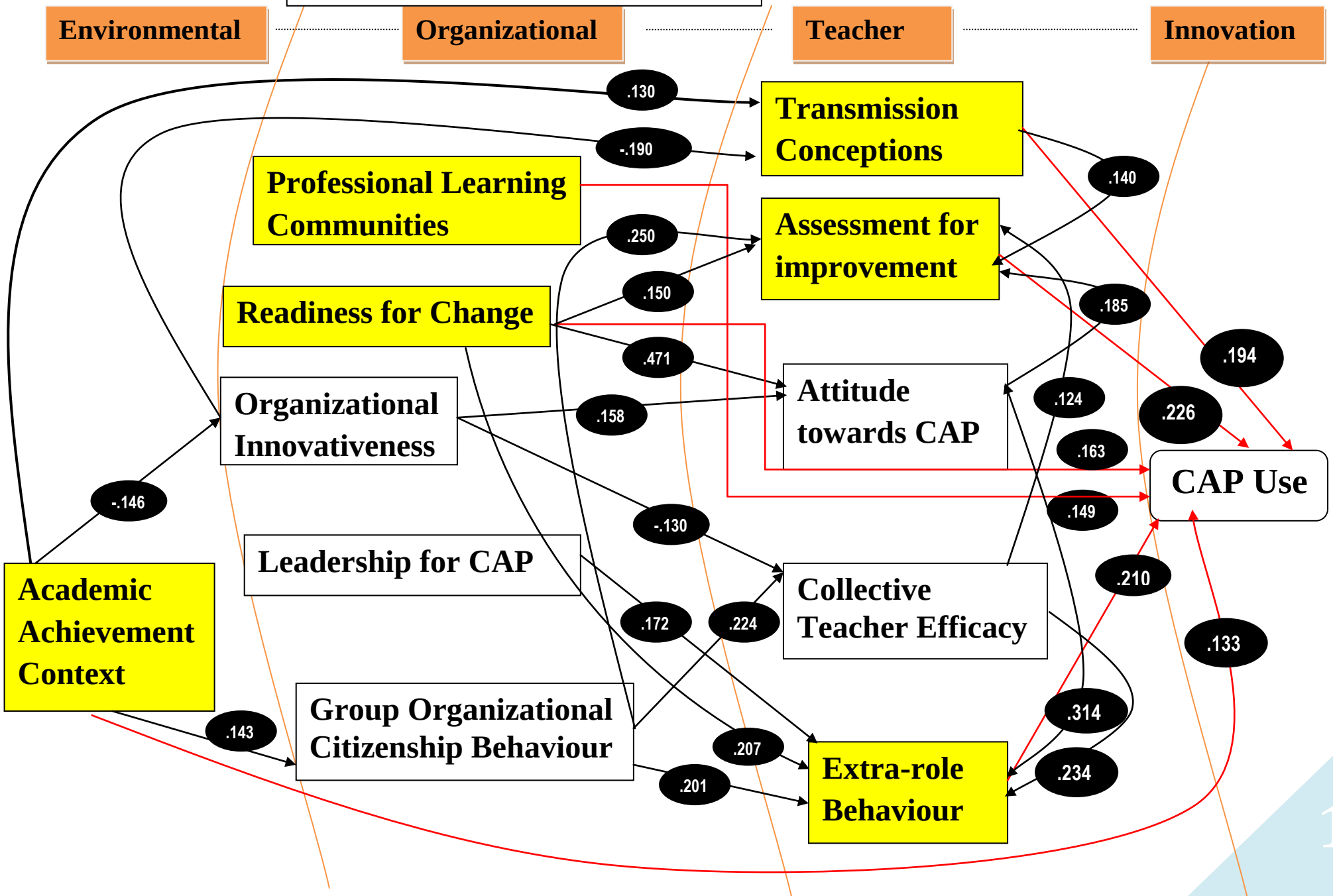


Figure 51: Path Analysis Model for Feedback

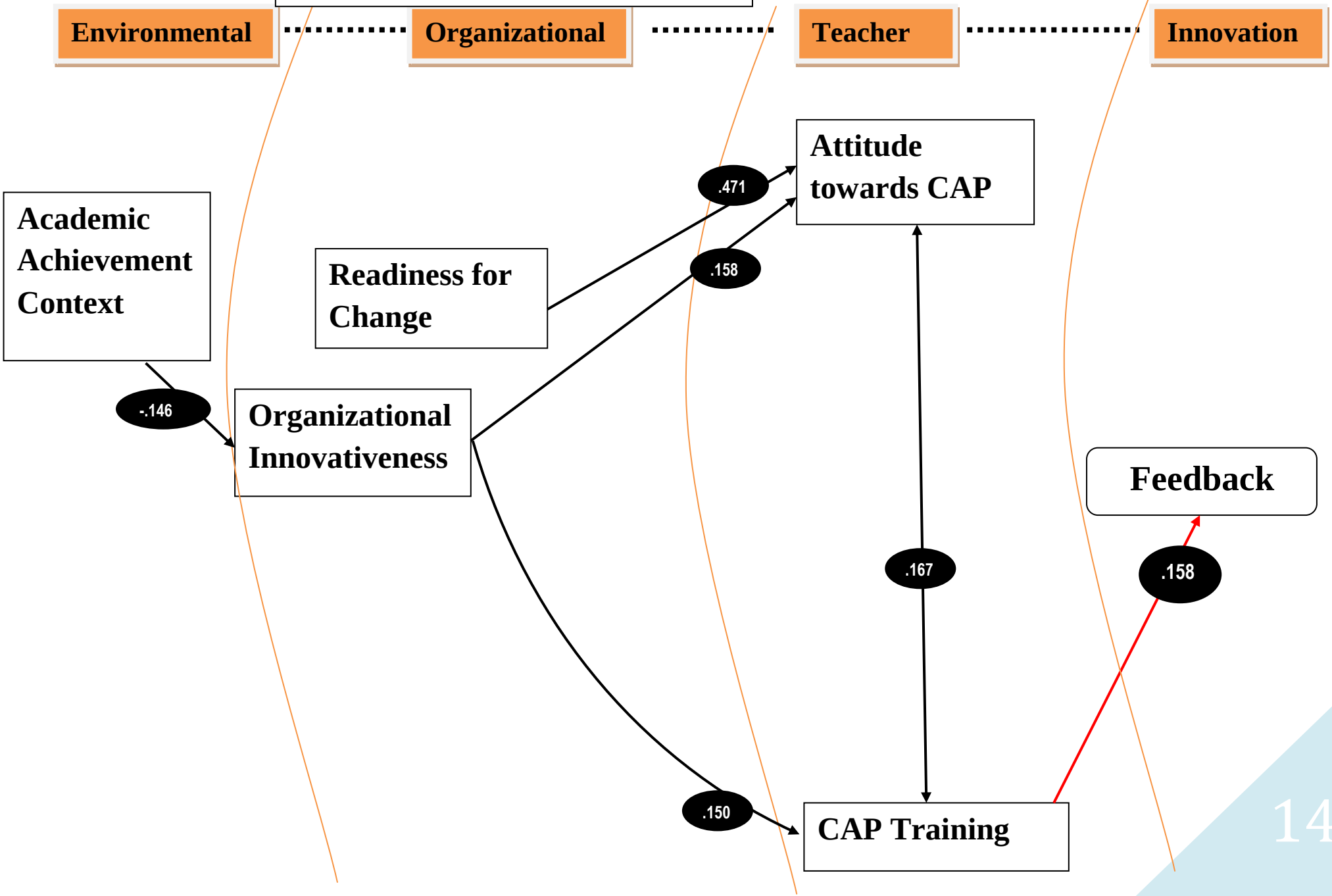
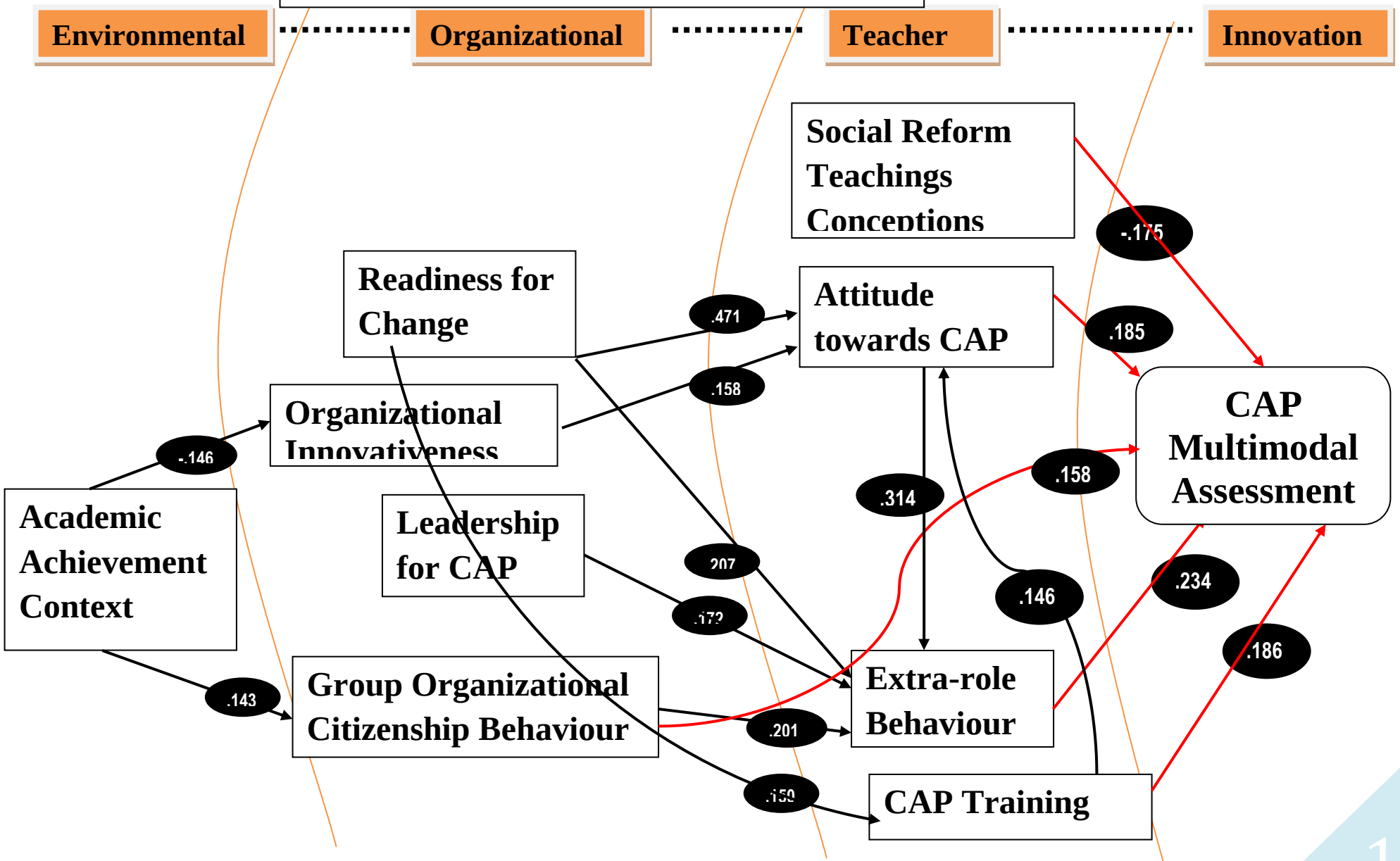


Figure 52: Path Analysis Model for Multimodal Assessment



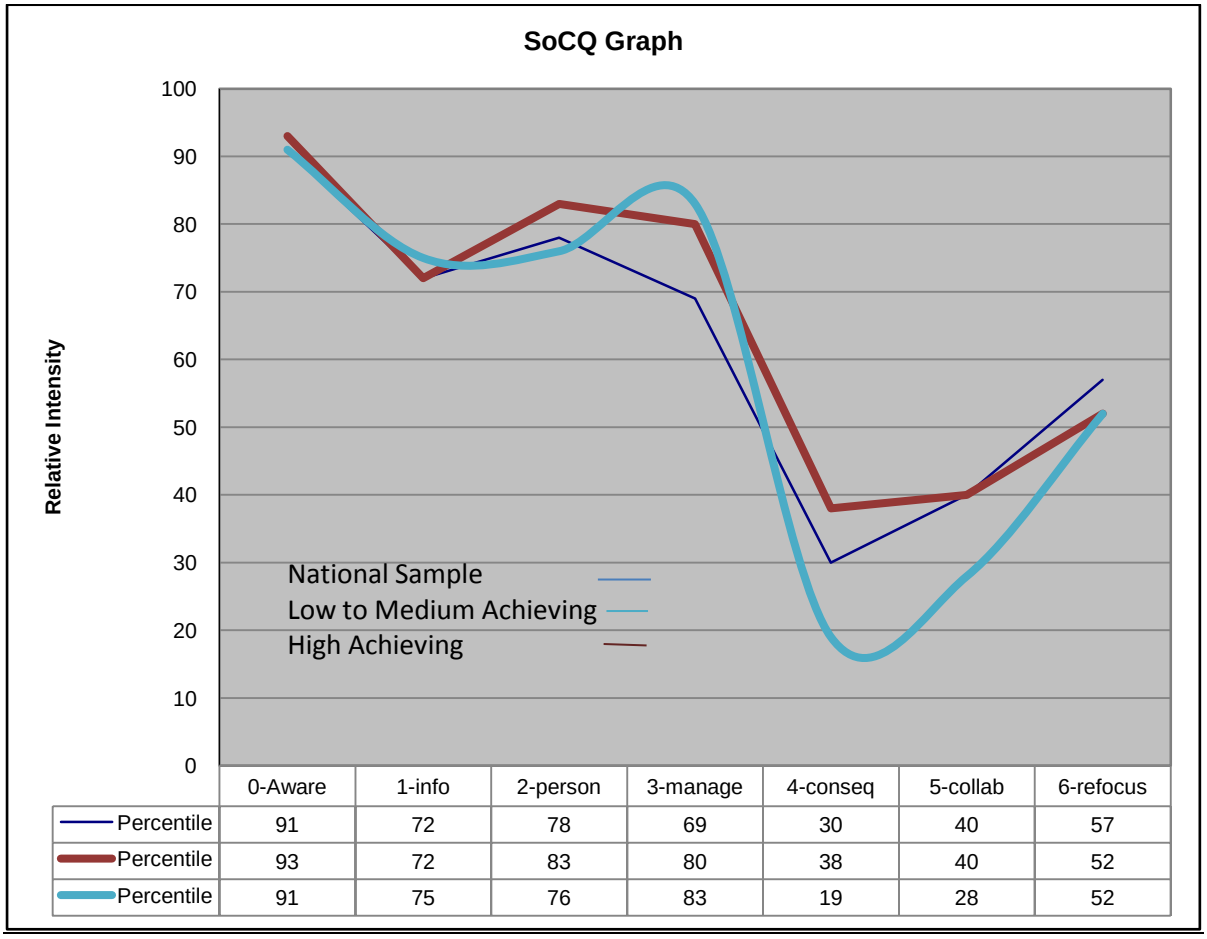
PHASE 3: Multisite Explanatory Case Studies

Qualitative Findings

THE PRINCIPALS

Selection Strategy for the school principals

To gather further insight into the administrative and organizational elements of the process, we interviewed two principals from high implementation sites together. Although the schools were located close by, one school had a low to medium achievement context and the other had a high achievement context; however, both schools reported high levels of implementation along with adequate levels of fidelity, which was low in the majority of the schools. The two schools were physically large, but the high achieving school was overcrowded.



	0-Aware	1-info	2-person	3-manage	4-conseq	5-collab	6-refocus
Percentile	91	72	78	69	30	40	57
Percentile	93	72	83	80	38	40	52
Percentile	91	75	76	83	19	28	52

Figure 53: SoC profile for schools led by the principals interviewed for Phase 3

The Sources of Concern profile shown in figure 53 suggested that there were slight differences between the profiles of these schools and the national profile. As shown, the low to medium school peaked at management concerns but had a sharper valley for the consequences compared to the national sample. The high achieving school had greater concerns on person and management than the national sample and also had higher consequences concerns. However, both schools still had high levels of a lack of concerns measured by Stage 0.

Two principals, two schools, one purpose

In the low to medium achieving school, the principal had been in the school since 2003. The school had finally been able to get school indiscipline under control and had turned its focus towards achievement and instruction, most notably reading. In the high achieving schools, the principal had been in her post since the year 2000. She had come to the school from a pilot school in which CAP had been done extensively and a whole school approach had been adopted for project work. In her current school, however, the projects were done by levels based on the curriculum. The pilot school was very low achieving and Student Support Services was not as yet in existence.

Both principals understood fully the purpose of CAP, but focused on different aspects of the process. The principal of the medium-achieving schools focused on the multimodal assessment but the principal of the high achieving school emphasized student readiness and remediation, a philosophy that was more aligned to the programme designer. Their instructional leadership consisted of tight, continuous monitoring and mechanisms and structures to ensure compliance although the feedback from teachers did not indicate that these were considered draconian measures. For example, the principal of the high achieving school noted:

> **Every term I do class checks and the completion of the cumulative record cards is part of it. It is on my checklist and I tell them I use the checklist when I am doing the confidential report.**

The principal of the high achieving schools recounted that her training as one of two individuals for a pilot school was extensive. In that pilot school, there was initially high buy in, but the extensive recording soon proved difficult and the intervention-action phases did not prove sustainable, primarily because of a lack of resources. Greater buy-in would have required more time and individual application, as the principal noted:

> **We did buy in to it but we found that when we identified students that needed intervention, it was very slow. . . We found that eventually teachers said, no. Initially teachers also found the recording was too much, but eventually they saw the importance of it. But at the beginning, it was very difficult.**

The current success of this principal was because she had been involved in the pilot training and was able to introduce and promote some elements, although these actions were not necessarily verified in the teachers' interviews. The principal of the medium-achieving school

also had some experience of CAP in a pilot school. Moreover, both teachers had Masters degrees in education or psychology.

The primary CAP activities in both schools were data collection and project work; however, even screening was still being done in the infants' school on entry. Nevertheless, the principals considered the projects to be the most successful aspects of CAP. The medium achievement schools conducted school-wide projects once per term and had implemented portfolios in the infants and journal writing in Standard 1. Both schools had whole school presentation sessions for projects, with students actively involved in presenting and observing other students' work. Both schools also collected much data, but there was definitely greater data collection with higher fidelity (in data collection) in the high achieving school. Both schools also had very high compliance with the Cumulative Record Cards completion and the high achieving school also regularly used the CAP booklet. The medium achieving school were no longer using the CAP student performance books for recording marks because it did not facilitate sharing of marks between and across years.

In both schools, physical resources such as photocopiers and books had been supplied for the CAP process, but the collection was neither extensive nor adequate. Human resources and physical space were also an issue in the high achieving denominational school. Training for staff was not extensive although the medium achieving school had conducted training in rubric development on its own. The lack of training appeared to be a sore point and was the first consideration given when the principals considered how the process might be improved:

> **First of all, all teachers need to be trained-Training of all teachers. Miss was saying that some teachers were trained about many teachers don't even understand what CAP is. Even miss has a different conception of what CAP is. If we know that the marks are being used to propel the children to a different level, then it might have had greater impact. [Interviewer: So sometimes, it's not clear?] It's not clear. So that we just doing this thing and we put in the marks and somebody says "But when they go to Secondary school, that's not necessary". Even if the marks won't use to help the students in the SEA - and in my view that is what continuous assessment is- but right now all you are doing is using the marks in the SEA and this right here says nothing. There is no continuity-it is supposed to be a continuous process, but right now, there is no continuity.**

However, the principal did not seem to recognize that the two data uses might be in conflict- data for remediation and intervention and data for high stakes selection processes. The latter would certainly require extensive statistical moderation. The principal of the high achieving school did recognize, however, that remediation and monitoring standards were an important reason and motivator in CAP. She noted:

> **You asked about the motivation? I have some standard 3 students who cannot read three letter words and I show the teachers that if we were tracking that at the beginning, it would not have reached standard 3, okay, we might have made them repeat some levels or call in and get some help. So**

I tell them that is the purpose –You find that it is a lot of work, but that is part of the purpose.

Of course, in reality, the extensive CAP data collected at the school was not often used for such purposes; instead, an annual diagnostic test was employed, which is surprising given the effort to collect continuous classroom data. The high stakes SEA interfered with projects in the senior school. Although in the medium achieving school, data use was not extensive, the institution had better integrated its human resources and collaborative remediation structures with the CAP data. This had brought some change in achievement in the national assessments of educational achievement.

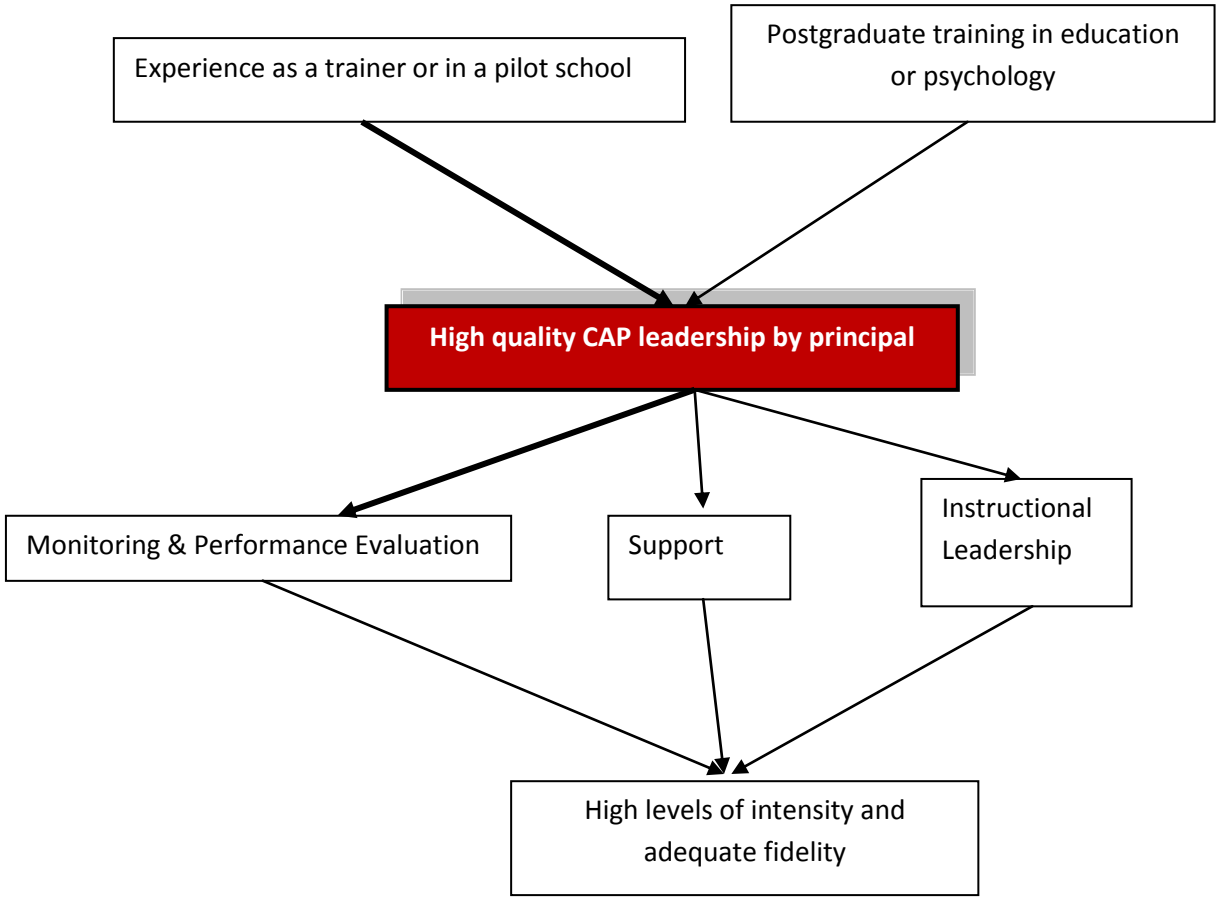


Figure 54: Antecedents and consequences of high quality leadership of the CAP.

SCHOOL SITES

Selection Strategy for schools

The schools chosen for the case studies were selected using a sequential mixed methods sampling strategy. The criteria for selection included Phase 2 scores for CAP programme strength and fidelity and selected organizational characteristics. Initially, three schools were selected, but access to San Fernando TML was denied at the time. The other two schools in the sample, however, provided very contrasting environmental and implementation contexts. Bien Venue Presbyterian reported medium to high levels of implementation from the

administrator, teachers and students. In contrast, the teachers in Nelson Street Girls' reported comparatively low levels of implementation. Nelson Street Girls' was a high achieving school in the Port-of-Spain area and Bien Venue was a low achieving CETT school in La Romaine, on the outskirts of San Fernando. The choice of schools was based on the concern that the achievement context might lead to differences in the intensity and fidelity of implementation at different sites. The intention was to confirm the integrated model generated from the quantitative and qualitative evidence in Phases 1 and 2.

Bien Venue: A low achieving, medium to high implementation site

CAP as our duty

Teachers in Bien Venue understood the CAP to be a system of multimodal continuous assessment. Most of the teachers had fully embraced this concept and had modified their own reporting so that it captured some of these areas for the termly tests. The staff conducted weekly tests religiously on Fridays. The teachers felt that the CAP concept was useful because it gave them multiple pieces of evidence about learning, enabling them to trace and monitor student progress, allowing some intervention. The concept of holistic assessment also resonated with the philosophies of most of the teachers at this site.

The teachers felt that they had benefitted from the multimodal approach used in the CAP which was an advantage to the predominant use of paper and pencil assessments in the traditional system. The teachers also saw the value of the student and parental involvement in the project and believed that is had improved the quality of interaction and ownership of learning. However, the traditional design of the report book still focused on scores and grades rather than on criterion-referenced levels. Nevertheless, the report book did capture a wide range of cognitive scores and work habits.

Pros outweigh cons

These perceived benefits were important because even at this site, teachers carefully weighed the pros and cons of the innovation and CAP had several challenges such as record keeping and historical opposition from the Teachers' Union. The level of record keeping in CAP was high enough to make the innovation untenable. Therefore, it seems that at this site, the benefits experienced outweighed these significant challenges. Bien Venue's current favourable experience with CAP was partly due to the fact that they had stuck with the innovation so that some of the more difficult tasks had become easier with time. For example, teachers found that initially parents were a bit apprehensive because students obtained low marks in the projects. In response, the school had made a concerted effort over the years to inform parents of the need for their involvement in the projects and this appeared to have helped.

The value and threat of time

Time, however, could be both a facilitator and an inhibitor. It appeared that the innovation was not supported externally by the training college experience and so new teachers coming in primarily learnt the procedures on site. In some instances, as teachers reflected on the innovation, the overall purpose could be lost as noted by one young male teacher:

> **I think the idea of CAP is a good one . . . but nevertheless, over the years, it [the purpose] kind of became fuzzy. While we were doing the work and everything in the end you have a booklet that reaches up to standard 5 and yet is not contributing to the SEA. It comes like you are doing all of this work and the work can't even be transferred over to secondary school.**

Perhaps the most significant factor that contributed to the routinization of the innovation at the site was the initial in-school sharing and training, since this had been a CAP school at the start. However, the advantage of this site-based training could also be limited because of high turnover of staff and the limited reach of this particular version of "train the trainer", as teachers discussed:

> **[Teacher 1] This was a pilot school so we had to do workshops [Teacher 2] We had teachers like {name called} and that is years now. And somebody else [Teacher 3] And that is one of my problems too because when I came into the school the training had ended and they had progressed already. And I remember that at the end of the term there were some pages that I had to do already. I wasn't even informed about the book as well. [Teacher 4] My experience was that I started teacher after CAP was implemented. It wasn't this one-I went to the school, they gave me a booklet, they said, "this is CAP, do it". It was up to me to read it and figure out what. So I had to inform myself of what to do so when I entered the classes I followed what other teachers had done. [Teacher 1] I don't think anyone even came back and told us about CAP.**

Part of the problem seems to be that the programme designers and system planners did not recognize that it is the schools (organizational learning) and not just the teachers that learn and therefore implementation and training must be primarily-site based, rather than directed remotely from the centre. As evidenced here, the model of training used in CAP might not penetrate deep down into an institution with knowledge transferred only through conversations between teachers and modelling.

The leadership at the school strongly supported CAP and made opportunities to allow teachers to share their ideas. Several former principals were also intimately involved in CAP, even giving suggestions in line with the basic principles of instructional leadership.

Everyone does it, but concerns and tensions remain

Thus several elements of the innovation had been retained within this institution at many levels, including the early school, as one middle aged female teacher noted:

> **At the infant class level where I have been from the start, we use the beginning of the CAP book, which assess almost everything about the child. I don't think they left out anything. So although there is a lot that you have to know about the child, to me it is very good because it helps you to understand the child and you can see where the child is [situated] at [compared] to where they are moving towards.**

Despite the teachers endorsement of the CAP as a process, there were still several concerns and fears similar to that observed in the Phase I schools. For example, teachers were very concerned about why the CAP scores were not used to support the high stakes examination at the end of the primary school cycle as promised. It would appear that this was still an important consideration. It seemed important, then, that policy considerations be finalized before interventions are initiated. The promise of a contribution from CAP to the high stakes assessment did not prove to be an advantage in the long run.

Another concern of teachers was the use of the capricious absolute cutscore of 60% in grading, which was not criterion referenced. This score could be demotivating for students who did not do well on the projects. In any case, the focus on simply using scores rather than on criterion-referenced grades suggested that significant elements of the old system had been retained within the innovation.

Fidelity sometimes rings true

There was some level of fidelity at Bien Venue and teachers understood to a degree the concept of formative assessment, especially with regards to adjusting teaching in response to student learning issues. This is captured in the response of one teacher to a question about the role of teaching-learning and assessment:

> **When you assessing the child, you checking to see what they have learnt and if you need to change your teaching strategy to accommodate weaker children because as they say not every child learns the same way.**

Teachers also understood the role of techniques such as re-teaching topics and item analysis in improving learning. However, there was no mention of peer and self assessment or more refined elements such as assessment as learning. On the other hand, there was a belief that continuous testing and competition was beneficial in itself.

An examination of the many artefacts collected from the school confirmed that there also was some use of the special education referral system, with some evidence used. However, the system did not make the best use of data, which would have been expected in a school recording data on student performance continuously.

The teachers understood that CAP had been lost in other part of the systems and that it was their own sense of duty and appreciation for some elements of the innovation that had sustained it up to then. They felt that their own efforts were never really supported by the ministry of Education and were quite surprised that the Ministry still supported the intervention.

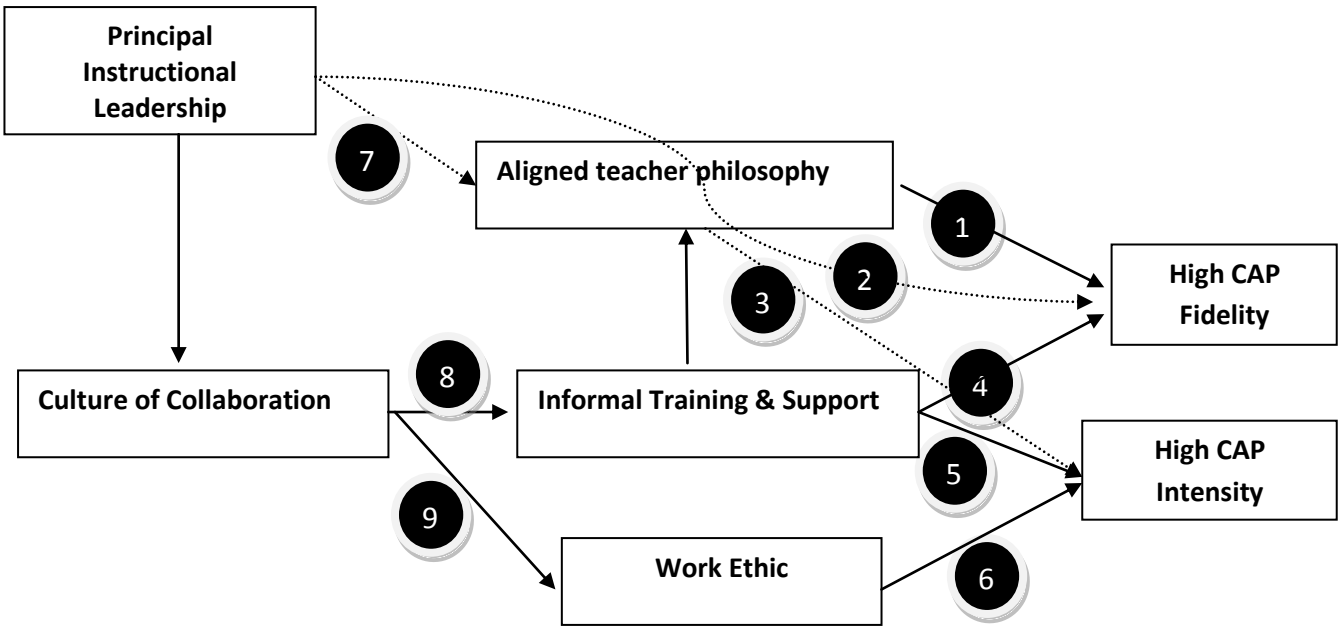


Figure 55: Relationship between key variables in Bien Venue Presbyterian based on qualitative data

Nelson Street Girls': A medium to high achieving, low implementation site

From high achieving to PEP

Although Nelson Street is currently listed as a PEP school because of their high numbers in of students below 30% in 2007, in the very recent past, the school had performed in the top tier in the educational district of Port of Spain and Environs as measured national assessments of educational achievement for the period 2005 to 2007. It was therefore originally classified as a high achieving Port of Spain school using the API indicator. This indicator is based on the numbers of students meeting or exceeding the standards in mathematics and language.

Stepping into CAP

The majority of teachers interviewed understood some of the basic concepts associated with CAP but were often fuzzy about its practical application in use. Additionally, there were very few artefacts available for verification of practice. There was little evidence that the many new teachers coming into the school were supplied with information formally or informally or material on CAP. Additionally, unlike in Bien Venue, the CAP was not applied within the early school, but instead most of the practice was concentrated within standards 1 to 5. CAP booklets or their adapted forms were not in common usage and there was no system for sharing knowledge on the approach, neither was there a shared philosophy among the teachers.

Teachers claimed that CAP booklets were still in use in Nelson Street Girls, but there was no strong institutional memory or knowledge of how the programme got started during the interview. Of the eleven teachers interviewed, only one vaguely remembered the introduction of the innovation in the school. She recounts:

> **[Interviewer] Has anybody been here from at least ten years ago? Did the principal introduce you to the programme? [Teacher] I don't think that the person was extremely clear on what was expected. Miss just gave us the white books to follow. But to say that is definitely what is to be done. . . . There was no such training.**

There was also little subsequent training to support the implementation of the innovation within the system or within schools once the pilot phase was completed. In any case, only a few teachers appeared to have had a philosophy that was attuned to monitoring student progress or focusing upon remediation and so much of the essence of CAP would have been lost as implementation progressed. One teacher, whose views appeared to be aligned noted:

> **I like the place in the CAP booklet where you have to put the required intervention. So if you chat with the parents, you have to write it down. Sometimes you tend to forget. . . . You can see if that child has progressed over the years.**

Another teacher talked about using the performance of the children on items in weekly tests to re-teach and re-assess topics.

Give me CAP, but leave that out

However, more often, teachers accepted some parts of the innovation but rejected other components that were incompatible with their philosophy of teaching-learning. For example, one teacher said:

> **[Teacher] I have been using the CAP for a while, but now I am in second year as well. . . . [Interviewer] Have you found any challenges in using it?)] . Teacher: The only thing I don't full out is that social behaviour thing in the back-I think there is some social behaviour thing at the back of the book. I will be very honest. [Interviewer] Why don't you full it out?] [Teacher] I just don't know why. But the projects, when the students bring it in, I will put it in. The monthly marks, the term marks, I will put it in. So probably that's the reason-It's not part of the final mark.**

From this teacher's perspective, then CAP was simply a recording strategy, a procedure to calculate an end of term mark, not a device for holistic assessment and remediation as seen among the teachers in Bien Venue.

Teachers claimed that assessments were conducted both weekly and termly but teachers did not always make use of multiple formats. There may have been a tendency, then, to over-emphasize paper and pencil tests for the weekly tests.

> **[Interviewer] Miss, how do you normally assess the students? [Teacher] Well to be honest, most times its pen and pencil. But at times depending upon the topic like in social studies, I say, do a poem. I put them in groups and they write poems or like for Carnival, they did a calypso. I am not very much musically inclined, but the girls in class they like music, drama and dance. So when you tell them do that, it is very good for them. They enjoy doing those things.**

Projects, projects everywhere, but none in the class

As with other schools, the student projects were very popular and used across a variety of subject disciplines. However, most of the projects were simply done at home with school occasionally used for the preparation phase. Thus again, the benefits and opportunities for formative assessment practice might have been lost. One teacher, for example, estimated the percentage of time spent doing the project at home and in school:

> **5% is done in school and that is discussing [the project], the materials you need and what is to be done, then to go home and do it. They may report on the project and get some feedback in class, but it is 95% at home.**

Teachers explained that one of the reasons for relegating the performance assessments to homework were the excessive demands on completing the syllabus and national tests. Thus, they did not see these alternative forms of classroom assessment as complementary to their work; instead, they believed that it competed directly for valuable time with more important activities. One teacher expressed the conflict in this way:

> **I try to finish my syllabus because although the Ministry say focus on this and focus on that, you still have your end of term test. If you say you are a teacher, and I am sure the Ministry is evaluating teachers also, the focus remains on the student's academic performance. So although most of my projects are done on topics within the syllabus-I discuss and I give them information to sort and put together-most of it is done at home-and some children don't do it still.**

Another teacher noted that the syllabus forced teachers into compromising use of the projects, which they perceive as useful primarily for holistic and aesthetic objectives:

These quotes support the finding in Phase I, which suggested that the way the projects were done did not allow for formative assessment, and it seems as well that, at least in some cases, these projects were often disconnected from the syllabus or scheme of work and designed to attach additional experiences, but in a supplementary way.

> **The syllabus is so vast, and I know from the bottom of my heart that we are supposed to cater for the whole child, but can you honestly tell me that a holistic educated child is not going to get into St. Joseph convent. That is it-We have to face facts now and I am sorry, as teachers we have to short-change the children. When we are pushing all the maths and English and the**

> **three Rs and I do think that the project takes a lot of time-it takes way from the teaching time. And there is so much competition for children to do well. So the holistic child is all well and good-It looks lovely in theory and on paper, but our children are competing.**

The pressures to excel in the school system were high, especially in this high achieving school and these pressures were compounded by the ignominy of being classified also as a PEP school. Thus, some teachers at the fourth and fifth standards would logically reduce their focus on these take-home projects, which were already mostly disconnected from the syllabus work. The challenge, it seems, would be for programme designers to integrate skills and competences related to performance assessments in an otherwise overly academic and packed curriculum.

As with the other schools as well, the use of rubrics was very rare, but teachers were adept at using standardized criteria for scoring the projects.

> **[Interviewer] This is an open question –when you do performance assessments, how do you award a mark? (Repeat twice after no answers). Most the time when we give projects we use a marking scheme like content and organization and you allot the marks to certain areas. So the criteria you looking for are in the project.**

After having experienced a recent workshop, one teacher was able to reflect on her own lack of training and reasoned that better training might have supported more efficient implementation of the projects. She stressed

> **I have attended some workshops on projects. I am in the first year department. And the workshops dealt with how to do performance assessments, how to construct rubrics. Really, we don't have any knowledge about these things. You asking people to give projects, but they going on their own whims, what they did in school. So there is nothing to grasp on to. Some of us don't have the degree as yet. We were never taught those things, we were not exposed to them in College, and so we don't know, some of us. It's a great need, workshops and training because that is the way we are going now. Traditional assessment to performance assessment and we are not too au courant about the techniques.**

Good practice exists

At the same time, there were instances of good practice in the use of performance tasks, formative assessment and feedback. In Mathematics and Social Studies, some use was made of performance-based tasks although these were often not effectively designed. However, this good practice was often located in specific classrooms. For example, one young teacher described how she had experimented with in-class projects in term 2 and found that the students did substantially better.

> **I started teaching here in September, so I was given the CAP and I would give weekly tests sometimes three of four times a month. At the end of the term, six of my girls did not do the project. It was as Miss said, we discussed in class only and you give them the rest to go home and do. Because of that, children's grades that first term were not very good. So what I did in my second term, I chose a project to do in class to ensure that everybody getting the project mark and girls that were very weak in the first term, I implemented a few things to assist them. A few motivational things, like motivation charts; put them to work in groups with the stronger girls in the class, give them individual work like the spelling words. . . . Some of my girls do remedial reading and I tried to liaise with the remedial teachers to see how they are doing in her class, and I tried group reading and things like that.**

This teacher was able to practice some elements of formative assessment, by providing scaffolding and motivation for her weaker learners. Another young teacher described how she provided feedback to her students based on informal observation and interactive formative assessment.

> **Even when I am teaching and I realize that they are silent or what not, I will call them aside and have a one to one chat with them. And I do that a lot. When it comes to test and I realize that they performed well this week and the next week I realized that the marks dipped, I will call them and ask them what is going on and have a rapport with them. I always do that**

Unlike at Bien Venue and St' Joseph's Government, however, these practices were not collaborative, but isolated and it might be that they would not be sustained in the current school culture.

Unsupported burdens and tensions

Like in Bien Venue, teachers also felt somewhat let down by the fact that CAP had not been used to supplement the SEA mark. It appears that some teachers may have begun the process in the genuine hope of reform to the high stakes selection process. Another burden was the demand on time, which itself was often at a premium for teachers. Some teachers believed that it was not always possible to advance the process further because of the absence of nonteaching contact or planning time. A final issue was the lack of resources with teachers having sometimes to supplement what was available. However, the implementation of school development planning did provide an opportunity to better organize the planning for resources.

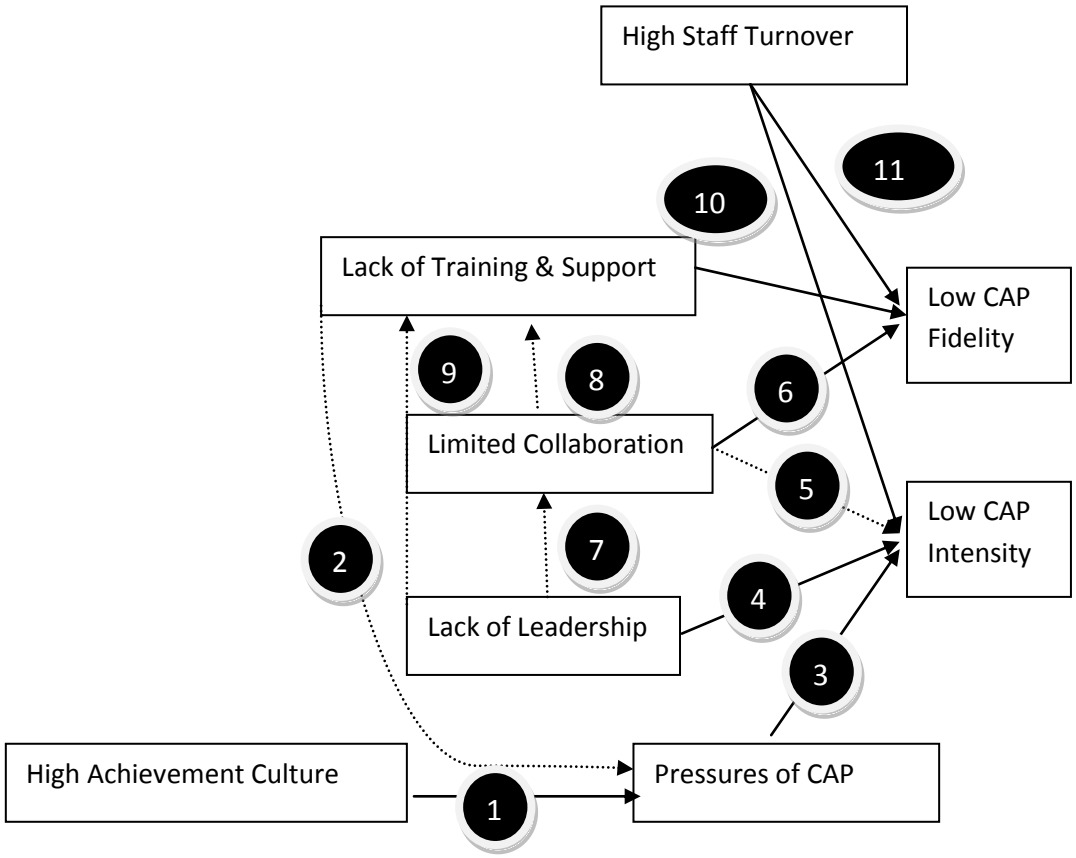


Figure 56: Relationship between key variables in Nelson Street Girls' based on qualitative data

Summary Analysis

Figure 56 summarizes the main causal links established in the comparative analysis of the two cases. There were contextual factors that influenced the uptake of the innovation and the level of intensity and degree of fidelity. The common factors were collegiality and collaboration, which had added importance in the context because there were no formal training programmes to support CAP at the building site. Collaboration in the form of sharing and modelling was a primary tool in the spread of the innovation at the site. This process could be constrained by high staff turnover over the period of implementation. Leadership proved a critical factor in ensuring such collaboration, both of which created a climate of support that allowed the innovation to scale-up. Nevertheless, fidelity remained relatively low at both sites primarily because of the absence of knowledge and the lack of training within the system.

Implications

Table 33 analyzes the different recurring patterns found at each site. In Bien Venue, teachers' worldviews were aligned with the CAP philosophy and principal leadership was very supportive of CAP. The culture of collaboration and work ethic fuelled extra-role behaviour and shared learning. Leadership and collaboration were lacking at Nelson Street Girls and the impact was accentuated by the high turnover. The high achievement culture further

constrained teachers' efforts to experiment with new approaches. The case studies partially confirmed the quantitative modelling although it gives greater prominence to the role of context. Building site leadership seems critical along with collaboration captured in the quantitative study in the form of professional learning community. The cases show that training can have impact if structures are in place to ensure that it is developed and applied at the site. Clarity of the innovation seems to be a recurring theme impinging on continued tensions and concerns for the innovation.

Table 33: Summary of causal relationships identified at the school sites

Bien Venue	**Nelson Street Girls'**
Teachers' worldviews aligned with CAP philosophy	Lack of leadership and limited collaboration lead to low CAP intensity and fidelity
Principal leadership supports CAP	High staff turnover and lack of collaboration limited intensity and fidelity
Culture of collaboration supports informal training and support which leads to shared purpose and CAP implementation	High achievement culture lead to several pressures and burdens that limited CAP use
Culture of collaboration supports shared work ethic, which maintains CAP intensity	

Integrated Findings (Meta-inferences) (With recommendations)

Meta-inferences were generated from the qualitative and quantitative findings and are discussed along with recommendations for education improvement below:

- *Programme strength was highly variable across sites, dependent upon both teacher belief systems and contextual variables such as achievement status.*

1. Both the qualitative and quantitative evidence confirmed that there was a high degree of variation in implementation strength or intensity across school sites. The qualitative data suggested that achievement context might be an important factor in this variation and the quantitative modelling and comparative analysis also partly confirmed this. It appears that the low achievement contexts of some schools could make the burden of assessment even more difficult. At the same time, there could be pressures in high achievement contexts to do frequent testing. Thus, the nature of the burdens in low and high achievement cultures appeared somewhat different.

2. The quantitative study suggested that teacher receptivity factors such as CAP attitude, perceptions of assessment, and individual extra-role behaviour were closely connected to the quality of implementation of the CAP. However, there were also critical organizational variables such as of organizational group behaviour, organizational innovativeness, readiness for change, and professional community which facilitated higher user receptivity and also directly influenced CAP use.

ORGANIZATIONAL PRIMING

1) *Focus building leadership on creating and managing change*
2) *Develop team-based structures in schools.*
3) *Organize professional development so that it is continuous and embedded in the work of teachers in the classroom*
4) *Reward innovation and experimentation at the district levels*

3. These findings suggest that the MoE cannot expect to create change remotely within school sites. Instead it must make the effort to improve the capacity of schools to cope with change. I would suggest a decentralized approach at the level of the education district and a strong site based change strategy. I also proposed a priming strategy, which would focus on making all schools in a district ready for change. The strategy includes the development of change based site leadership, team-based structures, ongoing professional development based on selected themes related to teachers' work, instructional coaches, and a focus on experimentation and innovation promoted at district level.

- *Once an innovation is installed, the pathways to change are variable and sometimes even regressive, strongly dependent upon the quality of leadership and key organizational characteristics.*

4. Essentially the quantitative evidence suggests that most teachers and sites were nonusers, even if they appeared to be at different levels (positive versus untrusting). Awareness concerns remained high even in schools that were judged to be high implementation sites. The qualitative evidence confirmed that the change process was not predictable and often varied for individuals and across different sites.

5. The rate of change at a site might be slowed substantially if there was a rapid turnover in staff. On the other hand, the qualitative evidence also suggests that in some cases there could be a return to early concerns, if support proved deficient. I would suggest that greater attention be paid to early implementation, with careful monitoring of an innovation. DCD officers should be (re)trained to monitor innovations using a simple model such as CBAM. This monitoring would lead to corrective actions to ensure that the innovation take root. However, the main manager of change must be the principal and training in this regard is critical. All the sites which had high levels of implantation were lead by principals who were trained in the CAP. When principal possessed this training, some level of routinization became apparent.

6. The quantitative evidence confirmed that users' initial concerns about the structure and function of CAP persisted throughout the life of the innovation and could still be notable ten years after initiation. Concerns about information do not go away it seems and in some cases, there was simply not enough knowledge about the nuts and bolts of the innovation to expect otherwise. Rather than preparing general material, the Department of Education Services might play a greater role in supporting system change by publishing practical material on the use of the innovation. It seemed paradoxical that several countries in Africa published significantly more material on continuous assessment than did Trinidad and Tobago.

7. Although CAP was initially installed with much fanfare and public marketing, support and resources was not always available at the building sites and this might have impeded widespread adoption. To reduce such variability in future reform efforts, the implementation infrastructure at head office and the districts must be greatly strengthened to monitor and direct the change process. This structural change must be supported by a comprehensive and flexible site-based change strategy. That change strategy must have, as a key element, the development of site-based assessment leadership (Guskey, 2009). This is the only mechanism that will ensure scale-up of an innovation at a site.

- *Teachers' knowledge, beliefs and attitudes were strongly correlated with implementation, strength and fidelity at sites.*

8. Both the qualitative and quantitative data confirmed that teachers' receptivity to change was a key factor influencing change. Teachers are the critical element in the intervention because teachers conduct classroom assessment. If teachers have concerns about an

innovation and they are not addressed, the innovation will not be implemented or implemented with fidelity.

9. Thus the low levels of assessment literacy and conflicting conceptions of curriculum, teaching and assessment were important sources of resistance and facilitation. The quantitative data suggested that some conceptions of assessment such as "assessment improves education" were critical elements in CAP implementation, strength and fidelity.

10. It may not be possible to install CAP effectively with the current understandings and practice of teaching-learning in the system. It seems advisable, then, to develop modern understandings of curriculum, teaching and assessment that are better aligned with the concept of continuous assessment. Notably, some countries have installed formative assessment as part of wider curriculum change and this may work to assure compatible teacher beliefs and intentions.

- *Fidelity was universally very low, related to inappropriate scoring approaches, misuse of performance assessments, the absence of feedback to students, and the neglect of data use strategies and routes.*

11. The qualitative evidence pointed to the fact that implementation strength differed from implementation fidelity. The quantitative study modelled CAP use and aspects such as the provision of feedback and use of multimodal assessment. The data showed that different factors explained these different variables supporting the original hypothesis. In the two qualitative studies, we found that even at sites where major CAP activities were routinely conducted, critical elements of both formative assessment and data use might be absent in the majority of classrooms. For example, there was no clear protocol for data use and project work often did not facilitate student-led assessment practice or feedback to students.

12. Likewise, peer and self-assessment were universally absent. Indeed, students were not adequately involved in the assessment process, which was kept apart from teaching-learning. Thus, although some schools reported relatively high levels of programme strength, in both the quantitative and qualitative studies, neither the essence of formative assessment nor effective use of the data was observed.

13. Consequently, quality outcomes could not be assured even when there was high programme strength. In the future, fidelity must be closely monitored when implementing innovations such as CAP because the quality of outcomes in an innovation such as CAP depends upon trueness in implementation. Several curriculum and violence prevention innovations will also make the same fidelity demand, so it seems useful to enhance the skill of the current evaluation unit in the DERE to serve this purpose.

- *Despite various concerns and contextual difficulties such as the lack of resources and support, teachers were able to complete most CAP tasks in schools with strong, informed leadership.*

14. In the Phase 1 exploratory study, two of the three schools with the highest implementation scores both had very effective leaders who were trained in the initial pilot programme. We were extremely impressed with the quality of their leadership and the level of CAP activity in the schools. Unfortunately, both schools were also high achieving, which meant that we could not truly separate the effects of context from the leadership at these sites. However, several medium achievement schools also reported relatively high programme strength when leadership was effective. In the Phase 3 explanatory study, the medium level implementation school had leaders who had been trained in the CAP pilot programme.

15. This role of assessment leadership on the success or failure of CAP in a school was also confirmed in the quantitative study with the indirect pathway of leadership of the CAP on CAP use often through other organizational priming variables. We had hypothesized that the influence was direct; however, the path model for CAP use suggests that the impact of this kind of leadership was on facilitating extra-role behaviour. It could be that such informed and supportive leadership enhanced teachers' ability to use the innovation by motivating the teacher and acting as a buffer for the multiple stressors of school life.

16. The qualitative evidence in Phase 1 and 3 also pointed to knowledge of the CAP assisting these principals to lend instructional and pedagogical leadership when necessary. However, this kind of knowledge was not readily available to everyone. Therefore, effective leadership must become the centrepiece of sustainable site based change and as such, greater efforts should be focused on training principals.

Towards a Mixed Methods Integrated Model

The original model organized the antecedent variables into three groupings, (1) teacher variables, (2) organizational variables, and (3) environmental variables. It was hypothesized that teachers' knowledge, attitude and feelings might be more directly related to the fidelity and intensity of the innovation, but this was not always evident in the quantitative modelling data. Nevertheless the teachers' receptivity to change remains a key variable for CAP use. A number of key organizational variables such as readiness for change and the ability of the school to promote innovation had both direct and indirect effects. For change to occur, it appears that an organization must be primed and prepared for change.

We hypothesized that teachers' attitude towards change might be a critical variable to which other teacher belief variables might be linked. The evidence suggests that the key variable is extra-role behaviour with both organizational and user variables directly impinging upon this personal willingness to go beyond the boundary in efforts to help students. These causal links are illustrated in Figure 57. The achievement climate was expected to be an important contextual variable, perhaps impacting on group organizational citizenship behaviour, collective teacher efficacy and individual extra-role behaviour. In the quantitative modelling, there was a negative relationship between academic achievement climate and organizational innovativeness but a positive relationship with group organizational citizenship behaviour. For CAP use, the achievement status also has a direct influence.

High achievement status could have provided a united vision in some "leaderless" schools, galvanizing teachers' efforts. It could be too that some teachers rationalized their effort using a cost-benefit analysis and were more willing to give the additional efforts if they were rewarded by greater student success. Thus, project work with low achieving students would be considered too difficult and completing Cumulative Record Cards too tedious in schools with many learning and behavioural problems.

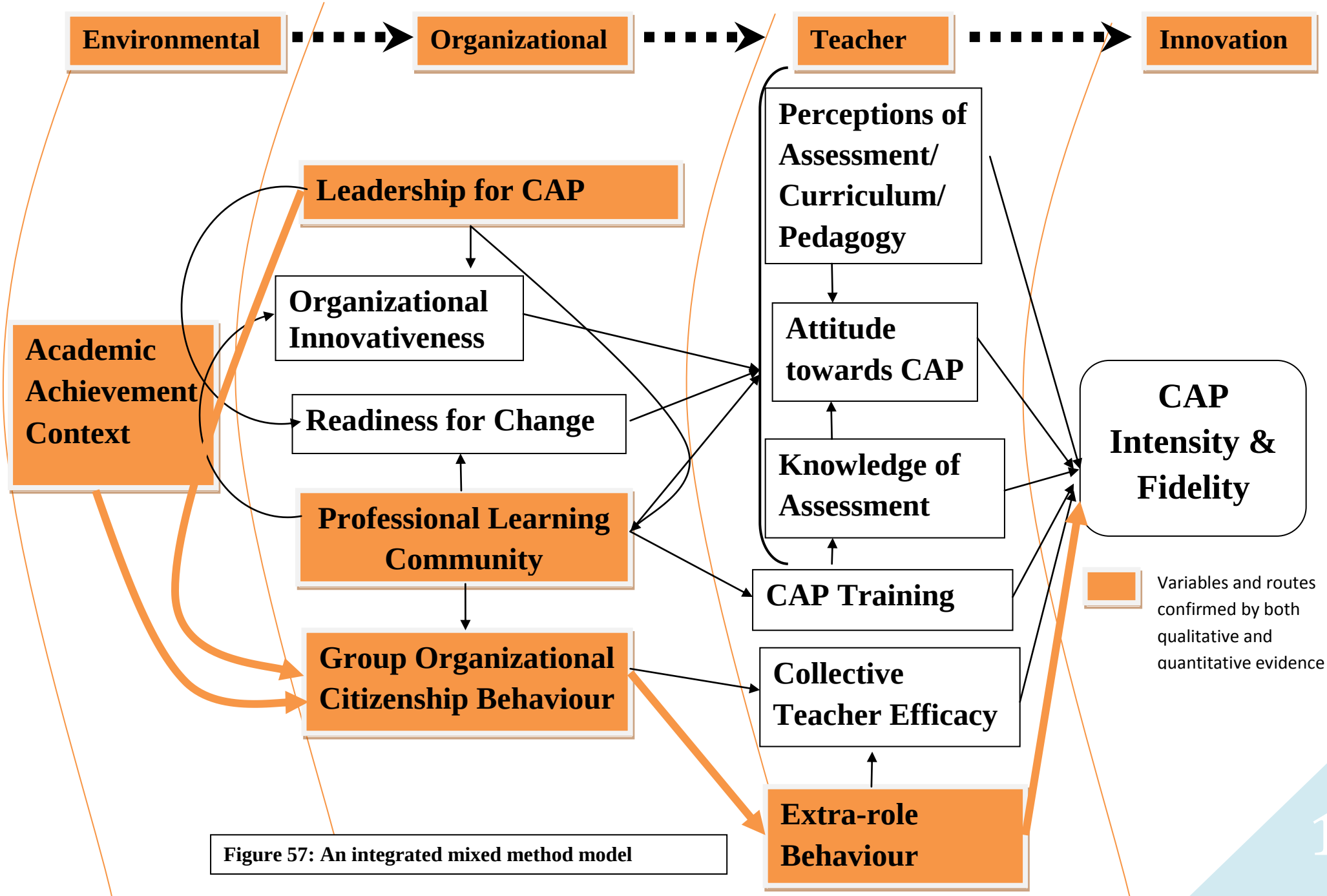


Figure 57: An integrated mixed method model

Future Solutions

I. **Reconfigure CAP as a simplified innovation (constituent parts) and focus on the key elements of formative assessment and data driven decision-making as the most likely factors to effect change in student learning. Re-emphasize a SEN procedure that builds on but is separate to CAP.**

II. **Encourage the integration of formative assessment with curriculum, teaching, and learning.**

III. **Clarify the different routes and multiple uses of data from CAP.**

IV. **Develop an administrative and organizational structure for the efficient implementation of CAP, including the provision of training integrated with the innovation.**

V. **Develop a training schedule for data-driven instruction and link this to the enhanced assessment system.**

VI. **Develop a site-based change for change strategy, which includes strategies by the Office of the DSS for priming or preparing schools for organizational change.**

VII. **Develop a nation-wide overall policy for the primary school that promotes a balanced, coherent, comprehensive assessment system that provides continuous data, but with an emphasis on formative assessment practice.**

VIII. **Clarify and simplify the primary school assessment cycle with a focus on formative assessment and appropriate data use.**

IX. **Construct a timeline for implementing the overall change strategy**

X. **Reconsider any intention to combine high stakes assessment with a CA programme. Conduct a cost-benefit analysis of any new system that either combines different assessments or makes use of multiple purposes for a single assessment.**

Detailed Recommendations & Best Practice for Future Solutions

---

I. Reconfigure CAP as a simplified innovation (constituent parts) and focus on the key elements of formative assessment and data driven decision-making as the most likely factors to effect change in student learning. Re-emphasize a SEN procedure that builds on but is separate to CAP.

II. Encourage the integration of formative assessment with curriculum, teaching, and learning.

III. Clarify the different routes and multiple uses of data from CAP.

IV. Develop an administrative and organizational structure for the efficient implementation of CAP, including the provision of training integrated with the innovation.

V. Develop a training schedule for data-driven instruction and link this to the enhanced assessment system.

VI. Develop a site-based change for change strategy, which includes strategies by the Office of the DSS for priming or preparing schools for organizational change.

1. We present a model for a reconfigured primary school CAP in Figure 58, which focuses upon installing the key elements of formative assessment and data use, but also incorporates as part of the innovation, professional development for assessment literacy. The model also considers the provision of onsite support at school sites. This configuration is similar to that installed in Western countries such as in Canada (Ontario), earlier considered. It acknowledges and recognizes the difficulty teachers might have in adopting formative assessment because it conflicts with their philosophy of teaching and is very different to what they currently do in the classroom.

2. I would argue that by itself, formative assessment is complicated enough to demand significant levels of professional development training. In the context of Trinidad and Tobago, professional development in assessment literacy becomes critical. For example, in the area of mathematics education, Webb et al. (2004) has noted:

> Although research supports the contention that formative assessment benefits student learning and can be used to facilitate learning with understanding, many mathematics teachers show limited understanding of the ways in which formative assessment can be incorporated into their classroom practices. As a result, teachers often have difficulties in making didactical decisions based on their students' work and therefore defer instructional activities to the sequence of activities in a textbook. Students in such classrooms are left with incomplete

> information about their progress. They frequently find themselves at a loss to self-assess what they know or don't know, and they continue to apply and reinforce faulty mathematical conceptions (p. 224)

3. I agree that high quality professional development for teachers must be team-based (Stiggins, 2000), providing individual practice on feedback, rubric development, and the conduct of pre-assessments. Therefore, part of the priming process is for organizations to develop a higher level of collaboration in professional matters, possibly by developing professional learning communities. A modern programme of professional development for assessment literacy would include recent work on training in this area (Lukin, 2004; Webb, 2009). The use of site-based training in teams would be critical. However, such training must be guided by national and school policy that clarifies the different role of assessments in the system.

4. Although the focus of the repackaged innovation is on formative assessment, we would expect that summative assessments might also be used as a tool to improve student learning. Such formative use of internal summative assessments is seen as a secondary component of the reconfigured CAP innovation. The summative assessments might include pencil and paper tests but the focus must be on using this data to identify where the students are in the cycle and creating strategies to ensure that students make progress. The focus on data-driven decision-making becomes critical, then, as the evidence in this evaluation study strongly suggests that teachers do not use internal assessment data even if they might be diligent at collecting information. National standards and school goals must govern what teachers do in assessment within each classroom. The national assessments of educational achievement are connected to the classroom primarily because they provide performance standards that can guide the use of internal summative and formative assessments. I believe that it is critical for the units and divisions at the Central Authority to develop local policy on assessment literacy.

5. The important element in classroom assessment is data use and I would suggest a greater focus on data-driven instruction within an overall framework of data-driven decision making in schools and district. Such an approach is aligned with the School Development Model being rolled out by the RDAU and the Cambridge Education consultancy, which emphasizes the use of data to plan and evaluate. Effective use of data involves using different types of data, but assessment data must also be fully explored along different routes. Figure 59 shows the multiple possible data use routes and pathways envisaged and are based on the work of Preuss (2007). The CAP emphasized decisions for remediation and special education needs but these might best be managed in a separate SEN system that must be installed by 2013. Meanwhile, there are other important uses to which the data must be put within efficient data-driven systems, including refining instructional systems.

6. Teachers must understand that different types of assessments might provide different types of information. For example, data from summative and interim assessments would be more useful in school development planning whereas diagnostic and classroom assessments will have greater valence when making decisions about remediation. These

are all examples of real-time data whereas the results of national assessments and the SEA will apply to the cohort of the preceding year. It is illogical for a school to use national assessment data for diagnostic purposes when real-time data is readily available.

7. The idea of individual and collaborative data-driven decision-making becomes critical as some data routes assume a team-based approach to the analysis and use of the data. Certainly interim and summative assessments from school based standardized tests would demand such collaborative activity on the part of teachers across similar year groups and even departments. Whole school improvement assumes that teachers are able to work together. Since data use is a critical outcome, processes, structures associated with data access, and policies for data use must be formulated (Schildkamp & Kuiper, 2010).

8. There is now some theory on how data-driven decision-making might work in schools, but a key element for installing such a system would be to ease the burden of assessment. Technology, then, becomes a critical element. Any proprietary warehousing system to be installed should be able to cope with local demands and needs (queries, etc) in recording and using continuous assessment data. A Student Information System usually cannot fulfil the needs covered by a warehousing system. Unfortunately, Trinidad and Tobago has neither a Student Information nor a Warehousing system. In the meantime, it might be possible to train teachers in the use of generic software for recording and processing data. However, new technology should not be installed separate from a formally developed plan for data-driven decision-making.

9. Implementation is a process, and therefore requires attention over a significant period. Efficient implementation leading to sustainability would require the presence of structures to manage and monitor an innovation such as CAP. A possible model for such an administrative structure is illustrated in Figure 60. The goal of any implementation is a sustainable innovation, implanted at a very high level and this demands efficient organization in the Central Administration (Noell & Gansle, 2008; Owston, 2007). The DERE currently has no staff specifically focused on the implementation and support of classroom assessment.

10. It seems advisable, then, to install an Assessment for Learning Unit to focus specifically on the elements of classroom assessment for learning. Activities of the unit will focus on promoting good practice in classrooms and will link and coordinate with the main implementation agencies, namely the Division of Curriculum Development (DCD), the Division of Educational Research and Evaluation (DERE), and the Teacher Education and Teacher Performance Project Unit (TETPPU). A strong monitoring and evaluation link will ensure that the Unit is responsive to the needs of schools and major stakeholders. The existence of such a unit will ensure that appropriate training is provided in this area and that policy is focused in a balanced, comprehensive assessment system.

11. The role of the DCD in this process is critical. The DCD seems better staffed and another approach might be for the DCD and the DERE to collaborate on developing a unit which will focus on supporting and managing curriculum-embedded assessment in the classroom. Curriculum officers at the District levels might be trained to function as

assessment specialists, supporting the use of formative assessment in the classroom. This understanding of assessment is quite different from traditional models, which see assessment as a separate activity; instead formative assessment is part of the instructional cycle.

12. Such an *Assessment for Learning Unit* might seek to redesign the use of projects in the CAP, focusing upon the critical design characteristics of (1) learning appropriate goals; (2) scaffolding for students and teacher learning; (3) frequent opportunities for formative assessment and revision; and (4) social organization to promote participation and a sense of agency (Fisher & Frey, 2007). Such a unit devoted to classroom practice will create materials such as rubrics and relevant forms to be used by teachers. The unit might promote national conferences and work with TETPPU to review and monitor training in assessment from service providers.

13. The MoE cannot expect to manage and monitor innovations in all schools from central administration. This idea is as ludicrous as imagining that the Greek gods could rule the world from Mount Olympus. A district and site-based focus is therefore critical to any new change strategy and is illustrated in Figure 61. A site-based policy means that the change process is actively supported and managed at different implementation locations, and not remotely from an education district or head office. A key element in this site-based change strategy would be principal leadership and support provided by instructional coaches (Knight, 2009).

14. Instructional coaches will be needed in addition to high quality leadership to manage personal and management concerns during the change process. The instructional coach function may be located in selected heads of departments and coaches may work in clusters of schools. What is required is strong training to support change agent skills along with in-depth knowledge of assessment. An instructional coach will likely have a Master's degree in Education and be strong in one or more content areas, preferably with a content degree. Another important element of the site-based change strategy is the development of Professional Learning Communities, which in this situation might focus upon teams of teachers learning and engaging in the development and practice of formative assessment in the classroom (De Four, De Four, & Eaker, 2009). Team and cluster-based learning must be supported by education district and national conferences on curriculum and formative assessment.

VII. Develop a nation-wide overall policy for the primary school that promotes a balanced, coherent, comprehensive assessment system that provides continuous data, but with an emphasis on formative assessment practice.

VIII. Clarify and simplify the primary school assessment cycle with a focus on formative assessment and appropriate data use.

IX. Construct a timeline for implementing the overall change strategy

15. I would argue that the original CAP problem statement was correct in highlighting the distorting effect of the dominance of the selection/placement examination at Eleven Plus. This distortion (negative washback) results not only from the high stakes nature of the examination, but the vacuum created by poor classroom assessment practice and the low level of assessment literacy among teachers. Under these conditions, any large-scale assessment will be misused, as currently seen in the case of national tests used for accountability. Unfortunately, the CAP documentation and rationale may have further blurred the distinction between national assessments of educational achievement and classroom assessment. These should have different roles and purposes within a comprehensive assessment system.

16. Stiggins (2008) reminded us that with assessment, purpose is everything and there is need at the central level for a total assessment solution. The overall goal, then, must be to achieve a comprehensive, balanced assessment system by clarifying the role and purposes of different kinds of assessment, as illustrated in Figure 62. The CAP documentation does include a graphic that alludes to the operation of a comprehensive assessment system. However, this diagram did not indicate how the different assessments are linked together in the pursuit of quality education.

17. As argued before, if readiness for learning is the goal, then, classroom assessments must be primarily *formative* in purpose, designed to improve student learning whereas national assessments of educational achievement are large-scale assessments used for accountability and monitoring student learning. Large-scale accountability assessments can be used formatively, but the degree of impact is likely to be much less than quality formative assessment in the classroom. The ideal practice, then, is to use the standards identified explicitly and implicitly in national assessments of educational achievement to guide the construction of meaningful and challenging performance tasks in the classroom. Such processes would demand that a teacher becomes familiar with several modes of assessment with the capacity to provide feedback and use data from different sources.

18. National assessments and placement testing therefore represent only part of the total solution to assessment used to improve education. Indeed, the emphasis in a balanced comprehensive system must be on classroom assessment, even if that practice is to be informed by large-scale assessment. Thus, data from national assessments of educational achievement are vital for ensuring high quality classroom assessments because they promulgate national standards and expectations. For example, a teacher must respond to the judged performance of his school and move on to design formative assessments that

help students bridge the gap identified in national assessment data. However, there is absolutely no reason for teachers to simply mimic large-scale assessment formats, even in their pre-assessments. Challenging and meaningful performance tasks will keep students fully engaged and provide multiple opportunities for high-quality planned and interactive feedback. However, even performance tasks might be based on the standards and achievement gaps identified from the evidence provided in large-scale assessments.

19. Thus, the policy on a balanced, comprehensive assessment system shown in Figure 62 will help to simplify and clarify the assessment cycle in primary schools shown in Figure 63. Predominant in the assessment cycle across the school year will be classroom assessment consisting of both formative and summative assessments conducted by the teacher. There is no need for teachers to record a score or grade for every planned formative assessment in the classroom. However, records of such assessments might be included in planning documents and lesson plans. This information along with data from summative assessments must be available to the teacher and to parents to show progress and to identify recurring difficulties. Standardized or common school assessments in terms 1 and 2 are a vital tool for reporting to and helping parents understand where their child is in terms of norm-referenced progress. However, we expect schools to develop standard-based reporting practices as well focusing on what the student has done. Such reports aid the development of action plans to help students.

20. Low achieving districts and schools will benefit from the opportunity to administer interim assessments in term 2, which provide a measure of students' progress towards the national standards. The common school assessments and the occasional interim assessments administered by the district are certainly sufficient for communicating to parents about the child's progress and information for parents can be further supplemented in parent-teachers' conference with data from classroom continuous formative assessment. Figure 65 shows a timeline for the implementation of these various facets of assessment reform. The change strategy should be first enacted by ensuring that leaders and coaches are fully trained. This will ensure a core of support once the innovation is fully initiated. Materials and guidelines should also be prepared prior to formal adoption and essential structures such as professional learning communities should be enacted early.

The question of combining CA with SEA (or using CA for selection)

X. Reconsider any intention to combine high stakes assessment with a CA programme. Conduct a cost-benefit analysis of any new system that either combines different assessments or makes use of multiple purposes for a single assessment.

21. It is an intoxicating idea to believe that continuous assessment might carry the final solution to continued use of a high stakes examination to select students at Eleven Plus. Early selection carries penalties in both system and individual performance and the full impact must be made clear by analyzing and comparing outcome and system configuration data for other countries. Such evaluations require analysing comparative assessment data and in the light of key outcomes, judging the impact of the degree of differentiation and stratification in the system. Such a study might consider equity, which is a common OECD theme. There are no easy solutions to removing a secondary school entrance examination, as is evident in the examination-oriented Hong Kong.

22. CA scores are used for selection and placement in some African countries, but in these contexts, universal secondary education has not been achieved; Trinidad and Tobago having claimed universal secondary education in 2001. The quality of evidence from these evaluation studies is insufficient to judge whether the systems in countries like Zambia work well. Hong Kong and South Africa's use of moderated school based assessment at the secondary level might provide a signpost, but in the end, early selection even in the best systems might be debilitating.

23. The study of classroom assessment in Western lands suggests that the relationship between continuous assessment and high stakes examinations is often not synergistic, especially in contexts of low assessment literacy (Harlen, 2005). Washback theory and models of test preparation for high stakes test suggest that high stakes purposes will inevitably exert an undue and negative influence on classroom assessment practice (Wall, 2000, 2005). The issue of using continuous assessment for high stakes placement created such a tension in this study, explicitly mentioned by several teachers. Teachers had expectations and some had doubts about how such a system might function.

24. Test justification and test validation theory suggest that the purpose of testing must be clarified and validated. A system in which continuous assessment is used for placement is a multiple purpose assessment designed to both enhance student learning and ensure valid placement decisions. Baker (2005) provided a useful analysis when considering the use of classroom assessments in high stakes systems of accountability. She argued:

> If teachers' judgments are to count in an accountability system, certain requirements need to be met. The first of these is the quality of teacher

> assessments, in particular, the degree to which their assessments adequately address cognitive demands and content in standards and the evidence that their estimates and scoring of student work are competent. Second, from the policy perspective, do teachers have a persistent conflict of interest such that their own views ought not to be employed in a system that has consequences for their workplace or remuneration? Third, can their assessments still function to provide instructional guidance for immediate action with students? Fourth, can teachers combine their own measures and those externally given in order to revise instruction in subsequent cycles? (pp. 367-368)

25. If we take the first requirement, we recognized that in the current scenario, the CAP projects certainly do not meet the criteria of quality and are plagued with multiple sources of construct-irrelevant variance (Haladyna & Downing, 2004). For the second requirement, it becomes clear that teachers do have a conflict of interest related to the prestige and performance of their institution. The third requirement is related to the provision of feedback, which we observed to be limited system-wide. The fourth requirement is also formative, suggesting use of data to improve teaching. Again, we found this to be relatively rare in our current context.

26. In the US, Meisels et al. (2001) found that teachers' judgements of performance on classroom assessments in the early school correlated with standardized tests. The work of Harlen and her colleagues on the validity of teachers' judgements in the UK is also very instructive. Apart from external moderation, several other strategies may be used to enhance the quality of teachers' judgements. The Queensland model in the senior school is heavy on assessment literacy. Generally, the conclusion is that dependability depends upon the level of specification and the provision of criteria (Harlen, 2005)

27. Easing the burden of high stakes test by relying on teachers' summative assessment is a wonderful idea on paper but not in practice, as Harlen (2005) noted:

> Unless teachers are prepared for taking advantage of the autonomy that is theoretically available to them, the tendency is for them to interpret ongoing or continuous assessment as a series of tests. Although these are teacher made, they tend to emulate the form and scope of external tests. This seems to be particularly so when the teachers' assessment is a component of a summative assessment with the remainder (often more than 50%) coming from an examination (p. 249)

28. The issue, however, is not just one of trustworthiness, dependability, or validity, but it is also one of fairness in the context of the high stakes decisions to be made. We cannot compare the use of classroom assessment scores in such low to medium stakes contexts with the proposed use of classroom scores to make critical high stakes decisions about the placement of students. Teachers are likely to respond differently when high stakes decisions are attached to their judgements as noted earlier. Another factor to consider is that such a system will operate at the first cycle of schooling rather than in the secondary school like successful systems in Queensland State, Australia.

29. The most favoured option for the role of continuous assessment, selected from the list in Nitko (1995), may be to separate continuous assessment from the high stakes placement process. Continuous assessment may still be given value through the certification process contained in the Cumulative Record Cards. At the same time, the quality of the placement process may be further enhanced by employing a multimodal strategy or using different assessment at different times. Nevertheless, even the perspective that continuous assessment will improve the validity of the selection process is quite flawed because assessments are not neutral tools. A better approach might be to consider differential placement through the lens of differences in "opportunities to learn", which will point to the urgent need to reduce the great variation in classrooms across the country.

30. A second short term option is presented here, assuming that the system of early selection might not be removed within three to five years. There are certainly significant improvements that can be made to the high stakes SEA to make it more efficient at differentiating learners while possibly reducing the washback effect. However, in a comprehensive system, for these changes to work, they must be aligned with reform in the national assessments of educational achievement, which functions as an accountability system. Figures 64 and 65 re-emphasize the need for the different assessments to serve different purposes. In light of the need for more efficient discrimination between learners, it might be useful to replace the writing assessment with a problem solving section in the SEA paper, bearing in mind that teachers can teach to the test even when the target examination is problem solving.

31. The place to experiment with large-scale performance assessment might be in the national tests where it is possible to introduce a writing portfolio at this stage. The assessment task would be in response to a standardized prompt and would involve several writing products illustrating the process of writing over a two week period. Scoring would be a trait rubric. This would remove writing from the high stakes SEA but reinforce writing as a process in the accountability system. The benefit of removing the high stakes writing from the national assessments is the disadvantage experienced by males, low ses students, and students from low quality schools

32. There are important benefits of creating an authentic writing assessment for the national tests. In theory, the use of this kind of assessment in an accountability system may be very desirable as noted by Maxwell (2002).

> A case can be made for moving in the direction of system monitoring that is founded on school-based teacher judgments rather than external standardised tests. School-based teacher judgments offer the possibility of richer, more authentic, more sensitive and more comprehensive assessments of student progress as compared to standardised tests. They also allow assessments to be devised according to modern understandings of knowledge and learning, especially in terms of developing capacities for lifelong learning, developing complex knowledge structures and repertoires, developing personally meaningful and useable knowledge, developing generalisable and transferable knowledge, and recognising the effects of context on performance (p. 9).

33. National assessments of educational achievement do not have to assess every subject area nor do they have to be conducted every year using the entire target population. Matrix sampling and biennial administrations are quite common. If data is needed for individual students, that should be left to the teachers' summative assessment who are responsive for reporting to parents and other stakeholders. The main focus of national assessments should be on language and mathematics, which are critical indicators of learning and the main targets of monitoring. Learning in science and social studies do not need to be evaluated each year, but such measurement should be restricted to standard 2.

CAP Reconfigured

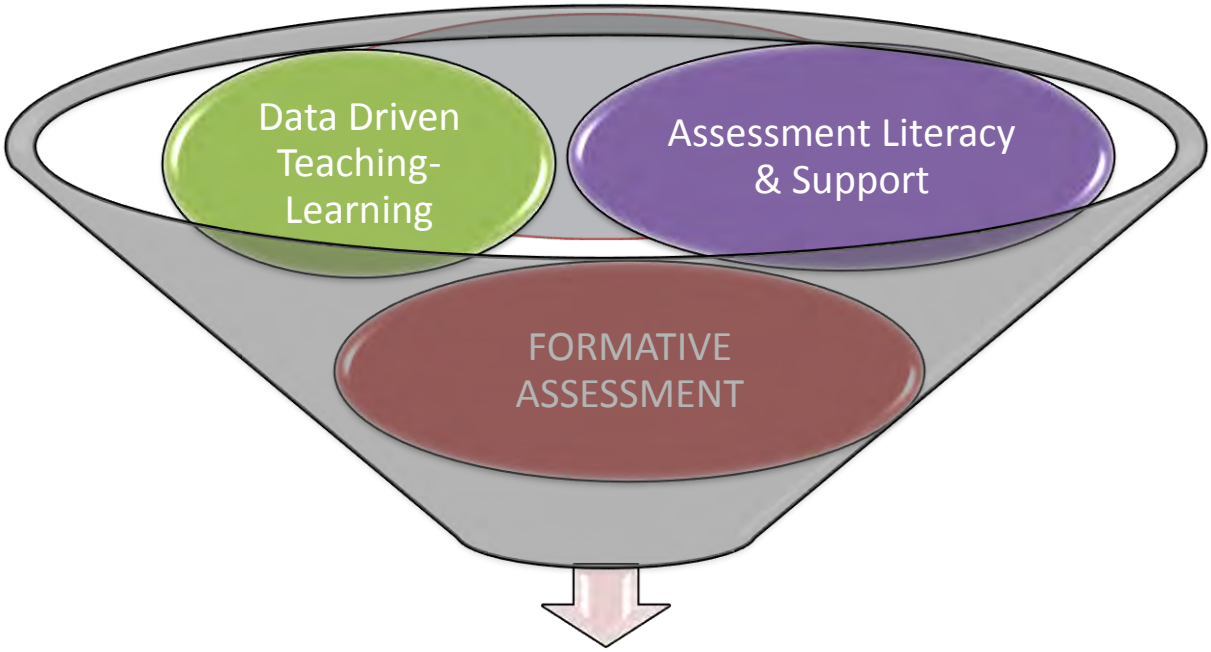


Figure 58: The essential elements in the reconfigured CAP

CAP Reconfigured

Data Routes

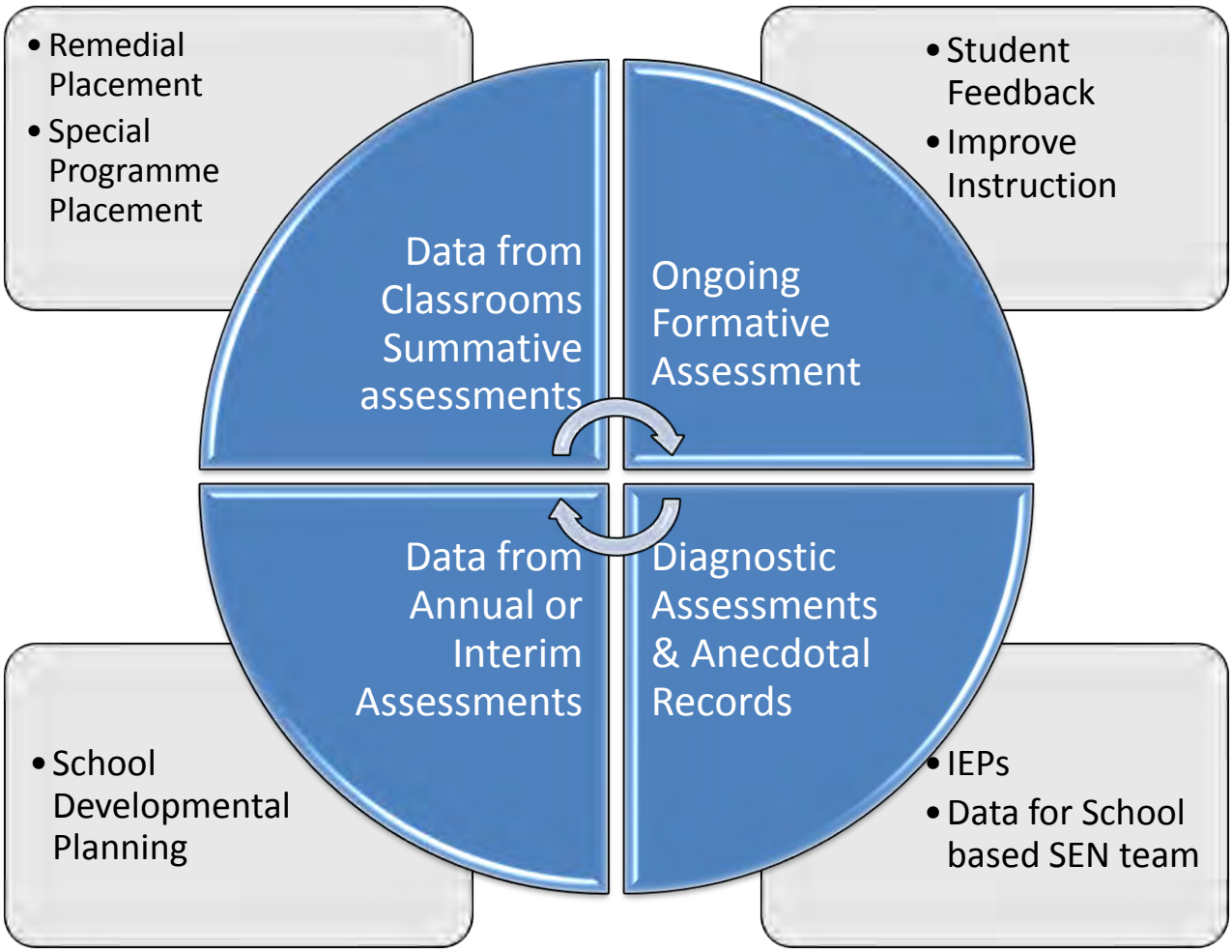


Figure 59: Data routes in the reconfigured CAP

CAP Reconfigured

Organization

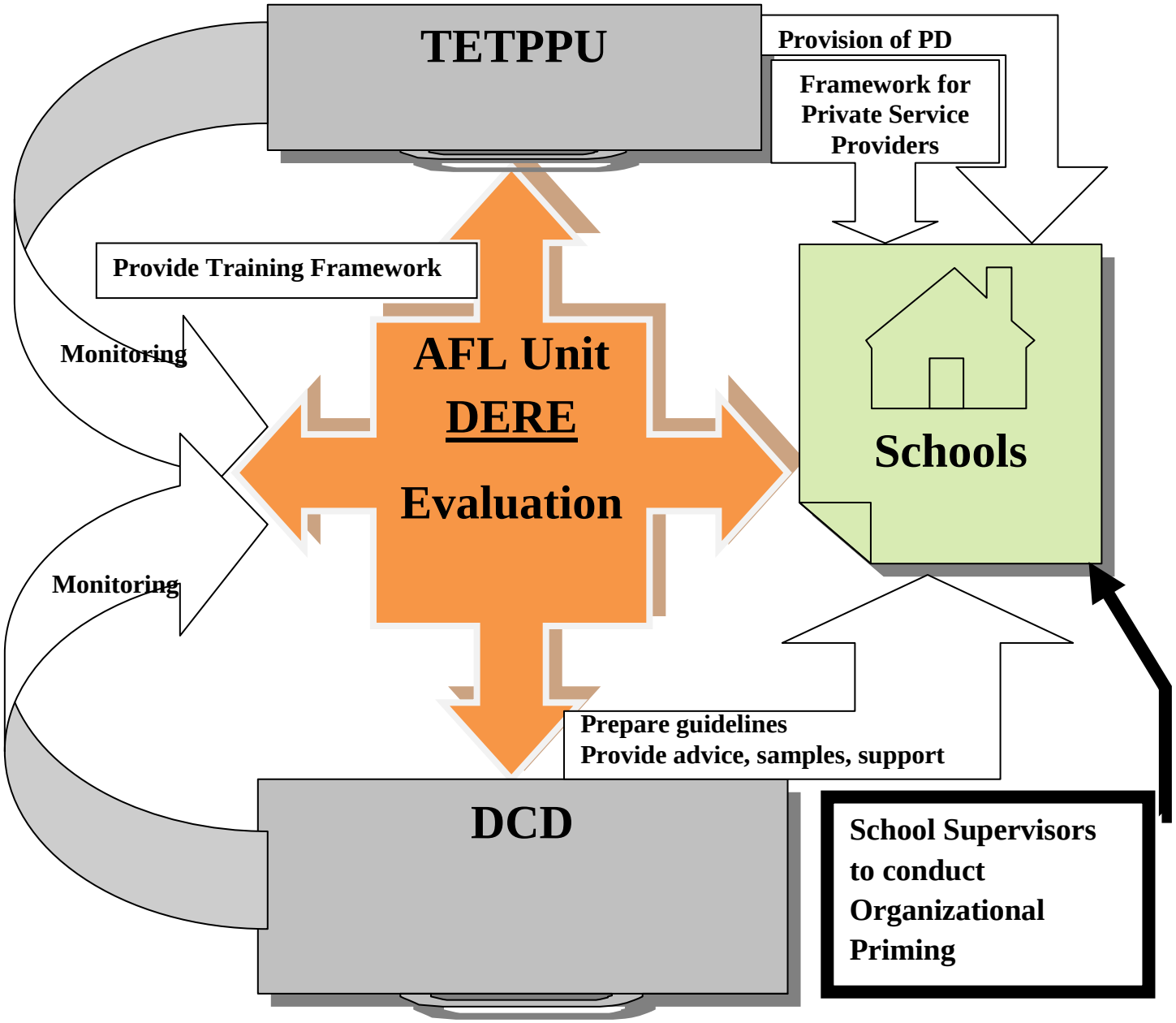


Figure 60: Administrative structures in the reconfigured CAP

Strategy for Change at Sites

Figure 61: Site-based change strategy

Developing a system-wide balanced assessment policy

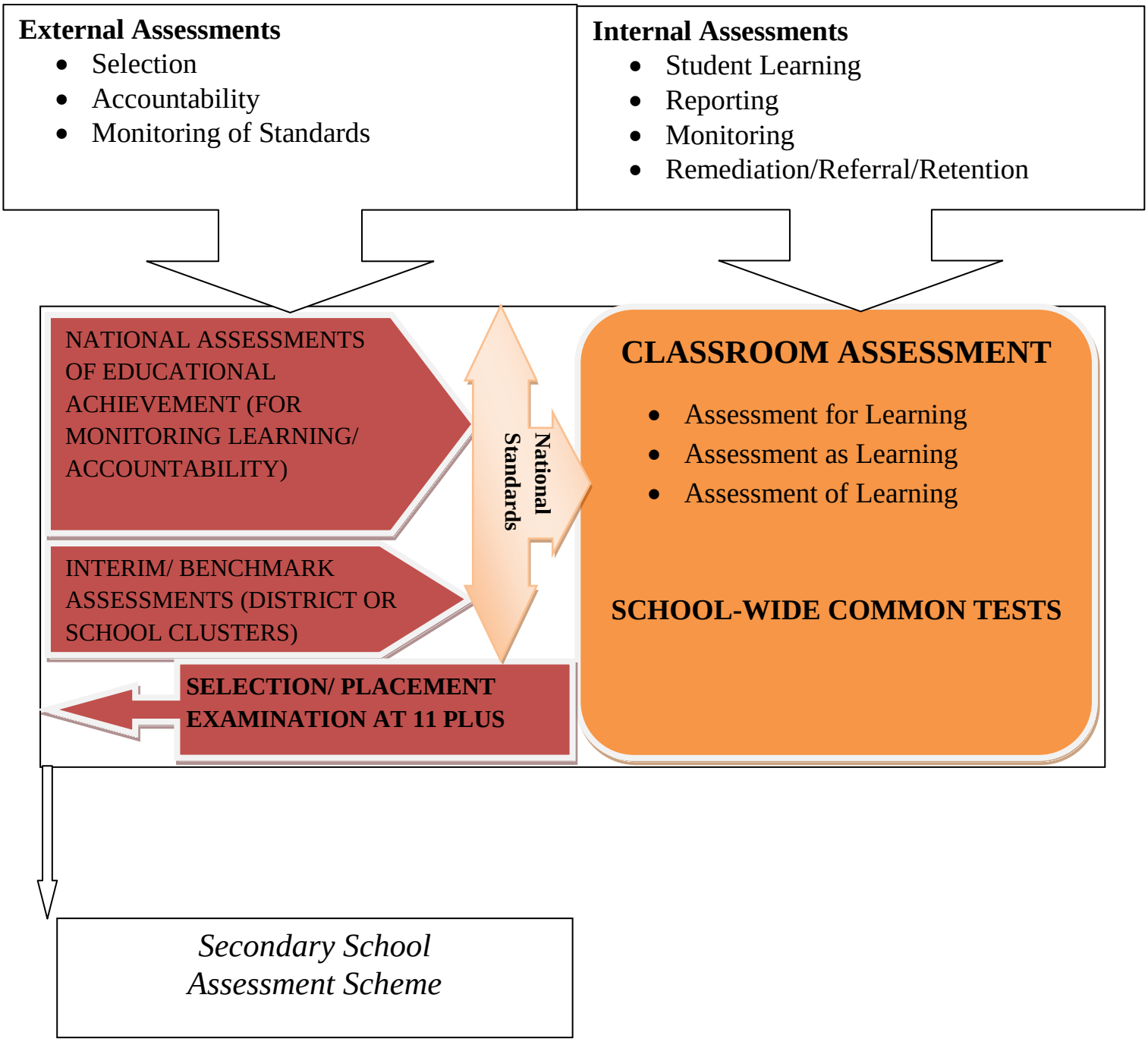


Figure 62: A balanced, comprehensive, nation-wide assessment policy

The Cycle of Formative &Summative Assessment in the Trinidad & Tobago Primary School (Proposed)

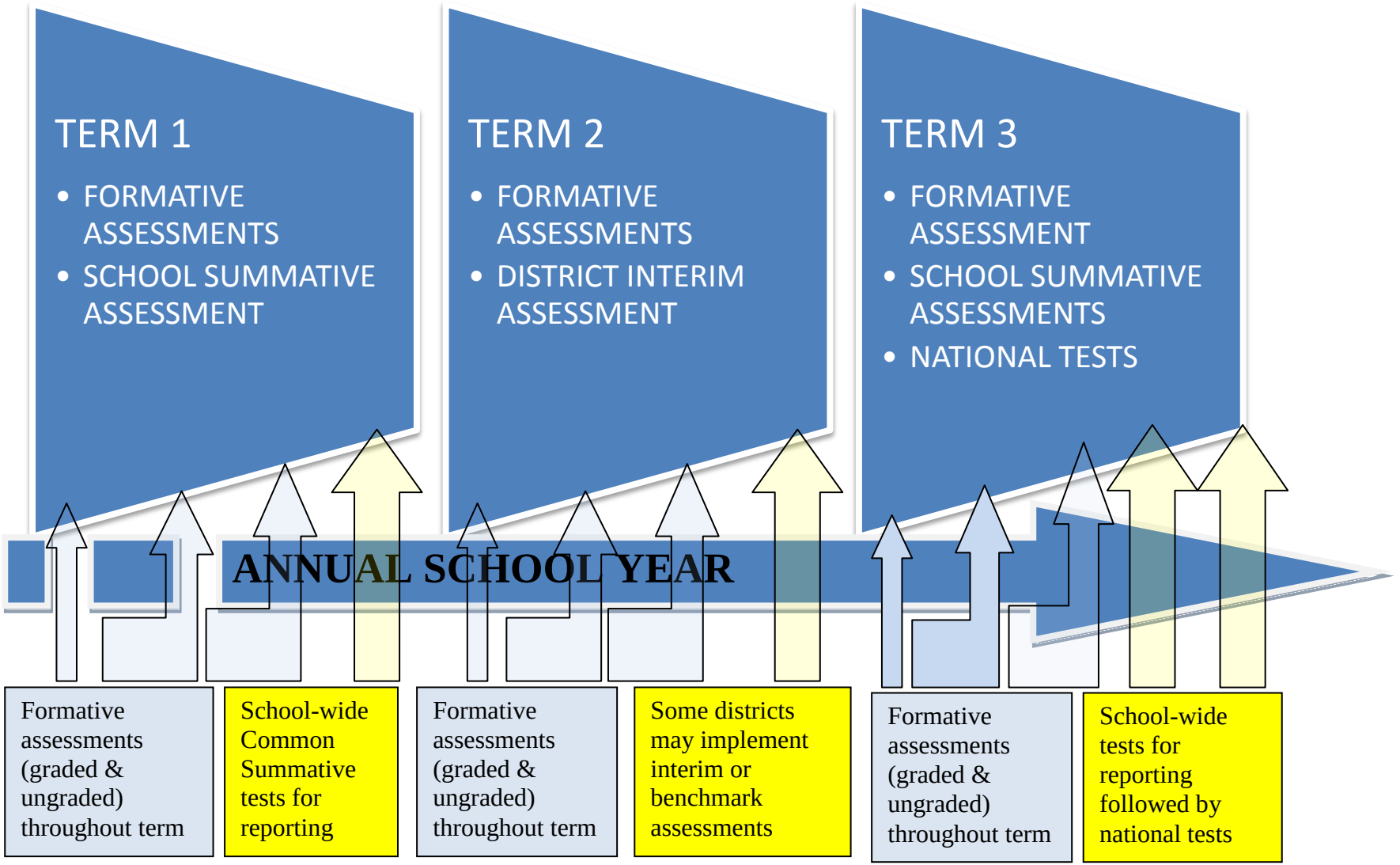


Figure 63: Assessment cycle in the primary school

Multiple Assessments, Unique purposes

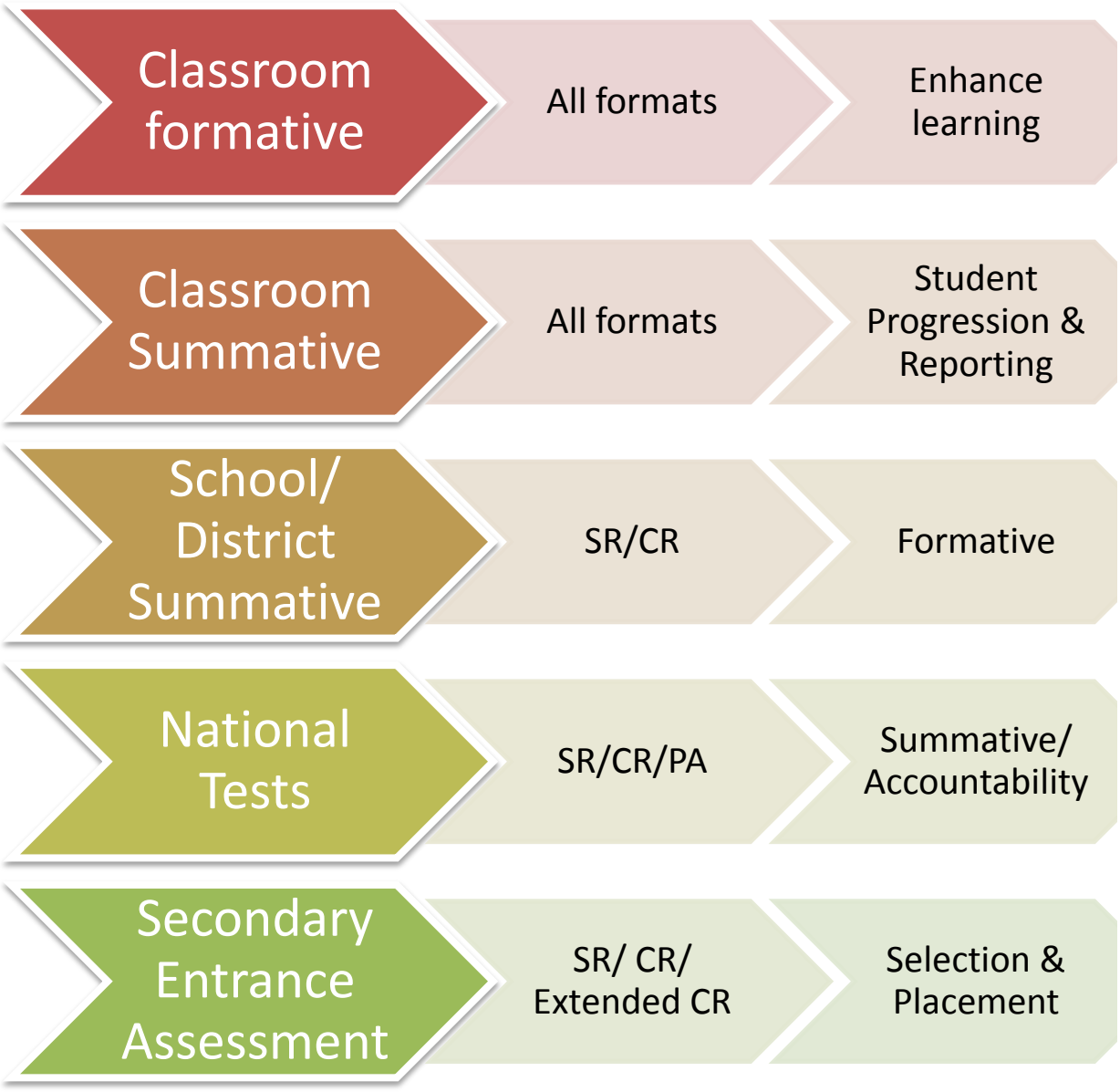


Figure 64: Purposes and assessments

New Assessment Designs

Assessment	Formats	When	Purpose & Use
1) Classroom Formative	CR/ SR/ PA	• Continuous	• To promote learning and metacognition
2) Classroom Summative	• CR/SR/ Occasional PA	• Monthly Assessments	• Progression/ Remediation/ • Reporting
3) School/District (Common Summative Assessments)	• CR/SA	• Annually or Biannual	• Formative • Reporting
4) National Assessments (Revised)	• CR/ SR in Language and Mathematics • CR/SR in Science and Social Studies • Authentic PA in Writing	• Annually Standards 1 & 3 • Biennial Standard 2 • Annually Standard 4	• Accountability • Accountability • Accountability & Formative
5) Secondary Entrance Assessment (Revised)	• CR/SR in Mathematics and Language • Extended CR problem solving paper	• Annually Standard 5	• Selection & Placement • Selection & Placement (Enhanced Differentiation & Validity)

Table 34: Assessments and formats

Timeline for Implementing Change Strategy

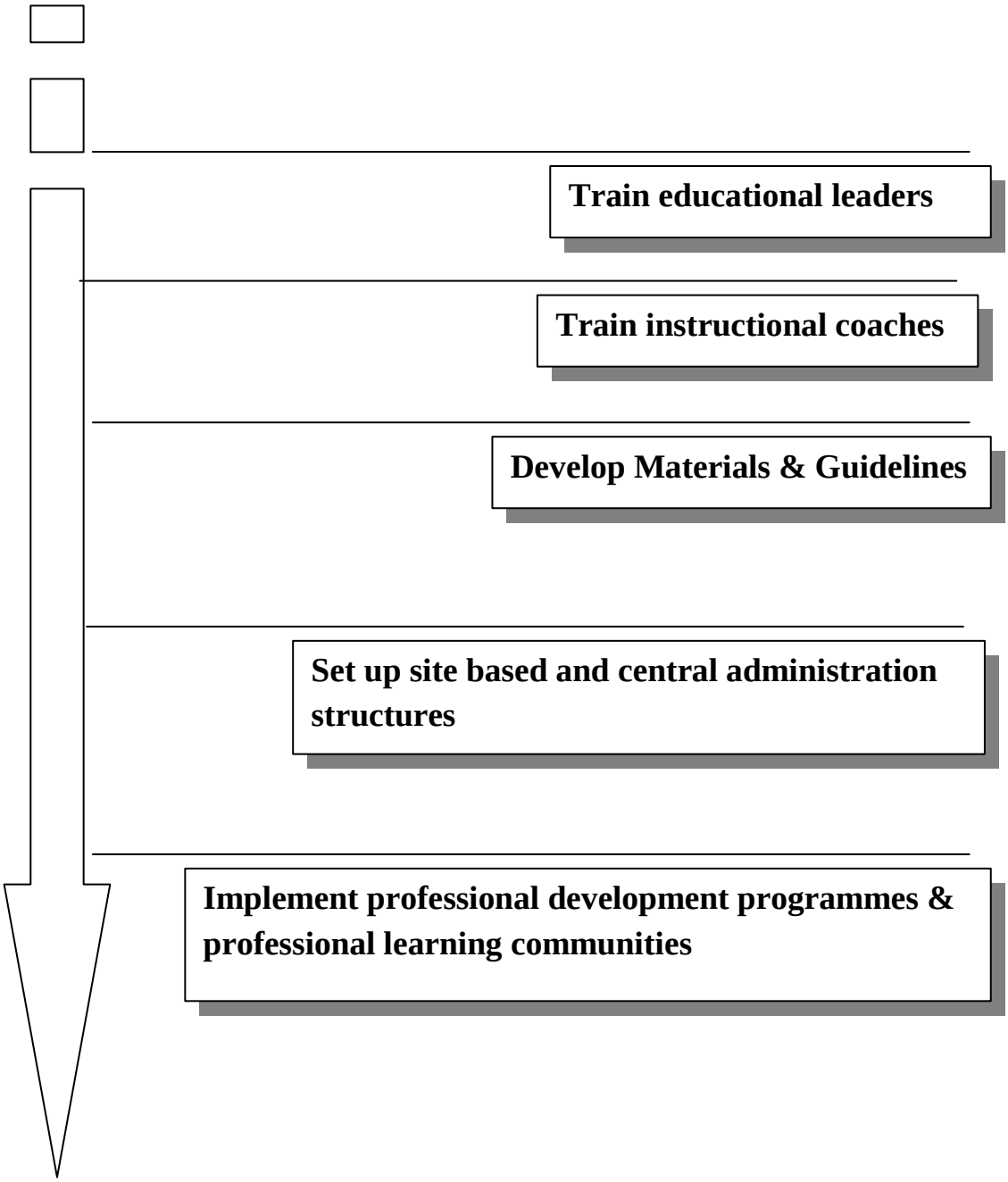


Figure 65: Timeline for implementing change strategy

MINISTRY OF EDUCATION
SEAMLESS EDUCATION SYSTEM PROJECT
Selected Sample Schools for Treatment under the Seamless Education System (SES) Project

Name and Address of School		Name of Principal	Staff Size	Enrolment								ENROL	FREE LUNCH				Agg. API	No.	CLASS	PILOT STATUS
				Infants		Standards					Total			2005	2006	2007				
				1	2	1	2	3	4	5										
37	Scarborough R C	Ms. Jacqueline Springer-Dillon	30	71	75	65	87	84	73	78	533	569	91	305	281	328	914	9	HIGH	N
41	Anstey Memorial Girls' A C	Mrs. Lenore Brewster-Mitchell	20	61	75	38	46	56	47	57	380	354	48	324	353	370	1047	10	HIGH	N
47	Holy Saviour Curepe Anglican	Mrs. Hazel Edwards	25	75	77	79	77	76	77	50	511	531	47	308	342	392	1042	12	HIGH	N
48	St Catherine Girls' Anglican	Ms. Jenny Archer	15	46	55	52	46	52	55	43	349	348	31	354	329	309	992	9	HIGH	N
10	Rampanalgas R C	Mrs. Naomi Marcano-Woodroffe	4	7	10	10	5	10	9	6	57	76	71	121	132	143	396	43	LOW	N
17	Carenage Boys' Government	Mrs. Lorna Simmons-Blackman	14	28	34	35	31	23	44	42	237	170	63	146	162	177	485	62	LOW	N
29	St Mary's A C	Mrs. Maureen Greene	16	45	49	44	44	43	44	43	312	298	61	233	229	283	745	61	LOW	N
52	Patna/River Estate Government	Mrs. Lynette Pierre-Mander	22	42	47	42	70	43	76	78	398	496	41	173	187	241	601	49	LOW	N
15	St Ann's R C	Mr. John Pantin	8	8	9	12	16	12	16	10	83	101	88	196	242	204	642	42	MEDIUM	N
27	Arima Presbyterian	Mr. Carlyle Azaad Gyan	19	48	76	60	48	60	44	70	406	403	46	253	290	294	837	47	MEDIUM	N
28	St Joseph Government	Ms. Patricia Hernandez	20	58	42	42	44	48	41	35	310	296	73	237	250	267	754	58	MEDIUM	N
34	Penal R C (St Dominic's)	Ms. Evangeline Vincent-Davis	19	35	31	38	24	27	38	30	223	261	77	241	285	297	823	33	MEDIUM	N

Table 35: Original Sample for Phase 1 Qualitative Exploratory Study

Bibliography

Adebowale, O.F. & Alao, K.A. (2008). Continuous assessment policy implementation in selected local government areas of Ondo state (Nigeria): Implications for a successful implementation of the UBE program. *KEDI Journal of Educational Policy, 5* (1), pp. 3-18. Retrieved from: http://eng.kedi.re.kr/07_journal/arc_main.php

Ainsworth, L. & Viegut, D. (2006). *Common formative assessments*. Thousand Oaks, CA: Corwin Press.

Airasian, P. W. (1988). Measurement-driven instruction: A closer look. *Educational Measurement: Issues & Practice, 7*(4), 6–11.

Ajidagba, U.A. (2004): Evaluation of students' learning outcomes IV: The use of continuous Assessment. In I. O. Abimbola & A.O. Ablade (Eds.), *Fundamental principles and practice of instruction*. Department of curriculum studies and educational technology, University of Ilorin.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *The standards for competence in the educational assessment of students.* Retrieved from www.unl.edu/buros/article3.html

Andrade, H., & Cizek, G. J. (Eds.). (2010). *Handbook of formative assessment*. New York: Taylor and Francis.

Assessment Reform Group (ARG). (2002). *Assessment is for learning: 10 principles*. Retrieved from http://www.assessment-reform-group.org

Baker, E. L. (2005). Technology and effective assessment systems. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement* (NSSE Yearbook, Vol. 104, Part 2, pp. 358-378).

Baldeosingh, K. (2010). Effective education in Trinidad and Tobago. *Trinidad Express*, Sunday March 28th 2010. Retrieved from http://www.trinidadexpress.com/index.pl/print?id=161616408

Baldeosingh, K. (2010b). Experimental sex education. *Trinidad Express*, Friday April 9th, 2010. Retrieved from http://www.trinidadexpress.com/index.pl/article_opinion?id=161624969

Bell, B. & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85*, 536-553.

Bernhardt, V. (2004). *Data analysis for continuous school improvement.* Larchmont, NY: Eye on Education.

Black, P., & Wiliam, D. (1998a). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan 80* (2), 139-144, 146-148.

Black, P. J., & Wiliam, D. (1998b). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7-73.

Black, P. & Wiliam, D. (2004). Classroom assessment is not (necessarily) formative assessment (and vice-versa). *Towards coherence between classroom assessment and accountability* - 103rd Yearbook of the National Society for the Study of Education (ed. M. Wilson), pp.183-188. Chicago: University of Chicago Press for the NSSE

Black, P., & Wiliam, D. (2007). Large-scale assessment systems: Design principles drawn from international comparisons. *Measurement: Interdisciplinary Research and Perspectives, 5*, 1-53.

Black, P.J., & Wiliam, D. (2009). Developing the theory of formative assessment.

Educational Assessment, Evaluation and Accountability 2, 5–31.
Black, P.J., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy, & Practice 17*(2), 215-232.
Bledsoe, K. L., & Graham, J. A. (2005). The use of multiple evaluation approaches in program evaluation. *American Journal of Evaluation, 26*(3), 302–319.
Bracey, G. W. (1987a). Measurement-driven instruction: Catchy phrase, dangerous practice. *Phi Delta Kappan, 68*(9), 683-686.
Broadfoot, P. (2002) Beware the consequences of assessment!, *Assessment in Education: Policy, Principles and Practice, 9*(3), 285-288.
Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Policy, Principles and Practice, 11*(3), 305-322.
Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports, 99*, 166-170.
Brown, G. T. L. & Lake, R. (2006, November). *Queensland teachers' conceptions of teaching, learning, curriculum and assessment: Comparisons with New Zealand teachers.* Paper presented at the Annual Conference of the Australian Association for Research in Education (AARE), Adelaide, Australia.
Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy and Practice. 16*(3), 347- 363
Campbell, C. and Levin, B. (2009). Using data to support educational improvement. *Educational Assessment, Evaluation and Accountability, 21*, 47-65.
Carless, D, (2005). Prospects for the implementation of assessment for learning. *Assessment in Education: Principles, Policy and Practice, 12*(1), 39-54.
Carless, D. (2007). Conceptualising pre-emptive formative assessment, *Assessment in Education: Principles, Policy and Practice, 14*(2), 171-184.
Chan, J. K. S., Kennedy, K., Yu, W. M., & Fok, P. K. (2006). *Assessment policy in Hong Kong: Implementation issues for new forms of assessment.* Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore, 21-26 May.
Chan, J. K. S. (2007). *We have various forms of assessments but only summative assessments count: Case studies of the implementation of an innovative assessment policy in Hong Kong*. Paper presented at the Redesigning pedagogy - Culture, understanding and practice conference Singapore.
Chapman, D. W. & Snyder, C. W. (2000). Can high stakes national testing improve instruction: reexamining conventional wisdom. *International Journal of Educational Development, 20*, 457-474.
Chappuis, S., Stiggins, R.J., Arter, J., & Chappuis, J. (2005). Assessment for learning: An action guide for school leaders, 2nd ed. Portland, OR: ETS Assessment Training
Chappuis, J. (2009). Seven strategies of assessment for learning. Portland, OR: ETS Assessment Training.
Chappuis, S, & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership, 60*(1), 40-44.
Chen, H.-T. (2005). *Practical program evaluation: Assessing and improve planning, implementation, and effectiveness*. Thousand Oaks, CA: Sage.

Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in language testing: Research contexts and methods.* Mahwah, NJ: Lawrence Erlbaum Associates.

Chewang, K. (1999). *Continuous assessment in Bhutan: Science teachers' perspectives*. Unpublished M.Ed dissertation, University of New Brunswick.

Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. Andrade & G. Cizek, (eds.) *Handbook of formative assessment* (pp. 3-17). New York: Taylor and Francis.

Clarke, S. (2001). *Unlocking formative assessment.* London: Hodder and Stoughton.

Clarke, S. (2005) *Formative assessment in the secondary classroom*. London: Hodder. Murray

Clarke, S. (2008). *Active learning through formative assessment*. London: Hodder Murray.

Clarke, S., McCallum, B., & Lopez-Charles, G. (2001). *Gillingham Partnership formative assessment project*: Interim report on the first term of the project - communicating learning intentions, developing success criteria and pupil self-evaluation. London, UK: Institute of Education, University of London

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Creswell, J. W. & Plano Clark, V. L. (2010). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.

Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*, 438–481.

Dappen, L., & Isernhagen, J. C. (2006). Nebraska STARS: Assessment for learning. *Planning & Changing*, *36*,147-156.

Darling-Hammond, L., & Pecheone, R. (2010). *Developing an internationally comparable balanced assessment system that supports high-quality learning*. Presented at the National Conference on Next Generation K–12 Assessment Systems, Center for K–12 Assessment & Performance Management with the Education Commission of the States (ECS) and the Council of Great City Schools (CGCS), Washington, DC.

Datnow, A., & Castellano, M. (2000). Teachers' responses to Success for All: How beliefs, experiences and adaptations shape implementation. *American Educational Research Journal*, *37*, 775-799.

Davies, A., et al (2009). *Position paper on assessment for learning from the 3rd conference on assessment for learning*, retrieved from http://bulldogcia.com/Documents/Articles/AfL_Position_Paper_Anne_Davies.pdf

De Four, R., De Four, R., & Eaker, B. (2009). New insights into professional learning communities at work. In Michael Fullan (Ed). *The challenge of change: Start school improvement now*. Thousand Oaks, CA: Corwin Press.

Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. Mahwah, NJ: Erlbaum.

Donaldson, S. I., & Gooler, L. E. (2003). Theory-driven evaluation in action: Lessons from a $20 million statewide work and health initiative. *Evaluation & Program Planning*, *26*, 355-366.

Earl, L. (2003). *Assessment as Learning: Using classroom assessment to maximize student learning*. Experts in Assessment series. Corwin Press, inc., Thousand Oaks, California.

Ebelechukwu, E. P. (2007). *School based assessment in Nigerian primary schools.* Paper Presented at the 33rd International Association for Educational Achievement, Baku 16th to 21st September, 2007. Retrieved from http://www.iaea.info/papers.aspx?id=70

Eisemon, T.O. (1990). Examinations policies to strengthen primary school in African

countries. *International Journal of Educational Development,10* (1), 69-82.
Examinations Council of Zambia (2003). Learner assessment for improved educational quality: An exchange of current ideas and best practices: Sub-regional conference on assessment held on 30 June to 2 July, 2003 at Zambezi Sun Hotel Livingstone, ZambiA
Fisher, D., & Frey, N. (2007). *Checking for understanding: Formative assessment techniques for your classroom.* Alexandria, VA: ASCD.
Fullan, M. (2007). *The NEW Meaning of Educational Change (4th ed.).* New York: Teachers' College Press.
Gascon, G. M. (2006). *An application of theory-driven evaluation in educational measurement.* Unpublished Ph.D Thesis, Ohio State University. Retrieved from http://etd.ohiolink.edu/view.cgi?acc_num=osu1165331507
Geddes, S. (2005). Helping teachers develop formative assessment strategies. New Zealand Journal of Teachers' Work, Volume 2, Issue 1, 49-54.
George A.A., Hall G.E., & Stiegelbauer S.M. (2006*). Measuring implementation in schools: The stages of concern questionnaire.* Austin, TX: SEDL.
Griffith, S. (2009). The truth about school-based assessment. *Sunday Gleaner, April 26th, 2009*. Retrieved from *www.jamaica-gleaner.com/gleaner/20090426/.../focus4.html* -
GTC (2004). *Perspectives on pupil assessment.* Paper presented to the GTC conference New Relationships: Teaching, Learning and Accountability London, 29 November 2004. Retrieved from www.gtce.org.uk/documents/publicationpdfs/policy_pnrcp1104.pdf
Guskey, T. R. (Ed.) (2009). *The principal as assessment leader.* Bloomington, IN: Solution Tree.
Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice, 23*(1), 17–27.
Hansen, H.F. (2005). Choosing evaluation models: A discussion on evaluation design. *Evaluation, 11*(4), 447-462.
Harlen, W. (2004). Can assessment by teachers be a dependable option for summative purposes? *GTC (2004) Perspectives on pupil assessment*. Paper presented to the GTC conference New Relationships: Teaching, Learning and Accountability London, 29 November 2004. Retrieved from www.gtce.org.uk/documents/publicationpdfs/policy_pnrcp1104.pdf
Harlen, W. (2005) Teachers' summative practices and assessment for learning –tensions and synergies, *Curriculum Journal*, *16*(2), 207 - 23.
Harris, A. & Pasinga, A. (1993). *Curriculum-based Assessment: Linking curriculum, assessment and learning in developing education systems.* IEQP. USAID document no. PN-ABZ-283.
Hayford, S. K. (2007). Continuous assessment and lower attaining pupils in primary and junior secondary schools in Ghana. Unpublished Ph.D Thesis, School of Education, University of Birmingham, UK.
Heritage, M. (October 2, 2007). *Formative assessment: What do teachers need to know and do? Phi Delta Kappan, 89* (2), 140-145
Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin Press.
Heyneman, S.P., (1987). Uses of examinations in developing countries: selection, research, and education sector management. *International Journal of Educational Development 7* (4), 251–263.
Hall, G. & Hord, S. (2006). *Implementing change: Patterns, principles, and potholes (2nd*

ed.). Boston, MA: Allyn & Bacon.
Hall, G. & Hord, S. (2006). *Implementing change: Patterns, principles, and potholes (3rd ed.)*. Boston, MA: Allyn & Bacon.
Israel, H. F. (2000). *The implementation and effects of continuous assessment in the English classrooms within the changing milieu of Education in South Africa.* Unpublished Ed.D dissertation. University of Baylor, Waco, Texas.
Johnson, D. (2006) Comparing the trajectories of educational change and policy transfer in developing countries. *Oxford Review of Education, 32*(5), 679-696.
Kalomba, L. S. & Mpaju, M. A. (2003). The impact of assessment and certification on teaching and learning: The Tanzania experience. Paper presented at the 21st Conference of the Association of Educational Assessment in Africa (AEAA), Cape Town, South Africa, August 25th to 30th, 2003. Retrieved from http://www.umalusi.org.za/ur/Conferences/AEAA/Kalomba%20L%20and%20Mpaju%20M%20A.PDF
Kamangira, Y. T. (2003). *Feasibility of a large-scale implementation of continuous assessment as a stimulus for teacher development in Malawi*. American Institutes for Research: Improving Educational Quality (IEQ) Project.
Kapambwe, W. M. (2010). The implementation of school based continuous assessment (CA) in Zambia. Educational Research and Reviews, 5(3), 99-107. Available online at http://www.academicjournals.org/ERR
Kassam, Y. (1994). Julius Kabarage Nyerere. *Prospects: The Quarterly Review of Comparative Education, XXX1V,*1/2, 247-259.
Kellaghan, T., & Greaney, V. (2001). *Using assessment to improve the quality of education*. Paris, UNESCO: International Institute for Educational Planning.
Kennedy, K. (2007). *Barriers to innovative school practice: A sociocultural framework for understanding assessment practices in Asia*. Paper prepared for the Symposium: "Student assessment and its social and cultural contexts: How teachers respond to assessment reforms". Redesigning pedagogy – culture, understanding and practice conference, Singapore, 28-30 May.
Knight, J. (2009). Coaches as leaders of change. In M. Fullan (Ed). *The challenge of change: Start school improvement now*. Thousand Oaks, CA: Corwin Press.
Kowalski, T. J. and Lasley, T J., (2008). Handbook of data-based decision making in education. New York, NY: Routledge
Kruger, G. (2004, September). *Continuous assessment in the lower primary school phase: From confusion to direction.* Paper presented at the meeting of the Reform Forum. Retrieved from www.nied.edu.na/publications/journals/journal18/Article3.pdf
Le Grange L. & Chris, R. (1998). *Continuous assessment: An introduction and guidelines to implementation. Juta: Cape Town.*
London, N. A. (1997). Socio-politics in effective curriculum change in a less developed country: Trinidad and Tobago. *Curriculum Inquiry, 27(1),* 63-80.
London, N. A. (1989). Selecting students for secondary education in a developing society: The case of Trinidad and Tobago. *McGill Journal of Education, 24*(3), 281-291.
Lukin, L. E. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues & Practice, 23*(2), 26-32.
Matthews, K. W. (2007). *The relationships among middle level teachers' assessment*

practices, instructional leadership, and student achievement. Unpublished Ph.D. Thesis, University of Missouri, Retrieved from the worldwide web @ http://gradworks.umi.com/33/22/3322724.html

Maxwell, G. (2002). *Moderation of teacher judgements in student assessment.* Discussion paper. Brisbane: Queensland School Curriculum Council.

Maxwell, J. A. (2004a). Causal explanation, qualitative research, and scientific inquiry in education. *Educational Researcher, 33*(2), 3-11.

Maxwell, J. A. (2004b). Using qualitative methods for causal explanation. *Field Methods, 16*(3), 243-264.

Mchazime, H. (2003). *Integrating primary school curriculum and continuous assessment in Malawi.* American Institutes for Research: Improving Educational Quality (IEQ) Project.

McManus, S. (2008). *Attributes of effective formative assessment*. Washington, D.C.: Council of Chief State School Officers.

Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in Kindergarten - Grade 3. *American Educational Research Journal, 38*(1), 73–95.

Mkhonta, L. A. (2003). *Continuous assessment at primary school in Swaziland*. Paper delivered at the sub-regional conference on assessment. Learner assessment for improved educational quality: An exchange of current ideas and best practices. 30 June to 2 July, 2003 Zambezi Sun Hotel Livingstone, Zambia. Conference Proceedings, pp. 33-40.

Moroz, R., & Waugh, R. F. (2000). Teacher receptivity to system-wide educational change. *Journal of Educational Administration, 38*(2), 159–178.

Moss, P. A., Pullin, D. C., Gee, J. P., Haertel, E. H., & Young, L. J. (Eds.). (2008). *Assessment, equity, and opportunity to learn*. Cambridge: Cambridge UP.

Motshekga, A. (2010). *Statement by the Minister of Basic Education, Mrs Angie Motshekga, MP on the progress of the review of the National Curriculum Statement*, Tuesday 06 July 2010. Retrieved from http://go2.wordpress.com/?id=725X1342&site=edulibpretoria.wordpress.com&url=http%3A%2F%2Fedulibpretoria.files.wordpress.com%2F2010%2F07%2Fmotshekga-statement-6-july-2010.pdf&sref=http%3A%2F%2Fedulibpretoria.wordpress.com%2F2010%2F07%2F07%2Fnew-curriculum-for-south-african-schools%2F

Mulenga, M. B. and Kapambwe, W. M. (2008). *The implementation of school based continuous assessment (CA) in Zambia. Paper presented at the 2008 IAEA Conference.*

NECTA (2003). *History of the National Examinations Council of Tanzania*. Retrieved from http://www.matokeo.necta.go.tz/history.htm

New Zealand Ministry of Education (2008). *National education findings of Assess to Learn (AtoL) Report*. Retrieved from the worldwide web @ http://www.educationcounts.govt.nz/__data/assets/pdf_file/0004/28165/Eval_Nat_Edu_Findings_Formatted2.pdf

Nitko, A., (1995). Curriculum-based continuous assessment: a framework for concepts, procedures and policy. *Assessment in Education, 2*(3), 321–338.

Noell, G. H., & Gansle, K. A., (2008). Moving from good ideas in educational systems change to sustainable program implementation: Coming to terms with some of the realities. *Psychology in the Schools, 46*, 78-88.

Nsibande, R. (2007) Knowledge and practice of continuous assessment: The barriers for

policy transfer. Unpublished doctoral dissertation. University of South Africa. Retrieved from http://wiredspace.wits.ac.za/handle/10539/2187

Nyerere, J.K. (1967). *Education for self-reliance*. Reprinted in Nyerere (1968) Freedom and Socialism. Dar es Salaam, Oxford University Press.

OECD. (2005). *Formative assessment: improving learning in secondary classrooms*. Paris: Author.

Ontario Ministry of Education (2010). *Growing success: Assessment, evaluation, and reporting in Ontario schools.* First Education Covering Grades 1 to 12. Retrieved from *ww.edu.gov.on.ca/eng/policyfunding/grow**Success**.pdf.*

Owston, R. (2007). Contextual factors that sustain innovative pedagogical practice using technology: an international study. *Journal of Educational Change, 8,* 61-77.

Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan, 68,* 679-682.

Popham, W. J. (2008a). *Classroom assessment: What teachers need to know* (5th ed.). Boston: Allyn & Bacon.

Popham, W. J. (2008b). *Transformative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Popham, W. J. (2009). *Is assessment literacy the "magic bullet"?* The Blog of Harvard Education Publishing. Retrieved from http://www.hepg.org/blog/19

Preuss, P. (2007). Data-driven decision-making and dynamic planning: A school leader's guide. Larchmont, NY: Eye on Education.

Pryor, J. & Crossouard, B. (2005). *A sociocultural theorization of formative assessment.* Paper presented at Sociocultural Theory in Educational Research and Practice Conference, University of Manchester, September 8-9. Retrieved from http://orgs.man.ac.uk/projects/include/experiment/pryor_crossouard.pdf

Pryor, J. & Crossouard, B. (2008) A sociocultural theorization of formative assessment. *Oxford Review of Education, 33*(5) 1-20.

Pryor, J. & Lubisi, C. (2002) Reconceptualising educational assessment in South Africa - Testing times for teachers, *International Journal of Educational Development 22*(6). 673-686.

Redfield, D., Roeber, E., & Stiggins, R. (2008*). Building balanced assessment systems to guide educational improvement.* Paper presented at the Council of Chief State School Officers 38th National Conference on Student Assessment, Orlando, FL. Retrieved June24, 2008, from http://www.ccsso.org/content/PDFs/OpeningSessionPaper-Final.pdf

Rogers, E.M. (1995). Diffusion of innovations (4th edition). The Free Press. New York.

Rogers, P. J., Hacsi, T. A., Petrosino, A., & Huebner, T. A. (Eds.) (2000). *Program theory in evaluation.* (New Directions for Program Evaluation, no. 87). San Francisco: Jossey-Bass.

Rogers, P., Petrosino, A., Hacsi, T., & Huebner, T. (2000). Program theory evaluation: practice, promise and problems. In: P. Rogers, A. Petrosino, T. Hacsi and T. Huebner (Eds.), *Program theory evaluation: Challenges and opportunities* (pp. 5–13), New Directions in Evaluation series. San Francisco, CA: Jossey-Bass.

Rogers, P. J. (2000). Causal models in program theory evaluation. In P. J. Rogers, A. J. Petrosino, T. Hacsi and T. A. Huebner (Eds.). *Program theory evaluation: Challenges and opportunities (pp. 47–55)*, New Directions in Evaluation Series. San Francisco, CA: Jossey-Bass.

Rogers, P. J. (2004). Logic models. In S. Mathison (Ed.) *Encyclopaedia of evaluation*, pp. 232–4. Newbury Park, CA: SAGE.

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation, 14*(1), 293-48.

Sadler, R. (1998). Formative assessment: revisiting the territory. *Assessment in Education, 5*(1), 77-84

Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice, 32*(2), 118-126.

Schein, E. H. (1996). *Three cultures of management: The key to organizational learning in the 21st Century*. Paper presented to the MIT. Sloan School of Management. Retrieved from http://www.solonline.org/res/wp/three.html

Schildkamp, K. & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching & Teacher Education, 26* (3), 482-496

Sebatane, M. (2000) International transfers of assessment: Recent trends and strategies. *Assessment in Educational Principles, Policy and Practice, 7*(3), 401-416.

Senge, P.M. (1996), Leading learning organizations: The bold, the powerful, and the invisible. In M. Goldsmith and F. Hasselbein (Eds.), *The leader of the future* (pp 41 -58). San Francisco, CA: Jossey-Bass Inc.

Senge, P. (1990) The fifth discipline: The art and practice of the learning organization. N.Y: Doubleday

Shandomo, H. (2008). *Continuous assessment in Swaziland: The predictable fate of western innovation in Africa. Saarbrücken: VDM Verlag Dr. Müller Aktiengesellschaft & Co. KG*

Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Shepard, L.A. (2005). Linking formative assessment to scaffolding. *Educational Leadership, 63*(3), 66-70.

Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), Educational measurement (4th ed.).Westport, CT: Praeger.
Shute (2008), pg 28, para 1, line 1

Singapore Examinations and Assessment Board (2006). *Can statistical and qualitative modes of moderation co-exist in a model for the quality assurance of school-based assessment – A South African perspective?* Paper presented at the 32nd IAEA Conference, 21-26 May, Singapore, Singapore. Retrieved from *www.iaea2006.seab.gov.sg/.../Can%20**statistical**%20and%20**qualitative**%20modes%20of%20**moderation**...*

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1),153-189.

Smith, M.L., Edelsky, C., Draper, K., Rottenberg, C., & Cherland, M. (1989). *The role of testing in elementary schools.* Los Angeles, CA: Center for Research on Educational Standards and Student Tests, Graduate School of Education, UCLA..

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. *Educational Researcher, 20(5),* 8-11.

Smith, M.L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues & Practice, 10,* 7-11.

Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan, 72,* 534-539.

Stiggins, R. J. (1994). *Student-centered classroom assessment*. New York: Macmillan Publishing Company.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan, 77*(3), 238-245.

Stiggins, R. J. (1999a). Are you assessment literate? *The High School Journal, 6*(5), 20-23.

Stiggins, R. J. (1999b). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues & Practice, 18*(1), 23-27.
Stiggins, R. J. (2000). Learning teams for assessment literacy. *OISE/UT's Magazine for Schools, 30*(4), 5-7.
Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice, 20*(3), 5-15.
Stiggins, R. J. (2008). *Assessment manifesto: A call for the development of balanced assessment systems*. Portland, OR: Educational Testing Service, Assessment Training Institute.
Stufflebeam, D. (ed.). (2001). *Evaluation models. New directions for program evaluation, 89*. San Francisco: Jossey-Bass.
Susuwele-Banda, W. J. (2005). Classroom assessment in Malawi: Teachers; perceptions and practices in Malawi. Unpublished Ph.D. Dissertation, Virginia Polytechnic Institue.
Swallow, J. R., Nielson, E. B., Chakufyali, P. N. (2009). *USAID/Zambia education program evaluation*. United States Agency for International Development. Prepared by DevTech Systems.
Teddlie, C. & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage.
Tierney. R. J. (1998). Literacy assessment reform: Shifting beliefs, principles possibilities, and emerging practices. *The Reading Teacher, 51*, 374-390.
Tuttle, H. G. (2009). *Formative assessment: Responding to your students*. Larchmont, NY: Eye on Education, Inc.
UNESCO. (1990). *World Declaration on Education For All*. Retreived on July 24 2010 from http://www.unesco.org/education/efa/ed_for_all/background/jomtien_declaration.shtml
UNESCO (2008). *Education for All by 2015: Will we make it? EFA Global Monitoring Report*. Paris: UNESCO and Oxford University Press
Vandeyar, S., & Killen, R. (2003). Has curriculum reform in South Africa really changed assessment practices, and what promise does the revised national curriculum statement hold? *Perspectives in Education (21)*, 119-143.
Vandeyar, S. & Killen, R. (2007). Educators' conceptions and practice of classroom assessment in post-apartheid South Africa. South African Journal of Education 27(1), 101-115.
Vedung, E. (1997) *Public policy and program evaluation*. New Brunswick, NJ: Transaction Publishers.
Vlaardingerbroek, B. & Taylor, N. (Eds.) (2009). *Secondary school external examination systems – Reliability, robustness and resilience*. NY: Cambria Press
Volante, L. (2010). Assessment of, for, and as learning within schools: Implications for transforming classroom practice. *Action in Teacher Education, 31*(4), 66-75.
Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System, 28*, 499-509.
Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: A case study using insights from testing and innovation theory*. New York: Cambridge University Press.
Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing, 10* (1), 41-69.
Walshe, K. (2007). Understanding what works -and why- in quality improvement: the need

for theory-driven evaluation. *International Journal for Quality in Health Care, 19*(2),57-59.

Wangchuk, K. (2000). *The feasibility of rubrics and portfolio assessment for use in continuous. assessment in Bhutan*. Master of Education Dissertation, University of New Brunswick.

Waugh, R.F. (2000). Towards a model of teacher receptivity to planned system-wide educational change in a centrally controlled system. *Journal of Educational Administration, 38*(4), 350-367.

Waugh, R.F. & Punch, K.F. (1987). Teacher receptivity to system wide change in the implementation stage. *Review of Educational Research, 57*(3), 237-254

Webb, N. (2002, April). *Assessment literacy in a standards-based urban education setting.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Webb, D. C., Romberg, T. A., Dekker, T., de Lange, J., & Abels, M. (2004). Classroom assessment as a basis for teacher change. In T. A. Romberg (Ed.), *Insight stories: Assessing middle school Mathematics* (pp. 223-235). New York: Teachers College Press.

Webb, D.C. (2009) Designing professional development for assessment. *Educational Designer, 1*(2) (26 pages). Retrieved from the worldwide web @ *http://www.educationaldesigner.org/ed/volume1/issue2/article6*

Weiss, C. (1995). Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community based initiatives for children and families. In Connell, J. P., A. C. Kubisch, L. B. Schorr & C. H. Weiss, (Eds.), *New approaches to evaluating community initiatives: Concepts, methods, and contexts (pp 65-92).* Washington, DC: The Aspen Institute.

Wilson, M., & K. Sloane. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education 13*(2), 181–208.

Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy, & Practice, 17*(1), 59-75.

Wylie, E. C., Lyon, C. J., & Goe, L. (2009). *Teacher professional development focused on formative assessment: changing teachers, changing schools* (Vol. RR-09-10). Princeton, NJ: Educational Testing Service...

Wolff, L. (2004). Educational assessments in Latin America: The state of the art. *Applied Psychology: An International Review, 53*(2), 192–214.

World Bank (1995). *Report No. 14865-TR. Staff Appraisal Report, Trinidad and Tobago Basic Education Report,* October 26, 1995. Washington, DC: Author.

Yu, W. M., Kennedy, K. J., Fok, P. K., & Chan, K. S. J. (2006, May). *Assessment reform in basic education in Hong Kong: The emergence of assessment for learning*. Paper presented at the 32nd Annual Conference of International Association for Educational Assessment: Assessment in an Era of Rapid Change: Innovations and Best Practices, Singapore.

Yu,W. M. (2007). Teachers' responses to student assessment reform in a Chinese context. Paper presented at the Redesigning Pedagogy – Culture, Understanding and Practice Conference, May 28–30, in Singapore.

Zambia Ministry of Education (1996) *Educating our Future*. Retrieved from planipolis.iiep.unesco.org/upload/**Zambia**/**Zambiaeducatingourfuture**.pdf

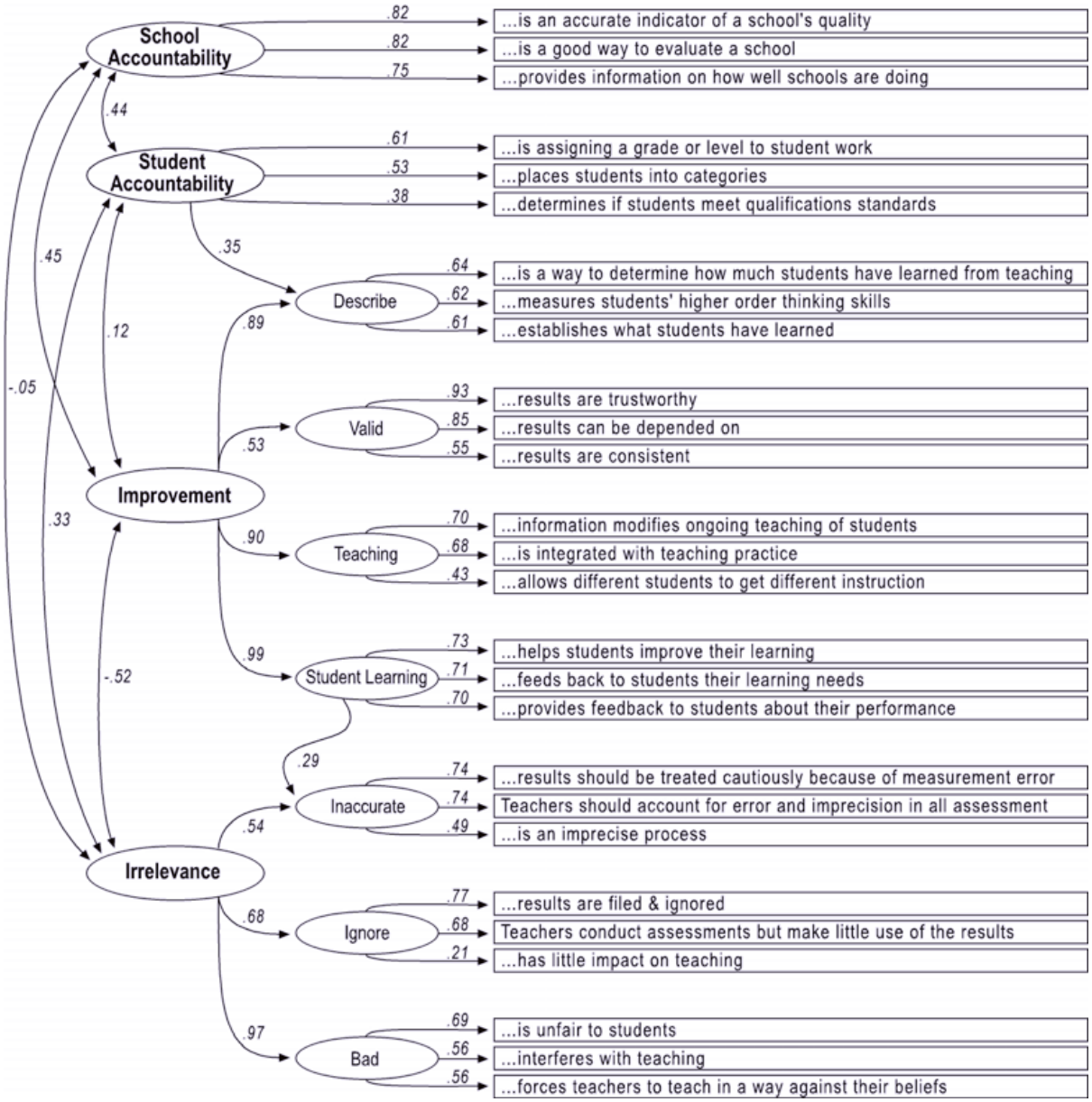
Appendix 1

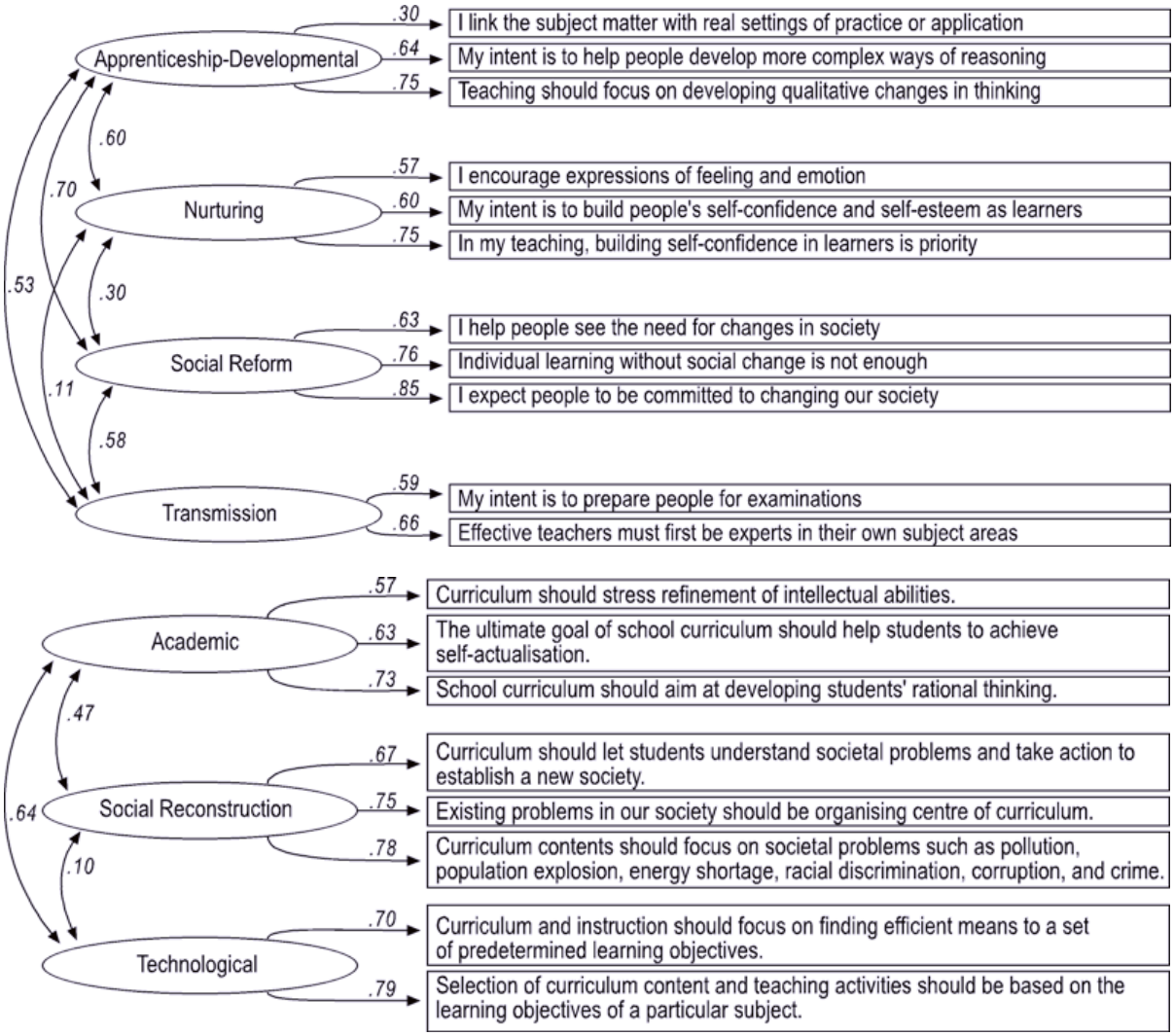
SCHOOLS IN QUANTITATIVE SAMPLE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Anstey Memorial A.C	14	3.7	3.7	3.7
	Aranguez Hindu	9	2.4	2.4	6.1
	Ascension A.C.	8	2.1	2.1	8.2
	Barrackpore Vedic	3	.8	.8	9.0
	Bien Venue Presbyterian	7	1.9	1.9	10.8
	Brighton Anglican	2	.5	.5	11.4
	Carapichaima ASJA	16	4.2	4.2	15.6
	Carenage Boys' Government	10	2.6	2.6	18.3
	Cocoyea Government	6	1.6	1.6	19.8
	Curepe Anglican	17	4.5	4.5	24.3
	D'Abadie Government	14	3.7	3.7	28.0
	Diego Martin Government Primary	5	1.3	1.3	29.4
	El Dorado South Hindu	17	4.5	4.5	33.9
	Felicity SDMS Hindu Primary	15	4.0	4.0	37.8
	Fifth Company	12	3.2	3.2	41.0
	Freeport Presbyterian	17	4.5	4.5	45.5
	Guayaguayare R.C.	16	4.2	4.2	49.7
	Harmony Hall	13	3.4	3.4	53.2
	Inverness Presbyterian	10	2.6	2.6	55.8
	La Lune R.C	6	1.6	1.6	57.4
	La Puerta Government Primary School	11	2.9	2.9	60.3
	Macaulay Government	7	1.9	1.9	62.2
	Mon Repos R.C	7	1.9	1.9	64.0
	Moriah Government	9	2.4	2.4	66.4
	Nelson St. Girls	12	3.2	3.2	69.6
	North Trace Government	9	2.4	2.4	72.0
	Orange Field Hindu School	18	4.8	4.8	76.7
	Palmiste Government	7	1.9	1.9	78.6
	Patna River Estate	4	1.1	1.1	79.6
	Rio Claro Presbyterian	13	3.4	3.4	83.1
	Rio Claro Vedic	6	1.6	1.6	84.7
	San Fernando TML	15	4.0	4.0	88.6
	Sixth Company Anglican	5	1.3	1.3	89.9
	St. Joseph Government Primary	13	3.4	3.4	93.4
	St. Mary's Anglican	15	4.0	4.0	97.4
	Unknown	3	.8	.8	98.1
	Waterloo SDMS Hindu	7	1.9	1.9	100.0
	Total	378	100.0	100.0	

Appendix 2

Conceptions of Assessment, Teaching, & Curriculum




Apprenticeship-Developmental
Nurturing
Social Reform
Transmission
Academic
Social Reconstruction
Technological
I link the subject matter with real settings of practice or application
My intent is to help people develop more complex ways of reasoning
Teaching should focus on developing qualitative changes in thinking
I encourage expressions of feeling and emotion
My intent is to build people's self-confidence and self-esteem as learners
In my teaching, building self-confidence in learners is priority
I help people see the need for changes in society
Individual learning without social change is not enough
I expect people to be committed to changing our society
My intent is to prepare people for examinations
Effective teachers must first be experts in their own subject areas
Curriculum should stress refinement of intellectual abilities.
The ultimate goal of school curriculum should help students to achieve self-actualisation.
School curriculum should aim at developing students' rational thinking.
Curriculum should let students understand societal problems and take action to establish a new society.
Existing problems in our society should be organising centre of curriculum.
Curriculum contents should focus on societal problems such as pollution, population explosion, energy shortage, racial discrimination, corruption, and crime.
Curriculum and instruction should focus on finding efficient means to a set of predetermined learning objectives.
Selection of curriculum content and teaching activities should be based on the learning objectives of a particular subject.

Appendix 3-Completed Field Researchers Form

FORM 1-MOE CAP EVALUATION-

Judgement of Implementation Levels

NAME OF INSITUTION: Patna River Estate Government School
EDUCATION DISTRICT
NAME OF RATER: Gwendolyn Ahyoung
JOB POSITION OF RATER: Principal Primary
INFORMATION

Question	----------SCALES----------			
1. Was the School part of the CAP pilot?	**YES**	**NO**		
2. Is the CAP currently in operation at the school?	**YES**	**NO**	**SOME-WHAT**	
3. Overall, how successful was this school in implementing the CAP	**0-25%**	**26-50%**	**51-75%**	**76-100%**
4. Approximately, in what percentage of classrooms was the CAP practiced?	**0-25%**	**26-50%**	**51-75%**	**76-100%**
5. Approximately, in what percentage of classrooms is the CAP still practiced?	**0-25%**	**26-50%**	**51-75%**	**76-100%**
6. What percentages of Commutative Record Cards are completed regularly?	**0-25%**	**26-50%**	**51-75%**	**76-100%**
7. To what extent does the school still train teachers for the CAP?	**Never**	**Biennially**	**Annually**	**By each term**
8. Currently, are there structures and resources in the schools dedicated to the CAP (Cupboards, manuals, etc)	**None**	**A few**	**Some**	**Many**
9. How much money in the school's budget is still allocated to the CAP process?	**0-5%**	**6-10%**	**10-15%**	**>16%**
10. To what extent is the data from CAP now used by the school in making decisions about students and teaching-learning?	**Never**	**Some-times**	**Often**	**A great deal**
Overall, how will you describe the level of implementation of CAP in this school	**Very Poor**	**So-so**	**Relatively High**	**Superior to most**

Additional Comments on why this school is a good or poor example of CAP implementation

The Principal has high expectations for her school but there is some form of division among the staff, as she tries to develop certain structures within the school. Some teachers still want to maintain the traditional mode of operation.

(Use next side if necessary)

FORM 2-MOE CAP EVALUATION-

Formative Assessment in Classroom

NAME OF INSTITUTION: Patna River Estate Government School

CLASSROOM: Standard One

EDUCATION DISTRICT:

NAME OF RATER: Gwendolyn Ahyoung

Instructions:

Observe the classroom at intervals during the day. You should make 5 observations of 15 minutes each. You may ask the teacher for artefacts and you should make substantial field notes in your field researcher's diary to be submitted. You may also ask the teacher to keep a diary of his classroom activity for the week of your data collection.

Core Question	**Evidence to Look for**	**Judgement (High=5; Low=1)**
7. To what extent does the classroom culture encourage multiple interactions and the use of assessment tools?	Assessments used by the teacher/ group work activity/ projects/ portfolios/ teacher diary/student interviews	**3**
8. To what extent have learning goals for the class been established and is individual progress tracked?	Cumulative Record Cards/Record, Forecast and Evaluation	**3**
9. To what extent is instruction varied to meet the needs of different learners	Record And Forecast	**3**
10. To what extent is assessment varied to meet the needs of different learners	Record And Forecast	**3**
11. To what extent is feedback to learners provided and to what extent is instruction adapted in response to the data?	Teacher diary/interview	**1**
12. To what extent are student active in all phases of teaching and learning?	Record And Forecast/Observation	**3**

FORM 3-MOE CAP EVALUATION-

The "Am I Doing CAP" Checklist

NAME OF INSTITUTION: Patna River Estate Government School

CLASSROOM: Standard One

EDUCATION DISTRICT:_____

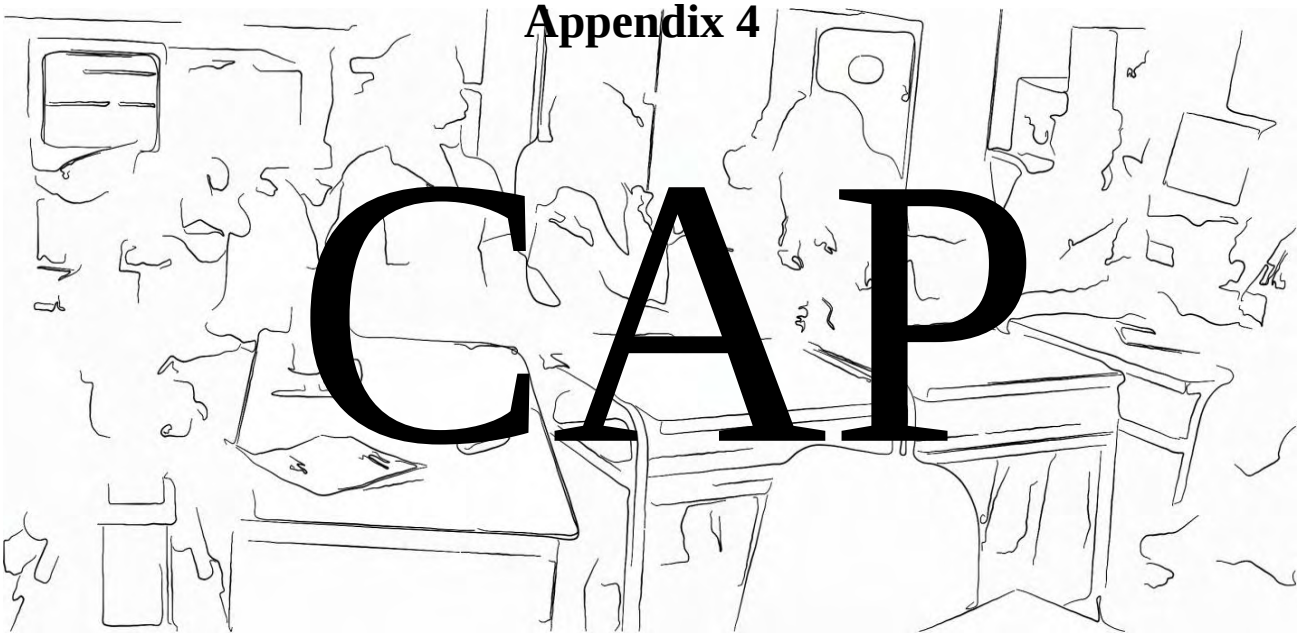
NAME OF RATER: Gwendolyn Ahyoung

Instructions:

Observe the classroom at intervals during the day. You should make about 5 observations of 15 minutes each. You may ask the teacher for artefacts and you should make substantial field notes in your field researcher's diary to be submitted. You may also ask the teacher to keep a diary of his classroom activity for the week of your data collection.

Activity	Never	Rarely	Some-times	Often	Always
11. Ongoing, continuous testing			-		
12. Maintains Cumulative Record Cards		-			
13. Gives varied feedback			-		
14. Sets goals/standards for learners				-	
15. Uses multiple forms of delivery			-		
16. Uses data from students			-		
17. Uses multiple modes of assessment		-			
18. Uses quality performance assessment		-			
19. Keeps Anecdotal Records and Journals			-		
20. Uses assessments diagnostically		-			
21.					

Appendix 4



The Continuous Assessment Programme in Trinidad & Tobago:

An Evaluation

Teachers' Questionnaire

Conducted for and on behalf the Seamless Education Project Unit and DERE, Trinidad and Tobago Ministry of Education

Demographic Information

1. **What is your current age?**
   - ☐ < 20 Years
   - ☐ 20-25 Years
   - ☐ 26-30 Years
   - ☐ 31-34 Years
   - ☐ 35-40 Years
   - ☐ 41-45 Years
   - ☐ 46-50 Years
   - ☐ 51-55 Years
   - ☐ 56-60 Years
   - ☐ > 60 Years

2. **How would you identify yourself? (TICK ONE)**
   - ☐ African-Trinidadian/Tobagonian
   - ☐ Indian Trinidadian
   - ☐ African-Indian Trinidadian (Mixed)
   - ☐ Syrian-Lebanese Descent
   - ☐ Chinese Descent
   - ☐ European Descent
   - ☐ Other_____

3. **What is your current marital status?**
   - ☐ Single
   - ☐ Separated
   - ☐ Divorced
   - ☐ Married
   - ☐ Widowed

4. **How long have been you a teacher?**
   - ☐ Less than 5 year
   - ☐ 5-9 years
   - ☐ 10-14 years
   - ☐ 15-19 years
   - ☐ 19-24 years
   - ☐ 25-29 years
   - ☐ > 30 years

5. **Gender**
   - ☐ Male
   - ☐ Female

6. **How long have been you a teacher at your current school?**
   - ☐ Less than 5 year
   - ☐ 5-9 years
   - ☐ 10-14 years
   - ☐ 15-19 years
   - ☐ 19-24 years
   - ☐ 25-29 years
   - ☐ > 30 years

7. **Which of the following academic qualifications have you attained?**
   - ☐ O-Levels
   - ☐ A-Levels
   - ☐ Tertiary Level Certificate
   - ☐ Undergraduate Diploma
   - ☐ Associate Degree
   - ☐ Undergraduate Degree
   - ☐ Postgraduate Diploma
   - ☐ Masters Degree
   - ☐ Doctoral Degree

8. **Which of the following professional qualifications have you attained?**
   - ☐ ACP/LCP or similar
   - ☐ Training College Certificate
   - ☐ Bachelor of Education Degree
   - ☐ Postgraduate Diploma of Education
   - ☐ Masters in Education Degree
   - ☐ Ed.D.
   - ☐ Ph.D. in Education
   - ☐ Other_____

9. What class do you currently teach?

- ☐ Infants 1
- ☐ Infants 2
- ☐ Standard 1
- ☐ Standard 2
- ☐ Standard 3
- ☐ Standard 4
- ☐ Standard 5
- ☐ Repeat Standard 5
- ☐ Floating Teacher
- ☐ Specialist Teacher
- ☐ Remedial Teacher
- ☐ Other_____
- ☐ Other_____

10. Which of the following classes have you taught in the last 5 years?

- ☐ Infants 1
- ☐ Infants 2
- ☐ Standard 1
- ☐ Standard 2
- ☐ Standard 3
- ☐ Standard 4
- ☐ Standard 5
- ☐ Repeat Standard 5
- ☐ Floating Teacher
- ☐ Other_____
- ☐ Other_____

11. How many students are in your current class?

- ☐ 0-10
- ☐ 11-20
- ☐ 21-30
- ☐ 31-40
- ☐ 41-50
- ☐ >50

12. In your own words, please describe what the Continuous Assessment Programme (CAP) means to you

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

About the CAP

13. At what school did you first experience the Continuous Assessment Programme?

- ☐ At my current school
- ☐ At another school (Name of school)

  _____

14. Have you received formal training specifically for the Continuous Assessment Programme?

- ☐ Yes
- ☐ No

15. If YES, who was the service provider? (Tick as many as you wish)

- ☐ The Ministry of Education
- ☐ My School PD programme
- ☐ The University of the West Indies
- ☐ The Teachers' College
- ☐ Other
  Specify _____

16. If YES, how would you judge the effectiveness of that training?

- ☐ Very Effective
- ☐ Mostly Effective
- ☐ Somewhat Effective
- ☐ Minimally Effective
- ☐ Not Effective at all

17. Which ONE of the following best describes your current and overall level of use of the Continuous Assessment Programme? (Tick ONE only)

- ☐ I have little or no knowledge of CAP, no involvement with it, and I am doing nothing toward becoming involved.
- ☐ I am seeking or acquiring information about CAP.
- ☐ I am preparing for the first use of CAP.
- ☐ I focus most effort on the short-term, day-to-day use of CAP with little time for reflection. My effort is primarily directed toward mastering tasks required to use the CAP.
- ☐ I feel comfortable using CAP. However, I am putting forth little effort and thought to improve CAP or deal with its consequences.
- ☐ I vary the use of CAP to increase the expected benefits within the classroom. I am working on using CAP to maximize the effects with my students.
- ☐ I am combining my own efforts in using CAP with related activities of other teachers and colleagues to achieve greater impact in the classroom.
- ☐ I reevaluate the quality of use of CAP, seek major modifications of, or alternatives to, present innovation to achieve increased impact, examine new developments in the field, and explore new goals for myself and my school or education district.

18. Please indicate the extent to which you currently engage in the following activities. (Tick in the appropriate box)

----------Activity----------	Never	Rarely	Some-times	Often	Always
22. Conduct ongoing, continuous testing					
23. Complete Cumulative Record Cards					
24. Complete Student Admission Records					
25. Complete Student Performance Records					
26. Give feedback to students based on in-class assessment results					
27. Set goals/standards for learners after pre-assessing their skills					
28. Uses multiple forms of delivery in instruction					
29. Integrate assessment with teaching					
30. Use informal classroom assessments such as observation					
31. Use different types of assessments in the classroom					
32. Use performance assessment such as projects and portfolios					
33. Keep Anecdotal Records of students					
34. Use assessments diagnostically					
35. Recommend screening and referral based on assessment data.					
36. Use assessment data to plan classes.					
37. Use assessment data to advise parents and students.					
38. Record information on the class register.					
39. Keep a Teacher Journal.					

19. How often did you use each of the following assessments in the last school year? (Tick in the appropriate box)

----------Assessment-------	Never	Once Each Year	Once Each Term	Monthly	Weekly	Daily
True-False/ Alternate						
Matching						
MCQs						
Short Answers						
Essay Type						
Portfolios						
Projects						
Oral Presentations						
Exhibitions						
Performance-Song/ Play/ Dance						
Other?						

20. In the last five years, for which of the following assessments have you received any sort of training? (Tick in the appropriate box)

---------Assessment --------	No training	A bit of training	Some Training	Lots of Training
True-False/ Alternate				
Matching				
MCQs				
Short Answers				
Essay Type				
Portfolios				
Projects				
Oral Presentations				
Exhibitions				
Performance-Song/Play/Dance				
Other?				

Organizational Innovation Scale

Instructions: Here are 14 statements about the way institutions and people deal with change. Please indicate the extent to which you agree with each statement by ticking in the appropriate box.

---------- Statement ----------	Strongly Disagree						Strongly Agree
	1	2	3	4	5	6	7
1. Experimenting with innovative ideas will get me nowhere in this school.							
2. This school will undertake something innovative only when it is clear that it will be successful.							
3. If I make a mistake at work, it will be held against me.							
4. In this school, it is important to do things right without ever making mistakes.							
5. By and large, the way we do things in this school will remain unchanged in the next 5 years.							
6. Leaders of this school favourably recognize those who try innovative things, even if they fail.							
7. The leaders of this school emphasize the importance of being on the cutting edge of innovation.							
8. The leaders of this school tend to uphold well-tested, traditional ways of doing things.							
9. When it comes to the work I do, there is usually one best way to achieve a particular outcome.							
10. New methods with uncertain potential for success should be avoided at all costs.							
11. I try to reduce unpredictability and uncertainty in my work as much as possible.							
12. Errors and mistakes are signs of failure.							
13. I try innovative techniques in my work, even if I don't know whether they will work or not.							

Classroom Assessment Practice Inventory

Instructions: Following are 67 classroom assessment practices. Please indicate the extent to which you use and are skilled in each practice by ticking in the appropriate box. Some skills might not be practiced.

Use Scale:
1 = not at all used, 2 = seldom used, 3 = used occasionally, 4 = used often, and 5 = used very often
Skill Scale:
1 = not at all skilled, 2 = a little skilled, 3 = somewhat skilled, 4 = skilled, and 5 = very skilled

----------ASSESSMENT PRACTICE----------	Use					Skill				
	1	2	3	4	5	1	2	3	4	5
1. Choosing appropriate assessment methods for instructional decisions.	1	2	3	4	5	1	2	3	4	5
2. Selecting textbook-provided test items for classroom assessment.	1	2	3	4	5	1	2	3	4	5
3. Revising previously produced teacher-made tests to match current instructional emphasis.	1	2	3	4	5	1	2	3	4	5
4. Administering announced quizzes.	1	2	3	4	5	1	2	3	4	5
5. Administering unannounced quizzes.	1	2	3	4	5	1	2	3	4	5
6. Evaluating oral questions from students.	1	2	3	4	5	1	2	3	4	5
7. Assessing students through observation.	1	2	3	4	5	1	2	3	4	5
8. Determining if a large-scale achievement test is valid for classroom assessment.	1	2	3	4	5	1	2	3	4	5
9. Using a table of specifications to plan assessments.	1	2	3	4	5	1	2	3	4	5
10. Developing assessments based on clearly defined course objectives.	1	2	3	4	5	1	2	3	4	5
11. Matching assessments with instruction.	1	2	3	4	5	1	2	3	4	5
12. Writing paper-pencil tests.	1	2	3	4	5	1	2	3	4	5
13. Writing multiple-choice questions.	1	2	3	4	5	1	2	3	4	5
14. Writing matching questions.	1	2	3	4	5	1	2	3	4	5
15. Writing true/false questions.	1	2	3	4	5	1	2	3	4	5
16. Writing fill-in-the-blank or short answer questions.	1	2	3	4	5	1	2	3	4	5
17. Writing essay questions.	1	2	3	4	5	1	2	3	4	5
18. Writing test items for higher cognitive levels.	1	2	3	4	5	1	2	3	4	5
19. Constructing a model answer for scoring essay questions.	1	2	3	4	5	1	2	3	4	5
20. Ensuring adequate content sampling for a test.	1	2	3	4	5	1	2	3	4	5
21. Matching performance tasks to instruction and course objectives.	1	2	3	4	5	1	2	3	4	5

22. Defining a rating scale for performance criteria in advance.	1	2	3	4	5	1	2	3	4	5
23. Communicating performance assessment criteria to students in advance.	1	2	3	4	5	1	2	3	4	5
24. Recording assessment result on the rating scale/checklist while observing a student's performance.	1	2	3	4	5	1	2	3	4	5
25. Using concept mapping to assess student learning.	1	2	3	4	5	1	2	3	4	5
26. Assessing individual class participation.	1	2	3	4	5	1	2	3	4	5
27. Assessing group class participation.	1	2	3	4	5	1	2	3	4	5
28. Assessing individual hands-on activities.	1	2	3	4	5	1	2	3	4	5
29. Assessing group hands-on activities.	1	2	3	4	5	1	2	3	4	5
30. Assessing individual class participation.	1	2	3	4	5	1	2	3	4	5
31. Using portfolios to assess student progress.	1	2	3	4	5	1	2	3	4	5
32. Following required procedures (time limit, no hints, no interpretation) when administering standardized tests.	1	2	3	4	5	1	2	3	4	5
33. Interpreting standardized test scores (e.g., Stanines, Percentile Ranks) to students and parents.	1	2	3	4	5	1	2	3	4	5

----------ASSESSMENT PRACTICE----------	**Use**					**Skill**				
	1	2	3	4	5	1	2	3	4	5
34. Interpreting percentiles for students and parents.	1	2	3	4	5	1	2	3	4	5
35. Calculating and interpreting central tendency and variability for teacher-made tests.	1	2	3	4	5	1	2	3	4	5
36. Conducting item analysis (difficulty & discrimination indices) for teacher-made tests.	1	2	3	4	5	1	2	3	4	5
37. Revising a test based on item analysis.	1	2	3	4	5	1	2	3	4	5
38. Obtaining diagnostic information from standardized tests.	1	2	3	4	5	1	2	3	4	5
39. Using assessment results when planning teaching.	1	2	3	4	5	1	2	3	4	5
40. Using assessment results when developing curriculum.	1	2	3	4	5	1	2	3	4	5
41. Using assessment results when making decisions (e.g., placement, promotion) about individual students.	1	2	3	4	5	1	2	3	4	5
42. Using assessment results when evaluating class improvement.	1	2	3	4	5	1	2	3	4	5
43. Using assessment results when evaluating school improvement.	1	2	3	4	5	1	2	3	4	5
44. Developing systematic grading procedures.	1	2	3	4	5	1	2	3	4	5
45. Developing a grading philosophy.	1	2	3	4	5	1	2	3	4	5

46. Using norm-referenced grading model.	1	2	3	4	5	1	2	3	4	5
47. Using criteria-referenced grading model.	1	2	3	4	5	1	2	3	4	5
48. Using systematic procedures to determine borderline grades.	1	2	3	4	5	1	2	3	4	5
49. Informing students in advance how grades are to be assigned.	1	2	3	4	5	1	2	3	4	5
50. Establishing student expectations for determining grades for special education students.	1	2	3	4	5	1	2	3	4	5
51. Weighing differently projects, exams, homework, etc. when assigning semester grades.	1	2	3	4	5	1	2	3	4	5
52. Incorporating extra credit activities in the calculation of grades.	1	2	3	4	5	1	2	3	4	5
53. Incorporating ability in the calculation of grades.	1	2	3	4	5	1	2	3	4	5
54. Incorporating classroom behaviour in the calculation of grades.	1	2	3	4	5	1	2	3	4	5
55. Incorporating improvement in the calculation of grades.	1	2	3	4	5	1	2	3	4	5
56. Incorporating effort in the calculation of grades.	1	2	3	4	5	1	2	3	4	5
57. Incorporating attendance in the calculation of grades.	1	2	3	4	5	1	2	3	4	5
58. Assigning grades.	1	2	3	4	5	1	2	3	4	5
59. Providing oral feedback to students.	1	2	3	4	5	1	2	3	4	5
60. Providing written feedback to students.	1	2	3	4	5	1	2	3	4	5
61. Communicating classroom assessment results to students.	1	2	3	4	5	1	2	3	4	5
62. Communicating classroom assessment results to parents.	1	2	3	4	5	1	2	3	4	5
63. Communicating classroom assessment results to other educators.	1	2	3	4	5	1	2	3	4	5
64. Avoiding teaching to the test when preparing students for tests.	1	2	3	4	5	1	2	3	4	5
65. Protecting students' confidentiality with regard to test scores.	1	2	3	4	5	1	2	3	4	5
66. Recognizing unethical, illegal, or otherwise inappropriate assessment methods.	1	2	3	4	5	1	2	3	4	5
67. Recognizing unethical, illegal, or otherwise inappropriate uses of assessment information.	1	2	3	4	5	1	2	3	4	5

Teachers' Conceptions of Assessment

Instructions: Please indicate to what extent you agree with each statement by ticking in the appropriate box

---------Statements about Assessment----------	Disagree		Agree			
	Strongly	Mostly	Slightly	Moderately	Mostly	Strongly
1. Assessment provides information on how well schools are doing.						
2. Assessment is an accurate indicator of a school's quality.						
3. Assessment is a good way to evaluate a school.						
4. Assessment places students into categories.						
5. Assessment is assigning a grade or level to student work.						
6. Assessment determines if students meet qualifications standards.						
7. Assessment is a way to determine how much students have learned from teaching.						
8. Assessment establishes what students have learned.						
9. Assessment measures students' higher order thinking skills.						
10. Assessment provides feedback to students about their performance.						
11. Assessment feeds back to students their learning needs.						
12. Assessment is integrated with teaching practice.						
13. Assessment information modifies ongoing teaching of students.						
14. Assessment allows different students to get different instruction.						
15. Assessment results are trustworthy.						
16. Assessment results are consistent.						
17. Assessment results can be depended on.						
18. Assessment forces teachers to teach in a way against their beliefs.						
19. Assessment is unfair to students.						
20. Assessment interferes with teaching.						
21. Teachers conduct assessments but make little use of the results.						

22. Assessment results are filed and ignored.						
23. Assessment has little impact on teaching.						
24. Assessment results should be treated cautiously given measurement error.						
25. Teachers should take into account the error and imprecision in all assessment.						
26. Assessment is an imprecise process.						

Stages of Concern CAP Questionnaire

Instructions
The purpose of this questionnaire is to determine what teachers, who are using or even thinking about using CAP, are most concerned about.

The items were developed from typical responses of school teachers who ranged from no knowledge at all about various programs to many years' experience using them. Therefore, **many of the items may appear to be of little relevance or irrelevant to you at this time.**

For the completely irrelevant items, please circle "0" on the scale. Other items will represent those concerns that you do have, in varying degrees of intensity, and should be marked higher on the scale.

For example:

This statement is very true of me at this time.	0 1 2 3 4 5 6 ~~7~~
This statement is somewhat true of me now.	0 1 2 3 ~~4~~ 5 6 7
This statement is not at all true of me at this time.	0 ~~1~~ 2 3 4 5 6 7
This statement seems irrelevant to me.	~~0~~ 1 2 3 4 5 6 7

Please respond to the items in terms of **your present concerns,** or how you feel about your involvement with **the Continuous Assessment Programme at this time**. We do not hold to any one definition of an innovation like CAP, so please think of it in terms of your own perceptions of what is involved or required.

Remember to respond to each item in terms of your present concerns about your involvement or potential involvement with the CAP using the scale below. REMEMBER TO CIRCLE ONLY ONE NUMBER.

0	1	2	3	4	5	6	7
Irrelevant	Not True of Me Now		Somewhat True of Me Now			Very True of Me Now	

Statement								
1. I am concerned about students' attitudes toward CAP.	0	1	2	3	4	5	6	7
2. I now know of some other approaches that might work better.	0	1	2	3	4	5	6	7
3. I am more concerned about another innovation.	0	1	2	3	4	5	6	7
4. I am concerned about not having enough time to organize myself each day.	0	1	2	3	4	5	6	7
5. I would like to help other staff members in their use of CAP.	0	1	2	3	4	5	6	7
6. I have a very limited knowledge of the CAP.	0	1	2	3	4	5	6	7
7. I would like to know the effect of reorganization on my professional status.	0	1	2	3	4	5	6	7
8. I am concerned about conflict between my interests and my responsibilities.	0	1	2	3	4	5	6	7
9. I am concerned about revising my use of the innovations in CAP.	0	1	2	3	4	5	6	7
10. I would like to develop working relationships with both our staff and staff from other schools using CAP.	0	1	2	3	4	5	6	7
11. I am concerned about how CAP affects students.	0	1	2	3	4	5	6	7
12. I am not concerned about CAP at this time.	0	1	2	3	4	5	6	7
13. I would like to know who will make the decisions in the new system.	0	1	2	3	4	5	6	7
14. I would like to discuss the possibility of using the innovation.	0	1	2	3	4	5	6	7
15. I would like to know what resources are available if we decide to adopt CAP.	0	1	2	3	4	5	6	7
16. I am concerned about my inability to manage all that CAP requires.	0	1	2	3	4	5	6	7
17. I would like to know how my teaching or administration is supposed to change.	0	1	2	3	4	5	6	7
18. I would like to familiarize other departments or persons with the progress of CAP.	0	1	2	3	4	5	6	7
19. I am concerned about evaluating my impact on students.	0	1	2	3	4	5	6	7
20. I would like to revise the CAP's approach.	0	1	2	3	4	5	6	7
21. I am preoccupied with things other than CAP.	0	1	2	3	4	5	6	7
22. I would like to modify our use of CAP based on the experiences of our students.	0	1	2	3	4	5	6	7
23. I spend little time thinking about CAP.	0	1	2	3	4	5	6	7
24. I would like to excite my students about their part in CAP.	0	1	2	3	4	5	6	7
25. I am concerned about time spent working with non-academic problems related to the innovation.	0	1	2	3	4	5	6	7
26. I would like to know what the use of the innovation will require in the immediate future.	0	1	2	3	4	5	6	7

27. I would like to coordinate my efforts with others to maximize the innovation's effects.	0	1	2	3	4	5	6	7
28. I would like to have more information on time and energy commitments required by the innovation.	0	1	2	3	4	5	6	7
29. I would like to know what other schools are doing in this area.	0	1	2	3	4	5	6	7
30. Currently, other priorities prevent me from focusing my attention on CAP.	0	1	2	3	4	5	6	7
31. I would like to determine how to supplement, enhance, or replace the innovation.	0	1	2	3	4	5	6	7
32. I would like to use feedback from students to change the program.	0	1	2	3	4	5	6	7
33. I would like to know how my role will change when I am using CAP.	0	1	2	3	4	5	6	7
34. Coordination of tasks and people is taking too much of my time.	0	1	2	3	4	5	6	7
35. I would like to know how CAP is better than what we have now.	0	1	2	3	4	5	6	7

Leading CAP

Here are some statements about leading the CAP. Leadership may come from a variety of sources, including the principal, vice-principal, department head, senior teacher, or even from a colleague sent for training. Respond to leadership in general by ticking in the appropriate box.

--Statements about Assessment Leadership--	Disagree			Agree		
	Strongly	Moderately	Slightly	Slightly	Moderately	Strongly
1. Leaders were familiar with different approaches to assessment.						
2. Leaders had a sound knowledge of the principles of test construction						
3. Leaders were familiar with recent ideas and practices related to assessment.						
4. Leaders understood their own level of assessment literacy.						
5. Leaders knew where to obtain information and support concerning modern assessment practices.						
6. Leaders considered assessment leadership to be a core aspect of instructional leadership.						
7. Leaders considered teaching, learning and assessment to be intertwined.						
8. Leaders understood the role of classroom and student context in what is taught and tested.						
9. Leaders placed value on students work as an indicator of what students know and can do.						
10. Leaders engaged teachers in reflective dialogue about assessment practices.						
11. Leaders were able to gauge the levels of assessment literacy among the teaching staff.						

12. Leaders worked with teachers to develop school and classroom action plans using student assessment data.						
13. Leaders supported teachers in assessment using instructional supervision.						
14. Leaders entered into meaningful debate about the relative value of different assessment practices.						
15. Leaders were able to assess adequately the assessment practices throughout classrooms in the school.						
16. Leaders promoted discussion of assessment practices and beliefs beyond the immediate school environment.						
17. Leaders conducted a critical examination of their own assessment literacy.						

Feedback to Students Scale

INSTRUCTIONS: Here are some ways to give feedback to students within a classroom. For the teaching period in the last school term, please indicate how often you practiced each strategy within your classroom by ticking in the appropriate box.

----------Statements----------	Never	Once Per Term	Monthly	Weekly	Daily
1. I give an overall score or grade on the assessment					
2. I tell learners about the correctness of each response and whether they are wrong or right					
3. I provide the correct answers or solutions to problems and questions that I present.					
4. I use the "repeat-until-correct" strategy where I tell the learner the response is incorrect and then he or she has to try to correct it.					
5. I work closely with students to help them highlight specific errors in their solutions or answers.					
6. I provide detailed feedback by focusing primarily upon the target concept or required skill.					
7. I provide detailed feedback by focusing primarily upon the topic being taught.					
8. I provide detailed feedback by explaining why an answer is wrong or right.					
9. I provide detailed feedback with hints on what to do next or I demonstrate what to do, but I don't always provide the right answer.					

10. I provide detailed feedback by focusing upon the learners' errors and misconceptions.						
11. I explain the correctness of students' responses, helping them to locate the errors and giving hints about the best possible approach to obtaining the correct solution.						

The CAP Attitude Scale

INSTRUCTIONS: Here are 25 statements about CAP. Please indicate the extent to which you either agree or disagree with each statement. (Tick in the appropriate box)

----------Statements----------	Strongly Disagree						Strongly Agree
	1	**2**	**3**	**4**	**5**	**6**	**7**
1. When students do projects in my class, most of the work should be done within the classroom setting.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
2. I use performance assessments frequently in my classroom.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
3. A variety of assessment formats must be used for classroom assessment to be considered "good".	**1**	**2**	**3**	**4**	**5**	**6**	**7**
4. Peer assessment can't be used in the primary school because pupils do not mark each other properly.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
5. When you give a project, it is best to give them feedback as they do the activity or task.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
6. CAP can't work in my classroom because there are too many children	**1**	**2**	**3**	**4**	**5**	**6**	**7**
7. CAP can't work in my school because there are too many students performing poorly.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
8. CAP will require me to spend too much time outside my normal working hours.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
9. My school does not have enough resources or the right kind of leadership to support CAP.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
10. CAP simply interferes with real teaching and learning.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
11. Projects should be given for homework to support classroom instruction	**1**	**2**	**3**	**4**	**5**	**6**	**7**
12. I am interested and excited about doing CAP.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
13. I construct rubrics to score my assessment projects.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
14. I often use observations and anecdotal records to refer students for screening or diagnosis.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
15. For CAP to work well there must be greater parental support in my school.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
16. CAP has too many tedious and time-consuming tasks.	**1**	**2**	**3**	**4**	**5**	**6**	**7**

17. I find it impossible to full out all of those forms and records in CAP.	1	2	3	4	5	6	7
18. Doing projects have helped my students learn and make greater progress.	1	2	3	4	5	6	7
19. We work together to plan our projects in this school.	1	2	3	4	5	6	7
20. I don't have enough time to explain some of the work when they do a test or assessment.	1	2	3	4	5	6	7
21. You can't assess all students.	1	2	3	4	5	6	7
22. When I was in school, there was no continuous assessment and that was a good thing.	1	2	3	4	5	6	7
23. I try to mark some assessments but I never get all done.	1	2	3	4	5	6	7
24. There are too many students in my class to assess properly.	1	2	3	4	5	6	7
25. CAP simply interferes with preparation for the SEA (Secondary Entrance Assessment).	1	2	3	4	5	6	7

Conceptions of Curriculum & Teaching

Instructions: Here are different statements about teaching and curriculum frequently held by teachers. Please indicate the extent to which you agree or disagree with each statement by placing a tick in the appropriate box.

----Statements about Curriculum and Teaching---	Disagree		Agree			
	Strongly	Mostly	Slightly	Moderately	Mostly	Strongly
1. Curriculum should stress refinement of intellectual abilities.						
2. The ultimate goal of school curriculum should help students to achieve self-actualization						
3. School curriculum should aim at developing students' rational thinking.						
4. Curriculum should let students understand societal problems and take action to establish a new society.						
5. Existing problems in our society should be the organizing centre of the curriculum.						
6. Curriculum contents should focus on societal problems like pollution, racial discrimination, corruption, and crime.						
7. Curriculum and instruction should focus on finding efficient means to a set of predetermined learning objectives.						
8. Selection of curriculum content and teaching activities should be based on the learning objectives of a particular subject.						
9. I link the subject matter with real settings of practice or application.						

10. My intent is to help people develop more complex ways of reasoning.						
11. Teaching should focus on developing qualitative changes in thinking.						
12. I encourage expressions of feeling and emotion.						
13. My intent is to build student's self-confidence and self-esteem as learners.						
14. In my teaching, building self-confidence in learners is a priority.						
15. I help people see the need for changes in society.						
16. I expect people to be committed to changing our society.						
17. Individual learning without social change is not enough.						
18. My intent is to prepare people for examinations.						
19. Effective teachers must first be experts in their own subject areas.						

Collective Teacher Efficacy

INSTRUCTIONS: Please indicate the strength of your agreement with each statement below by placing a tick in the appropriate box. *The statements are about teachers in YOUR school.*

----------Statements----------	Strongly Disagree	←					Strongly Agree
	1	**2**	**3**	**4**	**5**	**6**	**7**
1. Teachers in this school have what it takes to get the children to learn.							
2. Teachers in this school are able to get through to difficult students.							
3. If a child doesn't learn something the first time, teachers will try another way.							
4. Teachers here are confident they will be able to motivate their students.							
5. Teachers in this school really believe every child can learn.							
6. If a child doesn't want to learn, teachers here give up.							
7. Teachers here need more training to know how to deal with these students.							
8. Teachers in this school think there are some students that no one can reach.							
9. Teachers here don't have the skills needed to produce meaningful student learning.							
10. Teachers here fail to reach some students because of poor teaching methods.							

11. These students come to school ready to learn.							
12. Homelife provides so many advantages they are bound to learn.							
13. The lack of instructional materials and supplies makes teaching very difficult.							
14. Students here just aren't motivated to learn.							
15. The quality of school facilities here really facilitates the teaching and learning process.							
16. The opportunities in this community help ensure that these students will learn.							
17. Teachers here are well prepared to teach the subjects they are assigned to teach.							
18. Teachers in this school are skilled in various methods of teaching.							
19. Learning is more difficult at this school because students are worried about their safety.							
20. Drug and alcohol abuse in the community make learning difficult for students here.							
21. Teachers in this school do not have skills to deal with student disciplinary problems.							

Individual Extra-Role Behaviour

Instructions: To what extent do the following statements apply to YOU? (Tick in the appropriate box)

----------Statements----------	Disagree			Agree		
	Strongly	Moderately	Slightly	Slightly	Moderately	Strongly
1) Volunteer for school committees.						
2) Stay after school hours to help students with class material.						
3) Orient new teachers even though it may not be required.						
4) Arrive early for classes.						
5) Organize social activities for school.						
6) Volunteer for roles and tasks that are not required.						
7) Acquire expertise in new subjects that contribute to my work.						
8) Stay in class during breaks in order to listen to my students.						
9) Offer my colleagues work sheets and materials that I have prepared for my class.						
10) Go to school on my free days to prevent problems in my class.						

11) Assume responsibilities that are not a prescribed part of my job.						
12) Prepare special assignments for higher and lower level students.						
13) Prepare record and forecast and materials for replacement teachers.						
14) Participate in private celebrations of my students (e.g., birthdays).						
15) Make innovative suggestions to improve the school.						
16) Attend functions not required but which may help the school's image.						
17) Invite students to my home.						
18) Help other teachers who have heavy workloads.						
19) Organize joint activities with parents beyond that which is expected.						
20) Decorate the school.						
21) Help an absent colleague by assigning learning tasks to the class.						
22) Participate actively in teachers' meetings.						
23) Assist the principal in my free hours						
24) Work collaboratively with others (planning assignments, joint projects, etc.)						

Readiness for Change Scale

INSTRUCTIONS: Here are 25 statements about change in a school. If CAP were to be re-introduced, how would you respond to each statement (Tick in the appropriate box)?

----------Statements----------	**Strongly Disagree**						**Strongly Agree**
	1	**2**	**3**	**4**	**5**	**6**	**7**
1. I think that the school will benefit from this change.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
2. It doesn't make much sense for us to initiate this change.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
3. There are legitimate reasons for us to make this change.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
4. This change will improve our school's overall efficiency.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
5. There are a number of rational reasons for this change to be made.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
6. In the long run, I feel it will be worthwhile for me if the school adopts this change.	**1**	**2**	**3**	**4**	**5**	**6**	**7**
7. This change makes my job easier.	**1**	**2**	**3**	**4**	**5**	**6**	**7**

8. When this change is implemented, I don't believe there is anything for me to gain.	1	2	3	4	5	6	7
9. The time we are spending on this change should be spent on something else.	1	2	3	4	5	6	7
10. This change matches the priorities set by our school.	1	2	3	4	5	6	7
11. Our senior leaders will encourage all of us to embrace this change.	1	2	3	4	5	6	7
12. Our organization's top decision makers will put all their support behind this change effort.	1	2	3	4	5	6	7
13. Every senior teacher will stress the importance of this change.	1	2	3	4	5	6	7
14. This organization's principal will be committed to this change.	1	2	3	4	5	6	7
15. I think we will spend a lot of time on this change when the senior teachers don't even want it implemented.	1	2	3	4	5	6	7
16. Management will send a clear signal that this organization is going to change.	1	2	3	4	5	6	7
17. I do not anticipate any problems adjusting to the work I will have when this change is adopted.	1	2	3	4	5	6	7
18. There are some tasks that will be required when we change that I don't think I can do well.	1	2	3	4	5	6	7
19. When we implement this change, I feel I can handle it with ease.	1	2	3	4	5	6	7
20. I have the skills that are needed to make this change work.	1	2	3	4	5	6	7
21. When I set my mind to it, I can learn everything that will be required when this change is adopted.	1	2	3	4	5	6	7
22. My past experiences make me confident that I will be able to perform successfully after this change is made.	1	2	3	4	5	6	7
23. I am worried I will lose some of my status in the school when this change is implemented.	1	2	3	4	5	6	7
24. This change will disrupt many of the personal relationships I have developed.	1	2	3	4	5	6	7
25. My future in teaching will be limited because of this change.	1	2	3	4	5	6	7

Group OCB Scale

INSTRUCTIONS: Please respond to each of the following statements about YOUR SCHOOL by placing a tick in the appropriate box

----------Statements----------	Disagree			Agree		
	Strongly	Moderately	Slightly	Slightly	Moderately	Strongly
1) Teachers here complete assigned duties adequately.						
2) Teachers here fulfill responsibilities specified in their job description.						
3) Teachers here fulfill expectations.						

4) Teachers here meet the formal performance requirements of the job.						
5) Teachers here engage in activities that will directly affect their performance evaluation.						
6) Teachers here neglect aspects of the job they are obligated to perform.						
7) Teachers here perform essential duties successfully.						
8) Teachers here help others who have been absent.						
9) Teachers here help teachers who have heavy workloads.						
10) Teachers here assist the principal with his or her work (when not asked).						
11) Teachers here take time to listen to other teachers' problems and worries.						
12) Teachers here go out of their way to help new teachers.						
13) Teachers here take personal interest in other teachers.						
14) Teachers here pass along information to other teachers.						
15) Teachers' attendance at work is above the norm (etc., staying after school hours to help students).						
16) Teachers here give advance notice when unable to come to work.						
17) Teachers here arrive at work on time and do not return late after work breaks.						
18) Teachers here spend a great deal of time on issues irrelevant to work.						
19) Teachers here complain about insignificant things at work.						
20) Teachers here conserve and protect the school's property.						
21) Teachers here help other teachers and parents who have no formal interactions with them.						
22) Teachers here cover for fellow teachers.						
23) Teachers here have a strong volunteer orientation.						
24) Teachers here make innovative suggestions to improve school life.						
25) Teachers here coast toward the end of the day.						

Professional Learning Community

Instructions: For each of the seven areas (organized vertically), place your school in the category that best describes your experiences at your school by crossing out the appropriate boxes.

1	School personnel view the mission of the school as teaching rather than learning.	An attempt has been made to identify learning outcomes for all classes or subject areas but this attempt has not impacted much on the practice of most teachers. Teachers respond to students who are not learning at their own discretion.	Teachers are clear regarding the learning outcomes their students are to achieve. Teachers are developing strategies to assess student mastery of these outcomes, monitor the results, and attempt to respond to students who are not learning.	Learning outcomes are clearly articulated and each student's attainment of the outcomes is carefully monitored. Practices, programs and policies of the school are continually assessed based on their impact on learning.
2	No effort has been made to engage staff in describing preferred conditions for their school.	A vision statement has been developed for the school, but most staff are unaware of, or unaffected by it.	Staff members are aware of the vision statement, endorse it, and feel a sense of ownership in it.	Staff members routinely articulate the major principles of the shared vision and use those principles to guide their day-to-day efforts and decisions.
3	No effort has been made to engage the staff in setting and defining school improvement goals related to student learning. If goals exist, they have been developed by the administration.	Staff members have participated in a process to establish goals but the goals are typically stated as projects to be accomplished or are written so broadly that they are impossible to measure. The goals do not yet influence instructional decisions in a meaningful way.	Staff members have worked together to establish long and short-term improvement goals for their school. The goals are clearly communicated. Assessment tools and strategies have been developed and implemented to measure progress toward the goals.	All members of staff pursue measurable school improvement goals as part of their routine responsibilities. Goals are clearly linked to the school's shared vision. Goal attainment is celebrated and staff members are willing to identify and pursue challenging stretch goals.
4	Teachers work in isolation. There is little awareness of what or how colleagues are teaching.	Teachers recognize a common curriculum that they are responsible for teaching but there is little exchange of ideas regarding instructional materials, teaching strategies or methods of assessment.	Teachers function in work groups that meet periodically to complete certain "business" tasks such as reviewing intended outcomes and coordinating calendars.	Teachers function as a team. They work collaboratively to identify collective goals, develop strategies to achieve those goals, gather relevant data and learn from one another.

Professional Learning Community (cont.)

For each of the seven areas (organized vertically), place your school in the category that best describes your experienced by ticking in the appropriate boxes.

5	Questions of power are a continuing source of controversy and friction. Relationships between teachers and administrators are often adversarial.	Efforts have been made to reduce friction by clarifying "administrator roles" and "teacher roles". Both parties are protective of intrusion on to their turf.	Administrators solicit and value teacher input as improvement initiatives are developed and considered but administrators are regarded as having primary responsibility for school improvement.	Administrators pose questions, delegate authority, create collaborative decision-making processes and provide staff with the information and training they need to be involved in making decisions.
6	Little attention is devoted to creating systems that enable either the school or individual teachers to track improvement.	A few people in the school are tracking general indicators of achievement, such as standardized tests. Results are then reported, but not acted upon.	Individual teachers and teaching teams gather information that enables them to identify and monitor individual and team goals.	Everyone in the school participates in reviewing data to identify discrepancies between actual and desired results, develop strategies to achieve the goals, and track improvement indicators.
7	The results the school seeks for each student have not been identified.	Results have been identified but are stated in such broad terms that they are impossible to measure. Improvement initiatives focus on inputs - projects or tasks to be completed, rather than on student achievements.	Desired results have been identified in terms of student outcomes and student achievement indicators have been identified. Data is being collected and monitored within the school system. Results of the analysis are shared with teachers.	Teams of teachers value and seek out information on results. They gather relevant data and use these data to identify improvement goals and to monitor progress towards goals.

Assessment Knowledge

INSTRUCTIONS: Here are 16 situations, each with 5 possible different possibilities. Choose the one that you think is right.

1) *A fourth standard teacher wants a classroom assessment to capture students' progress on literacy, provide interest-arousing motivation, and offer in-depth information on strengths and weaknesses.* **Which assessment is the best choice?**

- ☐ A reading comprehension assessment with 50 multiple-choice questions
- ☐ A reading comprehension assessment with 30 short answer questions
- ☐ A reading comprehension assessment with 30 multiple-choice and 20 short answer questions
- ☐ Observation of the students' classroom literacy activities supplemented by anecdotal records
- ☐ A literacy portfolio with multiple samples of writing, reading, and spelling

2) *A second standard teacher wants a reliable and valid classroom assessment so that he can report to students and parents on learning in science for the entire year.* **Which assessment should he use?**

- ☐ A science assessment with 50 multiple-choice questions
- ☐ A science assessment with 30 short answer questions and a project
- ☐ A science assessment with 30 multiple-choice, 20 short answer , and a project
- ☐ An extended project and exhibition of the most exciting topics
- ☐ A portfolio of the best science projects

3) What is the best grade to apply to a score of 60%?

- ☐ A
- ☐ B
- ☐ C
- ☐ D
- ☐ None of the above

4) Which of the following is the best strategy for assigning grades?

- ☐ Assign fixed numbers of students to each grade
- ☐ Use the number right raw score
- ☐ Use the percentage score
- ☐ Use criteria based on knowledge and skills
- ☐ None of the above

5) *A fourth standard teacher develops multiple-choice and short-answer questions for a reading comprehension test.* **Which strategy will improve the meaningfulness of information provided by the scores?**

- ☐ Choose the best items from national tests
- ☐ Choose the best items from the Secondary Entrance Assessment
- ☐ Choose the best items from the language arts textbook
- ☐ Develop a table of specifications
- ☐ Organize the items in the test from easiest to most difficult

6) *A teacher has developed a portfolio of students' written work in social studies to be used for summative assessment.* **Which is the best strategy to improve validity and reliability?**

- ☐ Use checklists
- ☐ Use a scoring key
- ☐ Use a rubric
- ☐ Include student self assessment scores
- ☐ Include student peer assessment scores

The following is the scoring device used for the SEA Creative Writing Component. Answer the questions that follow

SEA ESSAY - MARK SCHEME
Criteria for scoring:

SCORE CRITERIA

9 - 10 An outstanding response
- excellent development of ideas
- effective organization
- fluent, "lively" writing
- effective variety in sentence structure
- striking word choice
- accurate sentence structure, grammar, mechanics
- excellent in reader audience appeal

7 - 8 A good response
- ideas are well developed
- good organization
- fluent writing
- competent word choice
- generally accurate sentence structure, grammar
- mechanics
- some reader appeal

5 - 6 A competent response
- fairly good development of ideas; supporting detail may not be especially well selected or relevant
- some attempt at organization
- generally accurate structure, grammar, mechanics (errors are not intrusive)
- writing not especially lively or interesting

3 - 4 A weak response
- the ideas are conveyed but they are either not well developed or loosely organized
- errors (grammar, structure, mechanics, expression) occur frequently

2 A poor response
- ideas are confusing and undeveloped
- organization is loose or non - existent
- errors occur so frequently as to intrude in the reader's enjoyment of the story

1 A very poor response
- writing is essentially incoherent because of the writer's inability to communicate in English

0 An extremely poor response or no response
- writing is incomprehensible
- words cannot be recognized

7) The instrument above used to score the SEA Creative Writing essay is a:

- ☐ Checklist
- ☐ Point Score Key
- ☐ Holistic Rubric
- ☐ Analytical Rubric
- ☐ Trait Rubric

8) Which strategy will improve the ability of the SEA instrument to discriminate between students who are good and poor writers?

- ☐ Increase the total score
- ☐ Decrease the total score
- ☐ Increase the number of score points
- ☐ Decrease the number of score points
- ☐ Write more detailed descriptors

9) Which data set should be used for making a decision to retain a student at the same year level:

- ☐ Observation & Anecdotal Records
- ☐ Standardized Diagnostic Tests
- ☐ Teachers' Classroom Tests
- ☐ National Tests
- ☐ Secondary Entrance Assessment

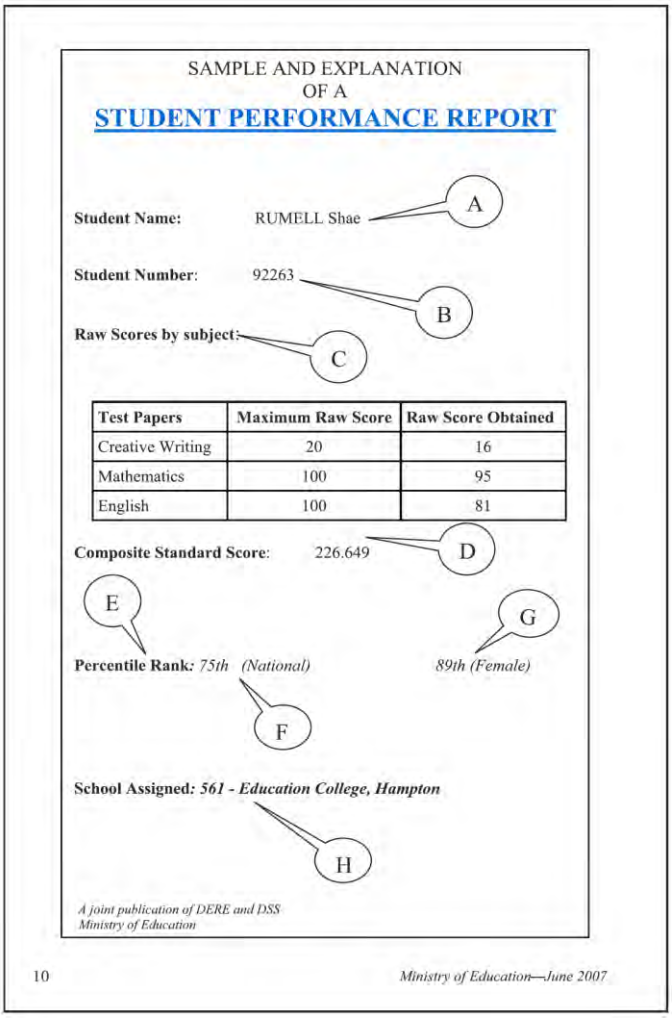
10) Which data set should be used to direct a student to remediation:

- ☐ Observations &Anecdotal Records
- ☐ Standardized Diagnostic Tests
- ☐ Teachers' Classroom Tests
- ☐ National Tests
- ☐ Secondary Entrance Assessment

11) Which of the following is the most appropriate use of data from national tests

- ☐ To determine the effectiveness of individual teachers
- ☐ To compare the effectiveness of different classrooms
- ☐ To determine the ability of individual students
- ☐ To make important decisions about individual students
- ☐ To indicate challenges and allow the setting of targets by the school

The following is a sample of an SEA Report. Questions 13 and 14 apply



SAMPLE AND EXPLANATION OF A

STUDENT PERFORMANCE REPORT

Student Name: RUMELL Shae (A)

Student Number: 92263 (B)

Raw Scores by subject: (C)

Test Papers	Maximum Raw Score	Raw Score Obtained
Creative Writing	20	16
Mathematics	100	95
English	100	81

Composite Standard Score: 226.649 (D)

(E) **Percentile Rank:** *75th (National)* (F) (G) *89th (Female)*

School Assigned: *561 - Education College, Hampton* (H)

A joint publication of DERE and DSS Ministry of Education

10 *Ministry of Education—June 2007*

12) *A teacher is explaining the above report to a parent.* **Which might be the most accurate explanation?**

- ☐ The student has an overall score of 75%
- ☐ The student has an overall score of 89%
- ☐ The student scores higher than 75% of the females in the country
- ☐ The student scored higher than 75% of the males and females in the country
- ☐ The students scores higher than 89% of the males and females in the country

13) Which score is used for assigning students to secondary school?

- ☐ Raw score
- ☐ National Percentile Rank
- ☐ Male of female Percentile Rank
- ☐ Composite standard score
- ☐ None of the above

14) What information is provided in the Secondary Entrance Assessment?

- ☐ Students' opportunity to succeed in secondary school
- ☐ Students opportunity to learn in the primary school
- ☐ Students overall academic ability
- ☐ Student's overall ability to learn
- ☐ Students' native intelligence

15) Which assessment data is best for grouping students into different classes of different abilities?

- ☐ Observations and Anecdotal Records
- ☐ Standardized Diagnostic Tests
- ☐ Teachers' Classroom Tests
- ☐ National Tests
- ☐ Students should not be grouped into different ability groups

16) When is it appropriate to provide practice for the national tests?

- ☐ If the students are weak
- ☐ If the school is doing poorly
- ☐ When parents cannot by the test booklets
- ☐ If the students need practice on the item format
- ☐ This is the best way to prepare students

---------- END OF QUESTIONNAIRE--------

THANK YOU