



**Consortium for Research on
Educational Access,
Transitions and Equity**

Education For All, The Quality Imperative and the Problem of Pedagogy

Robin Alexander

**CREATE PATHWAYS TO ACCESS
Research Monograph No 20**

April 2008



**Leading education
and social research**
Institute of Education
University of London

**Institute of Education
University of London**



Consortium for Research on
Educational Access, Transitions & Equity
Funded by DFID

The Consortium for Educational Access, Transitions and Equity (CREATE) is a Research Programme Consortium supported by the UK Department for International Development (DFID). Its purpose is to undertake research designed to improve access to basic education in developing countries. It seeks to achieve this through generating new knowledge and encouraging its application through effective communication and dissemination to national and international development agencies, national governments, education and development professionals, non-government organisations and other interested stakeholders.

Access to basic education lies at the heart of development. Lack of educational access, and securely acquired knowledge and skill, is both a part of the definition of poverty, and a means for its diminution. Sustained access to meaningful learning that has value is critical to long term improvements in productivity, the reduction of inter-generational cycles of poverty, demographic transition, preventive health care, the empowerment of women, and reductions in inequality.

The CREATE partners

CREATE is developing its research collaboratively with partners in Sub-Saharan Africa and South Asia. The lead partner of CREATE is the Centre for International Education at the University of Sussex. The partners are:

The Centre for International Education, University of Sussex: Professor Keith M Lewin (Director)
The Institute of Education and Development, BRAC University, Dhaka, Bangladesh: Dr Manzoor Ahmed
The National University of Educational Planning and Administration, Delhi, India: Professor R Govinda
The Education Policy Unit, University of the Witwatersrand, South Africa: Dr Shireen Motala
The Universities of Education at Winneba and Cape Coast, Ghana: Professor Jerome Djangmah
The Institute of Education, University of London: Professor Angela W Little

Disclaimer

This paper, and the research on which it is based, were commissioned by the UK Department for International Development (DFID). The paper is published, with DfID's approval, as a collaboration between the author and the CREATE team at the Institute of Education, University of London. CREATE is funded by the UK Department for International Development for the benefit of developing countries and is coordinated from the Centre for International Education, University of Sussex. The views expressed are those of the author and not necessarily those of DFID, the Institute of Education, University of London or the CREATE Team.

Copyright © 2008 Robin Alexander

Cover images Copyright © 2008 Karen Lennox. Enquiries to karenlennox@zen.co.uk

ISBN: 0-901881-25-2

Address for correspondence:
Professor Robin Alexander
Faculty of Education
184 Hills Road,
Cambridge CB2 8PQ, UK

Author's email: rja40@cam.ac.uk

Website: www.robinalexander.org.uk

Please email CREATE (create@sussex.ac.uk) if you require a hard copy of this publication.

Education for All, the Quality Imperative and the Problem of Pedagogy

Robin Alexander

**CREATE PATHWAYS TO ACCESS
Research Monograph No 20**

April 2008

Contents

Acknowledgements	v
Preface.....	vi
Summary	vii
1. Introduction	1
2. Quality as Indicators.....	3
2.1 Prototypical examples	3
2.2 From Millennium Development Goals to the 2005 EFA Global Monitoring Report	6
2.3 The confusion at the heart of ‘quality’	10
2.4 The NCERT Quality Monitoring Tools (QMT).....	11
2.5 Criteria for assessing the adequacy of quality monitoring frameworks	14
2.6 Indicators of quality: who needs to know what?	16
3. Sources for Accounts of Pedagogical Quality.....	18
3.1 National educational policy and its cultural context	18
3.2 National research on pedagogy	20
3.3 The international quality indicators literature	21
3.4 International research on pedagogy.....	22
3.5 Problems with the pedagogical research cited.....	23
3.6 Pedagogical research and multigrade reality	27
3.7 Alternative research sources, definitions and frameworks	29
3.7.1 Pedagogy as ideas.....	29
3.7.2 Pedagogy as practice	30
3.7.3 Act, form and frame	30
4. The Sting in the Tail: Indicators and Measures.....	36
5. Summary	39
5.1 Quality in EFA discourse	39
5.2 Indicators of quality.....	39
5.3 Criteria for assessing quality frameworks	39
5.4 Sources for defining quality and quality indicators.....	40
5.5 Pedagogy: definitions and frameworks	41
5.6 Indicators and measures	42
6. Where Do We Go From Here?.....	43
6.1 Principles of procedure.....	43
6.2 Priorities for action.....	45
References	46

List of Figures

Figure 1: Pedagogy as ideas (theories, values, evidence and justifications)	30
Figure 2: Pedagogy as practice.....	32

List and Explanation of Acronyms

AERA	American Educational Research Association
BRC	Block Resource Centre (India). The block is the administrative level below the district, above which is the state.
CRC	Cluster Resource Centre (India). CRCS are for a cluster of villages/schools at below-block level. CRCs and BRCs were a major innovation introduced under DPEP and carried forward under SSA (see below). They support teachers and their training/development at the local level.
DfID	Department for International Development (UK government)
DISE	District Information System for Education (India)
DPEP	District Primary Education Programme (India): The Government of India's programme for universalising primary education in a limited number of targeted states and rural districts, initiated in 1994 and superseded by SSA.
EC	European Commission
ECCE	Early Childhood Care and Education
EDI	Education Development Index: Index designed to measure progress towards EFA)
EFA	Education for All
GDP	Gross Domestic Product
IRF	Initiation-response-feedback: The form of the classic classroom exchange. Sometimes called IRE, (initiation-response-evaluation) because the teacher's feedback utterance is almost always evaluative.
ISCED	International Standard Classification of Education: UNESCO-devised system for classifying national education systems according to agreed international criteria including 'levels' (primary, lower secondary, upper secondary etc) and 'fields' (e.g. humanities and arts, business and law).
ISERP	International School Effectiveness Research Project
JRM	Joint Review Missions (India) of the Government of India and its development partners, for reviewing the progress of DPEP and SSA.
MDG	Millennium Development Goals of the United Nations, the second of which is 'Achieve universal primary education'.
MHRD	Ministry of Human Resource Development (Government of India)
NCERT	National Council of Educational Research and Training (India), the New Delhi 'apex' institution responsible for research, training and curriculum development and for development work on textbooks, testing and monitoring. NCERT is at the apex of a network of four regional institutes of education (RIE) and the various State Councils of Educational Research and Training (SCERT).
NER	Net Enrolment Ratio
NGO	Non-Governmental Organization
OECD	Organization for Economic Co-operation and Development
OFSTED	Office for Standards in Education (UK body responsible for inspection of schools and other educational institutions)
PIRLs	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
PTR	Pupil-teacher ratio
QMT	Quality Monitoring Tools, devised by NCERT, India.
SC	Scheduled Castes (India): Legal definition of those formerly known as <i>dalit</i> or <i>harijan</i> , listed under India's 1950 Constitution as entitled to receive positive discrimination. SCs, along with Scheduled Tribes (ST), are historically among India's most marginalized cultural groups.

SSA	Sarva Shiksha Abhiyan (India): The Government of India's nationwide programme for the universalisation of elementary education (6-14), initiated in 2002 as the successor to the District Primary Education Programme (DPEP).
ST	Scheduled Tribes (India): Tribal communities listed under the Indian Constitution and forming about 8 per cent of the country's total population.
TIMSS	Trends in International Mathematics and Science Study (formerly Third International Mathematics and Science Study)
TLMs	Teaching-Learning Materials
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UNICEF	United Nations Children's Fund
UEE	Universal Elementary Education
UPE	Universal Primary Education

Acknowledgements

I would like to thank DfID India for inviting me to write this monograph and CREATE for agreeing to publish it. It builds not just on my work as an academic in the fields of pedagogy and comparative research, but also on many years of work in different parts of India as an independent researcher and, on behalf of the EC and DfID, as a consultant to the Government of India's District Primary Education Programme (DPEP) and Sarva Shiksha Abhiyan (SSA). Michael Ward (DfID India), Venita Kaul (World Bank India) and Angela Little (CREATE) provided helpful comments on earlier drafts, and I benefited from conversations with colleagues at NCERT in Delhi on the SSA Quality Monitoring Tools which feature in the monograph. I am also grateful for comments from anonymous reviewers.

Preface

This CREATE research monograph by Robin Alexander explores how educational quality in the classroom is conceived, indicated and measured. It also explores how it might be conceived, indicated and measured better in the future. As the author explains, the monograph was written originally for DfID India for application in the context of Sarva Shiksha Abhiyan (SSA), the Education for All programme launched by the Government of India in 2002. However the content of the monograph has considerable relevance for the conceptual and methodological work of CREATE in India and elsewhere. Central to CREATE's concerns are the Zones of Exclusion in which many children are denied access to education of quality. One such is the zone of 'silent exclusion' in which children are enrolled in and attending school but are not participating in a learning and teaching process of good quality. Such children are at risk of learning very little and are likely to become poor attenders and drop-outs.

Robin Alexander's provocative discussion of quality, and of quality indicators and quality measures in the domain of pedagogy, offers a rich and detailed framework for thinking about and planning for education of quality for all children, including those at risk currently of falling within the zone of silent exclusion. The monograph complements well the CREATE Country Analytic Review titled *Access to Elementary Education in India* by R. Govinda and M. Bandhopadhyay and CREATE Research Monograph No. 17 *Small, Multigrade Schools and Increasing Access to Primary Education in India: National context and NGO initiatives* by N. Blum and R. Diwan. We are very pleased that Robin has agreed to have his work published within this CREATE series.

Professor Angela W. Little
Institute of Education, London
CREATE Partner Institute Convenor

Summary

This paper was commissioned by DfID India for use in the particular context of Sarva Shiksha Abhiyan (SSA), the Education for All programme which the Government of India launched in 2002 after its 86th Constitutional Amendment made education from age 6-14 the fundamental right of every Indian child. For this reason, although there are many references to non-Indian developments, SSA and the work of contingent agencies such as the Delhi-based National Council of Educational Research and Training (NCERT) provide both the central focus and the context within which problematic issues to do with quality and pedagogy in the EFA context are explored.

The paper contends that EFA discourse has moved from a commitment to quality to its measurement without adequate consideration of what quality entails, particularly in the vital domain of pedagogy. Pedagogy, indeed, is often the missing ingredient in EFA discussion of quality. Meanwhile, the demand for quality indicators has left important methodological questions unanswered. This paper provides a critique of typical quality indicator frameworks from international and EFA sources. It notes a concern with input and context at the expense of process, an arbitrariness in what is focused upon, an excessive use of proxies, neglect of international pedagogical research, and fundamental confusions about the key terms ‘quality’, ‘indicators’ and ‘measures’.

After considering an important and influential Indian framework (from NCERT) for monitoring quality via indicators, the paper proposes criteria for assessing this and similar initiatives, central to which are comprehensiveness, evidential basis, validity, reliability, impact, manageability and appositeness to level and context of use. In respect of the latter, the paper presses the need for the question ‘Who needs to know what?’ to be asked at each level of the system from national to local.

The paper then investigates the empirical and conceptual basis for accounts and indicators of quality, arguing the importance of national culture and circumstance alongside international pedagogical research. The paper finds that the interface between the latter and the EFA literature is extremely weak; and that school effectiveness research is given a prominence in EFA contexts which it is not accorded elsewhere and which its flawed methodology and cultural insensitivity do not justify. Underlying the selective use of evidence is a more general failure adequately to conceptualise what is being researched, and the paper sets out a map of the territory of pedagogy at the levels of ideas and action to remedy this deficiency, illustrating its value by considering just one of the various aspects of teaching identified, classroom interaction.

Pressing home the concern about the conceptual problems of the field as it currently stands, the paper points to the tendency to treat indicators and measures as synonymous, and to the distortion which follows an excessive preoccupation with measures. It argues that to jettison indicators on the grounds that they are non-measurable is to ignore much of what matters most in the quality of learning and teaching. This exclusivity also contradicts claims that the dangers of trying to represent quality as quantity are fully understood.

Finally, the paper proposes principles of procedure to guide future work on quality indicators and measures in the EFA context.

Education For All, the Quality Imperative and the Problem of Pedagogy

1. Introduction

Since the 1990 Jomtien World Declaration on Education for All (EFA), the EFA debate has witnessed some very different national responses but at an international level four broad shifts of focus:

- from a view that primary education is no more than a filter for secondary, and that the fate of the educationally unsuccessful majority matters less than the prospects for those who make the grade, to a recognition that universalising primary education benefits the nation as well as the individual¹;
- from a preoccupation with school access, enrolment and retention - the sine qua non of EFA - to a concern for educational outcomes and quality; or from getting children into school to addressing what they learn and how;
- from treating equity and quality as separate to recognising that they are sides of the same coin, in that education for all cannot reasonably mean quality for only some;
- from the assumption that it is sufficient to define quality via a handful of mainly proxy indicators to a dawning recognition that we need to engage much more directly with what lies at the core of the educational endeavour, that is to say, with pedagogy.

It seems fair to suggest, especially since the Dakar Framework for Action of 2000, that the first three of these are now part of mainstream EFA discourse, even though there is an inevitable time-lag between international commitment, national policy, professional culture and everyday practice. Thus, setting up infrastructures for universalising basic education is one thing; universalising genuine belief in a pattern of basic education which is well conceived in its own terms, regardless of what follows it, is quite another. More tenuous, too, is the further sense of 'universalisation' implicit in the third point above: that EFA is achieved only when the learning experiences in each classroom speak with equal meaning to the condition of every child present.

The time-lag increases as we move to the final shift. As yet, there is less consensus on what 'quality' actually entails, especially when we move from the conditions for quality (infrastructure, resources, teacher supply and of course access, enrolment and retention) to the pedagogy through which educational quality is most directly mediated. Further, whereas policies for access and enrolment may be subject to the closest scrutiny, and whereas outcomes, at least as judged by test scores, may be endlessly computed and forensically compared across time and space, pedagogy remains territory which is either cautiously avoided as too complex or is incautiously blundered into as ostensibly unproblematic. Indeed, some of those who insist that specialist expertise is necessary for handling the complexities of access, enrolment, retention and outcomes – at least in as far as these can be quantified –

¹ This should not be taken as accepting the tendency to equate EFA with UPE. In India, the country which provides the main context for this paper, EFA is now realised through Sarva Shiksha Abhiyan, a programme for universalising elementary (6-14) education which was initiated in 2002 in response to a specific Constitutional Amendment that made education from 6-14 the basic human right of every Indian child. The problem is that universalisation has to start somewhere, and in the UN Millennium Development Goals it is confined to the primary stage.

exercise no such caution about pedagogy, cheerfully peddling unexamined certainties about the conditions for effective teaching and learning.

This failure properly to engage with pedagogy creates a vacuum into which are sucked a plethora of claims about what constitutes ‘best practice’ in teaching and learning and about the virtues of this or that pedagogical nostrum – group work, activity methods, joyful learning, child-centred teaching, teaching-learning materials (TLMs), personalised learning, interactive teaching and so on. Such claims, often framed by the polarised discourse of ‘teacher-centred’ vs. ‘student-centred’, are rarely discussed, let alone evaluated against hard evidence, with the result that they rapidly acquire the status of unarguable pedagogical truth and become transmuted into policy.

The whole cycle then becomes self-reinforcing. If outcomes improve this is attributed to the success of the prescribed teaching practices, which thereby become all the more firmly entrenched. But if progress is slower than we would like, then it is assumed that the prescriptions are not being correctly operationalised by the teachers, rather than that the prescriptions need to be re-assessed and perhaps reconceived by those who train the teachers and provide the frameworks of policy and resource within which they work.

It is of course never that simple. Classroom outcomes are multi-factorial. In an ambitious governmental initiative like India’s Sarva Shiksha Abhiyan (SSA) which is an elementary (6-14) rather than a UPE initiative it is the sum and interaction of the programme’s many elements – civil works, school equipment, teacher recruitment and training, community mobilisation, local ownership, bridge schools and hostels for the hard-to-reach, free midday meals, positive discrimination in favour of girls and children from scheduled castes/scheduled tribes (SC/ST), textbooks, TLMs, pedagogical renewal and so on – that makes the difference.

So, for example, in weighing non-pedagogical SSA factors like the midday meal against pedagogical factors like TLMs, who is to say which makes the most difference to learning outcomes when for children who are under-nourished regular midday meals may affect not just enrolment and retention but also the levels of alertness and concentration on which children’s engagement with the TLMs, and their larger capacity to learn, depend?

In this paper, therefore, I propose that we take two steps back in order to try to take at least one forward. I shall compare and assess different ways of approaching quality in education, and shall argue that the specific problems of monitoring the quality of pedagogical process, arguably the core of educational quality as more widely defined, are problems of conception and evidence as much as procedure.

A word of explanation: the paper refers frequently to India – and indeed has already done so in this introduction – for India, or more specifically Sarva Shiksha Abhiyan – was the context for which it was originally prepared. Although extensive reference is also made to international contexts and sources, it is not claimed that India’s particular experience of aiming to universalise first primary (6-11) and now elementary (6-14) education, or the way it has grappled with the question of quality, is generalisable to other contexts. On the other hand there is more than enough in the Indian experience which has wider if not universal resonance and even application, and the author’s own experience of working in a number of countries in addition to India should provide some kind of corrective in this regard. Beyond that, it is for readers to take from the paper what they deem appropriate and helpful to their own circumstances.

2. Quality as Indicators

2.1 Prototypical examples

Because the international debate about the quality of education has been dominated by those who operate in the domains of policy, accountability and funding rather than in the arena of practice, quality has tended to be conceived not as what it actually is but as how it can be measured. That is why ‘indicators’ have come to occupy a place in the discourse on quality which those more closely involved in children’s education might find decidedly odd; and that is why the indicators thereby nominated speak to the preoccupations of providers rather than to those of teachers and learners.

In contrast, when learners themselves are asked about educational quality they tend to talk not about test scores but about the felt experience of learning, dwelling especially on their attitude to the tasks set (interesting/boring/easy/difficult) and the degree to which they find the context of peer and teacher-student relationships supportive and rewarding (Brown and McIntyre, 1993; Rudduck et al, 2006). For children, then, the preferred indicators are affective as much as cognitive and instrumental, and in the context of EFA, where student motivation and retention remain serious concerns, we would do well to remember this. Yet focussing on affectivity alone is as conceptually and empirically restrictive as is the treating of mathematics test scores and educational outcomes as synonymous. All quality frameworks, whether technical or common-sense, are selective, but some are more spectacularly, culpably or puzzlingly selective than others.

Here are some prototypical examples of the approach to quality through indicators. First, some ‘world education indicators’ from OECD:

- Context of education
- Financial and human resources invested in education
- Access to education, participation and progression
- The learning environment and organisation of schools
- Individual, social and market outcomes of education
- Student achievement

(OECD, 2000)

That gives the familiar mix of input, process and outcome which in one way or another frames most such efforts. ‘The learning environment’ and ‘student achievement’ indicators are then elaborated as:

- teaching time
- total intended instruction time
- student absenteeism
- computers in schools and their uses
- mathematics achievement, 4th and 8th grade
- students’ attitudes to science, 4th and 8th grades
- students’ beliefs about performing well in mathematics, 4th and 8th grades

(OECD, 2000)

The limitations of this are very evident and this initial model provides a paradigm for the difficulties of the indicators approach as a whole. Teaching is reduced to just four indicators, three of which are about time while the other fixates in an arbitrary fashion on computers. Student learning across the entire curriculum is defined through just three indicators, two of which are in the same curriculum area. Such a list, we might fairly suggest, has very little value, conceptually or in application.

In the same year, the European Commission produced an ostensibly more comprehensive list of 16 ‘quality indicators’:

Attainment

- Mathematics
- Reading
- Science
- ICT
- Foreign languages
- Learning to learn
- Civics

Success and transition

- Drop out rates
- Completion of upper secondary education
- Participation in tertiary education

Monitoring of education

- Evaluation and steering of school education
- Parental participation

Resources and structures

- Education and training of teachers
- Participation in pre-primary education
- Numbers of students per computer
- Educational expenditure per student

(EC, 2000)

The list is more comprehensive, that is, until one looks at the small print. The entire set is construed as input and outcome, with process nowhere to be seen. Further, when we examine the measures which are believed to be a necessary concomitant of indicators (I shall return to this important distinction towards the end of this paper), some of them are undeniably crude. For example, attainment in civics is measured by calculating the percentage of 24 year olds who agree with the statement ‘I’m glad that foreigners live in our country’ and disagree with ‘All foreigners should be sent back to their country of origin’. How can adult civic understanding – for these are 24 year olds – conceivably be reduced to an enforced choice between two such banal and loaded extremes?

An earlier attempt by OECD to plug the process gap raised different problems. The indicators in their 1994 study of quality in teaching were premised on the wholly reasonable view that the quality of the education the student receives is conditioned by the kind of teaching he/she experiences, which in turn was defined as being dependent upon:

- Content knowledge
- Pedagogic skill
- Reflection
- Empathy
- Managerial competence

(OECD, 1994)

The first two have high face validity: a teacher needs to understand what he/she is teaching and to have the concomitant skill. The other three could mean anything or nothing. Yet even if we can secure agreement that these are indeed the key indicators of quality in teaching, how can they be – indeed in the case of ‘reflection’ and ‘empathy’ can they be – measured? Yet OECD press on regardless, extending their list to include:

- Commitment
- Love children
- Set an example to their children
- Manage groups effectively
- Incorporate new technology
- Master multiple models of teaching and learning
- Adjust and improvise
- Know the students
- Exchange ideas with other teachers
- Reflect

(OECD, 1994)

By this time, such a list loses all remnants of credibility. How, in educational settings, is ‘love children’ indicated and measured, qualitatively let alone quantitatively? Exactly what kind of an ‘example to their children’ are teachers expected to set? What are they expected to ‘adjust’ and in what areas of their work are they expected to ‘improvise’? How many ‘multiple’ models of teaching does it take to demonstrate ‘mastery’, and may these be any models or are only certain models allowed? Does ‘know the students’ mean demonstrate deep insight into how they think and feel or merely know their names? And so on.

The same problem is exemplified even more startlingly in DfID’s EFA ‘goals for quality primary education’, published in 2000 (the italics are mine):

- Developing *committed* and *motivated* teachers
- Defining and implementing *appropriate* curricula
- Providing *appropriate* teaching and learning materials
- Using *appropriate* languages for learning
- Promoting community participation
- Managing physical assets *effectively*
- Strengthening site-based management
- Undertaking *meaningful* assessment
- Creating a *child-friendly* environment
- Harnessing technology

(DfID, 2000)

Nearly all of these pivot on a modifier which conspicuously lacks any kind of objective meaning. Perhaps the virtue in this list is the transparency of the problem of relativism which it illustrates. It is hard to quarrel with the claim that teachers, curricula, learning materials, language, resources and so on are prerequisites for successful education, so while the DfID list does not even pretend to provide objectively or consistently applicable indicators of quality, we might feel inclined to sit down and thrash out what, within a specific cultural context such as SSA, is meant by ‘meaningful’, ‘appropriate’, ‘effective’ and ‘child-friendly’. But that hardly meets the needs of those who for whom the imperative is to identify indicators for defining and monitoring educational quality which are as low-inference as possible.

2.2 From Millennium Development Goals to the 2005 EFA Global Monitoring Report

The United Nations second Millennium Development Goal (MDG) was immensely ambitious yet conceptually minimalist:

Ensure that by 2015 children everywhere, boys and girls alike, will be able to complete a full course of primary schooling.

In the earlier Dakar Framework for Action a ‘full course of primary schooling’ had been framed by six EFA goals covering early childhood care and education, free and compulsory primary education, life skills, literacy, gender equality and, especially, quality. It thereby seemed to herald a new era in the drive to define and assess quality through indicators.

Goal 6. Improving all aspects of the quality of education, and ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills. Quality is at the heart of education, and what takes place in classrooms and other learning environments is fundamentally important to the future well-being of children, young people and adults. *A quality education is one that satisfies basic learning needs, and enriches the lives of learners and their overall experience of living.* Evidence over the past decade has shown that efforts to expand enrolment must be accompanied by attempts to enhance educational quality if children are to be attracted to school, stay there and achieve meaningful learning outcomes. (UNESCO, 2000, Goal 6 and paras 42-3, my italics)

Note that quality is once again defined in terms of outcomes rather than process. Actually, ‘quality’ in the italicised sentence is probably redundant, for if ‘education’ means anything it means satisfying basic learning needs and enriching learners’ lives. So Dakar provides a broad prospectus for education rather than an enhanced view of educational quality.

In this vein the Dakar Framework went further, positing eight conditions for ‘basic education of quality for all, regardless of gender, wealth, location, language or ethnic origin’:

- healthy, well-nourished and motivated students;
- well-trained teachers and active learning techniques;
- adequate facilities and learning materials;
- a relevant curriculum that can be taught and learned in a local language and builds upon the knowledge and experience of the teachers and learners;
- an environment that not only encourages learning but is welcoming, gender-sensitive, healthy and safe;

- a clear definition and accurate assessment of learning outcomes, including knowledge, skills, attitudes and, values;
- participatory governance and management;
- respect for and engagement with local communities and cultures.

(UNESCO, 2000, para 44)

Though this list yet again begs the question of how ‘adequate’, ‘relevant’, ‘well-trained’, ‘welcoming’ and so on are defined, there are those who see the series of UNESCO EFA Global Monitoring Reports as representing a considerable advance on the difficulties of earlier efforts such as those illustrated above. The title *The Quality Imperative* of the 2005 report (UNESCO, 2004) is encouraging, as is the range and depth of much of the subsequent discussion of the problems of defining quality.

To reconcile the many traditions of educational thought in the international arena the report posits a broad framework of five dimensions of educational quality, each subdivided:

- learner characteristics dimension [aptitude, perseverance, school readiness, prior knowledge, barriers to learning]
- contextual dimension [economic, cultural, national policy, requirements and standards, resources, infrastructure, time, expectations, etc.]
- enabling inputs dimension [teaching and learning materials, physical facilities, human resources, school governance]
- teaching and learning dimension [learning time, teaching methods, assessment/feedback/incentives, class size]
- outcomes dimension [literacy, numeracy and life skills, creative and emotional skills, values, social benefits]

(UNESCO, 2004: 36)

Actually, except for the inclusion of learners, this is very close to the OECD list with which we started – context, input, process, outcome – so we are on familiar territory. Familiar, too, is the empirically problematic nature of many of the parenthesised indicators, and yet again those in the core domain of pedagogy are thin and arbitrary. Further, the ‘teaching and learning dimension’ is presented as a sub-set of ‘enabling inputs’, thus in effect reducing the complex and often unpredictable dynamics of pedagogic process to the four ostensibly fixed and measurable ‘inputs’ listed, only two of which are within the teacher’s control. This seems to signal that process remains, for the purveyors of quality indicators at least, a no-go area.

As if to underline this, children’s creative and emotional development, which by their nature dictate an open-ended pedagogy and the possibility of divergent outcomes, are reduced to ‘creative and emotional skills’, presumably on the basis that skills are more controllable and amenable to measurement than is development. But what exactly is an ‘emotional skill’? The ability to smile, rage or weep? And through the shedding of precisely how many tears for a suffering fellow-human is the emotional ‘skill’ of empathy measured and judged satisfactory?

This, then, is another example of a recurrent tendency in the literature on education for development: making pedagogy fit the available measures rather than the measures fit the pedagogy. Pedagogy is defined as a controllable input rather than as a process whose dynamic reflects the unique circumstances of each classroom and which is therefore variable and unpredictable; and the only aspects of pedagogy which are admitted as ‘inputs’ are those

which can be measured. The whole exercise becomes impossibly reductionist, and the educational endeavour itself is as a consequence trivialised.

Moving on to the central task of measuring progress towards EFA, the 2005 report reminds us of the chosen constituents and indicators of the EFA Education Development Index (EDI):

Constituent	Indicator
• Universal Primary Education (UPE)	Net enrolment ratio in primary education
• Adult literacy	Literacy rate of those aged 15 years and over
• Quality of education	Survival rate to grade 5 of primary education
• Gender parity	Mean of the gender parity index (GPI) for the gross enrolment ratios and adult literacy rates (UNESCO, 2004: 136)

Three of the EDI indicators are proxies, and each raises questions. The difficulties of measuring gender parity are openly acknowledged but are left unresolved. UPE as a whole is measured by enrolment but by nothing that happens to the child subsequently, which is hardly satisfactory, for universal enrolment in primary education indicates that children are entering schools but not that educative experiences of good and consistent quality are being provided once they settle in. Survival rate to grade 5 might be more appropriate, but it is used for ‘quality of education’ and is therefore presumably unavailable as an indicator for UPE as well. Thus we are left with the bizarre equating of ‘quality’ with ‘survival’, and the implication in that unfortunate choice of words that education is an ordeal rather than a pleasure. ‘How good was your school?’ ‘Outstanding: I survived to grade 5.’

Meanwhile, the fit between adult literacy and its proxy is reasonably close, provided that it is attended by a reasonably convincing measure of adult literacy (historically, defining what counts as functional literacy among adults has always been a minefield (Graff, 1991; OECD, 1995a).

If the EDI is uncomfortably reductionist on quality – and each successive EFA Global Monitoring Report repeats the phrase that ‘for the time being, the EDI incorporates only the four most quantifiable EFA goals’ (see, for example, the 2007 report (UNESCO, 2006: 348)), UNESCO at least acknowledges some of the difficulties. It notes that quality is usually indicated by student learning outcome measures, and accepts that these tell us nothing about outcomes in respect of goals other than the narrow selection which are conventionally measured, nor about the cognitive value added by schooling to home and social background (or, perhaps no less important, vice versa). In fact, UNESCO rejects learning outcome measures because across the countries covered by the EDI the data are unsafe. Instead, survival rate to grade 5 is proposed as the next best proxy, and is defended as correlating with outcome measures such as performance in international tests. UNESCO also considers the claims of pupil-teacher ratio (PTR), but finds the evidence ambiguous and confirms survival rate to grade 5 as the safest option:

The fifth year of primary schooling is often taken as the threshold for acquisition of sustainable literacy. The survival rate to grade 5 also captures aspects of grade repetition, promotion policy and early dropout, and thus incorporates some comparison of the internal efficiency of education systems. (UNESCO, 2007)

However, systemic efficiency and pedagogical quality are not at all the same. Further, survival rate to grade 5 could serve as a no less convincing proxy for several other factors: parental encouragement, student ambition, even the midday meal, to which in the context of SSA we have already referred.

Yet, despite these methodological difficulties, the 2005 EFA report on quality sticks to relatively conventional indicators, culminating in detailed statistics on international ‘trends in basic or proxy indicators to measure Dakar EFA Goal 6’ [Quality of Education], using:

- school life expectancy (expected number of years of formal schooling)
- survival rate to Grade 5
- pupil/teacher ratio
- female teachers as percent of total
- trained teachers as percent of total
- public current expenditure on primary education as percent of GDP
- public current expenditure on primary education per pupil

(UNESCO, 2004: 386-93)

The best discussion of quality in the UNESCO report comes when the authors escape from the constraints of indicators and range more freely across six areas which seem to have rather greater potency in the quality debate because they are recognisably about quality rather than merely, at one stage removed, indicative of it:

- appropriate, relevant and inclusive educational aims
- relevance and breadth in curriculum content
- actual time available for learning, and its use
- effective teaching styles
- appropriate language(s) of instruction
- regular, reliable and timely assessment, both summative and formative.

(UNESCO, 2004: 146-158)

This list brings us closer to the heart of the problem of quality. It is partly about values, for defining what aims for a public system of education are ‘appropriate’ and what content is ‘relevant’ to students and society at a time of rapid change is eminently and necessarily a matter for debate. But the problem is also one of evidence, and there is no shortage of evidence, as we shall see, about what kinds of teaching are most conducive to learning. Regrettably, however, the report barely scratches the surface of this literature, and dwells instead on superficially attractive notions like ‘time on task’ (which the distinguished American pedagogical researcher Nate Gage (1978) once called ‘a psychologically empty concept’) and school effectiveness, a line of research whose ‘smart bullets’ have proved irresistible to policy-makers and administrators but whose credibility is doubted by much of the educational research community (Hamilton, 1995).

We shall return later to the gulf between the research sources preferred by policy-makers and those peer-rated by the research community itself. Having said that, I acknowledge that if not treated with undue mathematical or scientific reverence, time on task can be a useful tool for both teachers and policy-makers/administrators. It encourages teachers to think harder about the gap which often opens up between the officially-stated hours of schooling and the time children actually spend on their learning, and indeed how that time is spent, and to work to

shift the administration/teaching balance towards the latter. It encourages policy-makers and administrators to start to think about classroom process and dynamics and to understand that plant and resources are only a starting point. Yet it remains the case – hence Gage’s objection – that time ostensibly devoted to learning is not necessarily time spent on learning, and we know very well that the difference in the cognitive ground covered over a week between students taught well and those taught indifferently or badly can be immense. In that sense, time on task hints at an issue of importance for the quality debate, but does not engage with it.

Nevertheless, one striking and fundamental corrective emerges from the 2005 EFA Report, that ‘quantity and quality in education are complements rather than substitutes’. They argue this having concluded that:

The countries that are farthest from achieving [the Dakar EFA] quantitative goals 1-5 [ECCE provision, UPE enrolment and completion, life skills, adult literacy and gender parity] are also farthest from achieving goal 6 [quality as defined in the previous quotation]’. (UNESCO, 2004: 225, my parentheses)

2.3 The confusion at the heart of ‘quality’

Thus far we have encountered a number of problems with accounts of educational quality based on indicators.

- Early models concentrated on input and outcome and ignored process.
- Later models attended to educational process but in an arbitrary and selective fashion, isolating only those aspects which were deemed readily amenable to measurement, regardless of their pedagogical significance.
- Yet the very act of isolating such aspects in effect conferred validity upon them, whether or not validity was merited, so quality was reduced to quantity.
- Some models, attempting to move beyond the crudity of early indicator frameworks, leavened their instrumentalism with the language of affectivity and focused on desired attributes of teachers themselves. However, this introduced unacceptably high levels of ambiguity and inference into an exercise which was ostensibly about achieving the opposite.
- On closer examination, most nominated indicators of process, of whatever persuasion, are really input or contextual variables.
- In general, pedagogy has been made to fit the available measures rather than the other way round.
- Where direct measures are not available, proxies are used, and the proxies for process quality tend to be, again, outcomes or inputs.
- The framing of educational process quality indicators is rarely, if at all, justified by reference to research on learning and teaching.

To these I want to add one not yet mentioned but which arises from all the examples cited thus far:

- A fundamental confusion in the use of the word ‘quality’.

This confusion surfaces whenever ‘quality’ is used as an adjective rather than as a noun, which etymologically it is, though the adjectival use is increasingly common, especially in policy and marketing circles where someone is trying to sell something, whether ‘quality healthcare’ or ‘quality garden furniture’. As a noun applied to a process like education, ‘quality’ can mean either an attribute, property or characteristic, in which case it is value-neutral, or it can mean a degree of excellence, as in ‘high’ or indeed ‘low’ quality. But when used as an adjective, as in phrases like ‘quality education’ (Dakar) or ‘the quality imperative’ (the 2005 UNESCO EFA monitoring report) ‘quality’ invariably designates or implies a standard or level of quality to be desired. Thus at first glance ‘teaching quality’ could mean either the qualities or attributes of teaching – all teaching – or only those particular features which differentiate the best teaching from the mundane. However, because it is the title of a 1990s UK policy initiative we are in marketing mode and ‘quality’ means ‘high standard’ rather than mere ‘attribute’.

Since policy makers have a responsibility to set and maintain high standards of public service, they might retort that there’s nothing wrong with using ‘quality’ in this aspirational way. However, the problem with ‘the quality imperative’, ‘a quality education’, and most similar adjectival uses of ‘quality’ is that they propose indicators of the standard to be desired without pausing to consider the attributes, or qualities, which characterise all education, let alone ‘quality education’. The problem is one of confusion, or elision, of the descriptive and prescriptive senses of ‘quality’. Almost always, prescription is privileged over description. Yet unless we are clear about the basic definitional qualities of – in this case – pedagogy, we are hardly in a position to propose what distinguishes a ‘quality pedagogy’ from an ordinary one.

It is the failure to engage descriptively with the attributes of education as pedagogic processes which produces many of the problems to which we have referred, specifically:

- the arbitrary selection of the qualities (descriptive) by which quality (prescriptive) is indicated, and
- presenting as indicators of qualitative excellence (prescriptive) what in reality are no more than basic attributes (descriptive).

If I am right in asserting that the favoured indicators of educational process quality favour prescription over description and that having done so they prescribe selectively and arbitrarily, then the task of improving the quality of education involves much closer attention to description and analysis, and rather less attention to problematic sound bites like ‘child-friendly teaching’ and tautologous minima like ‘effective teaching styles’.

2.4 The NCERT Quality Monitoring Tools (QMT)

A more ambitious approach to quality indicators than any cited so far, and one closer to home, is provided by the National Council of Educational Research and Training (NCERT) in conjunction with the Government of India’s Ministry of Human Resource Development (MHRD). This approach, currently at an early stage of implementation and subject to further

modification in the light of experience, provides a comprehensive indicator-based framework for defining educational quality and sets of instruments or ‘quality monitoring tools’ (hereafter QMT) for use at five levels:

- state
- district
- block
- cluster
- school/community

(NCERT, 2006: 4)

This, immediately, is an important advance on the models illustrated above, which tend to frame process quality, which by its nature is local and highly specific, in terms of very general policy preoccupations. It is relatively rare for an indicators framework to engage with classroom process at the level of detailed offered by NCERT, and rarer still for the various levels to be so carefully delineated and addressed.

The QMT framework starts with eight ‘dimensions’ of quality:

- School infrastructural facilities
- School management and community support
- School and classroom environment
- Curriculum and teaching learning materials (TLMs)
- Teacher and teacher preparation
- Classroom practices and processes
- Opportunity time (teaching-learning time)
- Learners’ assessment, monitoring and supervision.

(NCERT, 2006: 1)

This framework incorporates the familiar elements of input, context, process and outcome, though they do not correspond precisely with the dimensions listed – so, for example, dimension (2) above incorporates both input and outcome indicators. QMT can therefore be seen as partly overlapping but mainly complementing the data from SSA’s District Information System for Education (DISE) on infrastructure, facilities, enrolment and teacher supply and qualifications (Mehta, 2006).

For each of the eight dimensions NCERT identifies between six and twelve ‘key indicators’, giving a total indicator set of 62. Some of the indicators are further subdivided. Additionally, NCERT proposes six ‘major dimensions for improving quality of elementary education’, which overlap the above but with one important addition, children’s attendance:

- children’s attendance
- community support and participation
- teacher and teacher preparation
- curriculum and TLMs
- classroom practices and processes
- learners’ assessment, monitoring and supervision.

(NCERT, 2006: 7)

After a lengthy process of development and discussion nationally and at regional and state levels, fourteen draft monitoring tools or ‘formats’ and three analytical sheets were finalised for use across the five specified levels, together with instructions on how frequently the monitoring at each level should be undertaken, when and by whom, and which of the forms should be used in each case.

I do not propose to dwell in detail on the QMTs, nor to attempt to evaluate them, though below and in the latter part of this paper I shall suggest criteria by which all such instruments might be assessed before they are brought into service. What is of interest at this point is how, for the purposes of specifying indicators, educational quality is defined and described in the vital domain of pedagogy, where other indicator frameworks and instruments are so conspicuously weak.

Pedagogy in the QMTs is distributed across six of the eight ‘dimensions’:

- School and classroom environment (physical and social environment, including relationships among teachers and children).
- Curriculum and TLMs (curriculum coverage, teaching resources such as blackboard, textbooks, libraries and other equipment).
- Teacher and teacher preparation (teacher profiles and training, teacher competence and motivation, teacher support and relationships).
- Classroom practices and processes (classroom organisation, display of materials, pupil grouping, PTR, lesson introductions, ‘teaching-learning process (pedagogy)’, use of TLMs, student initiative in teaching-learning process, assessment procedure and frequency).
- Opportunity time (numbers of days, teaching hours, teachers and classes, pupil attendance, teacher attendance).
- Learners’ assessment, monitoring and supervision (state policies for assessment, recording, reward and punishment, feedback mechanisms).

(NCERT, 2006: 2-4)

If we press still further our search for the elusive process aspects of pedagogy, we again find them scattered across the dimensions, but concentrated under ‘classroom practices and processes’, one of whose indicators carries the specific label ‘pedagogy’. Teacher-student relationships, which some would regard as no less central, appears separately, under ‘school and classroom environment.’

Pedagogy thus described (for, again an advance on the frameworks exemplified earlier, the QMT has a descriptive dimension as well as a prescriptive intent) is monitored at block and cluster levels. The Block Level Analytical Sheet does not actually specify pedagogical indicators but invites the BRC (block resource centre) personnel completing the sheet merely to specify ‘five good examples of pedagogic practice’. At the core of the entire process, as far as pedagogy is concerned, we find the classroom observation schedule (Format CLF IIa) and the form for summarising the quality of teaching in different areas of the curriculum (mother tongue, mathematics, environmental studies and English) (Format CLF IIc). Both are administered quarterly by CRC (cluster resource centre) personnel.

The cluster-level tool is brief, subjective and high-inference: ‘level of learner’s participation in classroom teaching (high/moderate/low)’, ‘competence in using child-centred approach’ (yes/no)’, ‘difficulties in the classroom (please specify)’. In contrast, the block-level

observation tool is very detailed, and mixes descriptive with evaluative indicators, for example: ‘type of classroom setting (monograde/multigrade)’, ‘teaching methods (teacher dominated/child centred)’, ‘use of TLMs (adequate/inadequate/not at all)’.

Thus in the QMT suite of indicators and tools, pedagogy is defined as a combination of the following features of teaching and learning:

- overall view of teaching (teacher dominated/child centred)
- grouping (educational basis)
- seating (physical arrangements)
- resources (textbooks, TLMs, blackboard)
- interaction (defined as lesson introductions/questions/conclusions)
- organisation (class/group/individual)
- response (difficulties encountered)
- assessment (mode/incidence).

We might quibble about the vagueness of ‘child-centred’, its familiar opposition with ‘teacher dominated’ and the evident difficulties in measuring so slippery an indicator as ‘competence in using child-centred approach’. Given that the QMTs are to be used nation-wide, and that at the level of India’s one million schools quality monitoring is done by observers using checklists which include indicators of this kind, we might also wish to know more about procedures for training the observers and ensuring inter-judge reliability so that findings across schools, districts and states can legitimately be compared. Presumably the QMT development process will address such concerns.

2.5 Criteria for assessing the adequacy of quality monitoring frameworks

This account of the NCERT QMT, as one of the fuller and more carefully researched and trialled examples available, prompts five questions:

- Is the list of general pedagogical features in QMT comprehensive in relation to a coherent and sustainable account of pedagogy?
- Are the specific indicators by which each general feature is elaborated and operationalised appropriate?
- Are the indicators likely to be interpreted with sufficient consistency by their users to allow their monitoring purpose to be properly served?
- What exactly is the conceptual and/or empirical basis of the dimensions, features and indicators identified in this, or any other quality indicators framework? How far can it be justified?
- Is the QMT procedure – involving as it does fourteen monitoring formats and three analytical sheets at five levels from state to school up to four times each year – manageable?

‘Comprehensiveness’ ‘appropriateness’, ‘consistency’ and ‘justifiability’ are criteria not far removed from those used in the more familiar discourse of student assessment. The parallel is apt, since quality monitoring, no less than student assessment, entails procedures for the

assessment of educational performance in relation to stated criteria. That being so, three basic and long-established tests can usefully be invoked from this contingent evaluation literature (see especially Harlen 2007):

- the test of validity (how well the indicators correspond to that which they are deemed to indicate);
- the test of reliability (how consistently the various users apply the indicators when monitoring the same phenomena in different settings);
- the test of impact (the intended and/or unintended consequences of the use of the instrument, especially for those – teachers and learners – to whose activities the instruments are applied).

The validity test has two strands: (i) correspondence between the indicators and the nature of pedagogy; (ii) correspondence between what is said to represent quality and how quality might actually be defined. This distinguishes the descriptive aspect of an indicator from its evaluative aspect. In turn, this difference is both holistic and specific.

That is to say, we should check overall construct validity in respect of the relationship between the account of pedagogy expressed by the proposed pedagogical indicators and what actually happens in classrooms; and we should check validity at the level of specific indicators. In the latter case, taking the example of ‘interaction’, we would need to ask whether the complexity and significance of classroom interaction is appropriately represented by focusing on – in the case of the QMTs – the type of questions which teachers ask and the gender distribution of any questions which the students themselves ask (descriptive aspect). Having done that, we would need to know whether what observers judge to be ‘good’ classroom questions are indeed questions of the kind which promote student understanding and learning (evaluative aspect). Thus, in selecting what they choose to focus upon, and in making the judgements they make, are QMT users, and QMT authors, barking up the right tree, pedagogically speaking? Are TLMs (or whatever example one chooses) the correct focus for the attention of monitoring tools, or are they are a pedagogical blind alley?

The reliability test raises no less substantial challenges, especially in the context of instruments like the QMTs which are intended to be used by large numbers of people in a vast and highly diverse system. Indeed, achieving even statewise reliability for the QMTs may well be an impossible aspiration, and it might be sensible to work instead to secure reliability at the level of day-to-day use, for example among all BRC personnel within a given district, or all CRC personnel within a given block. This, I understand from discussions at NCERT, is the preferred context for the application of the QMTs. However, given that QMT data may be aggregated to provide accounts of statewise and national qualitative trends in SSA, or, if not aggregated, then at least compared across clusters, blocks, districts and states on the assumption that they are reasonably stable both semantically and methodologically, then reliability remains a problematic aspect of the QMT.

As for impact, we must make an important distinction here too, this time between impact in application and impact in reporting. We need to know, that is to say, what effect the imminence of the various monitoring procedures has upon those to whose work they are applied, and, most obviously, the impact on teachers and students of having CRC observers in their classrooms using the two tools discussed above. How far, in expectation or reality, do

the QMTs distort what they monitor? Once we move to the reporting stage, we need to know how the information collected through the fourteen QMT formats and three analytical sheets is analysed and used, and especially what decisions about educational policy and practice might be taken as a result.

Impact cannot be considered in isolation. The use of QMT data is framed and perhaps constrained by questions of validity and reliability. The data should not be used for purposes beyond what their validity and reliability allow. No less important, especially given what we have said about the relationship between quality and equity, instruments designed to monitor and assess professional activity of any kind should not be used in a way which is unfair to those to whose work they are applied. Equity in this context is partly about the need for users to stay firmly within the boundaries of what the indicators can sustain; and partly about the attribution of responsibility for what the monitoring process discovers. Thus, if a CRC officer identifies weaknesses in the classroom practice of teacher x, who is to blame – teacher x, or the various other individuals and institutions involved in his/her recruitment, training and support (including perhaps the said CRC officer)? In this context we would do well to remind ourselves that top-down accountability procedures tend to weigh heaviest on those at the bottom of the heap, and to absolve most readily those at the top, whether or not those at the bottom are most culpable and those at the top most innocent. And though there is a strong argument for bottom-up evaluation, such procedures are almost always top-down.

2.6 Indicators of quality: who needs to know what?

In the context of our discussion about the pros and cons of quality indicators, the NCERT QMT framework serves another useful purpose. By proposing that quality be monitored at different levels and in different ways, it provokes the question ‘Who, at each level of the system, needs to know what in order that quality can be assured?’

I would suggest that this is a question which some of the authors of earlier frameworks illustrated appear not to have asked themselves. Do policy makers really need to engage with matters like learning time, the management of groups, teacher empathy, the teacher’s improvisatory skills and the quality of TLMs? Are they competent to do so?

It could of course be argued that the neglect of pedagogy in indicators of quality arises precisely because policy-makers have made the running and they feel that this level of detail is not their concern. That argument would be more convincing if the indicator frameworks were consistent in the level of detail they address, but this is not so. Many are a curious mix of the highly generalised and the very specific, which suggests not so much a deliberate decision to leave pedagogy to those closer to the classroom as a failure to address our question as put, combined with an absence of the pedagogical understanding or advice which would enable them to answer it.

Thus, to propose ‘child-centred teaching methods’ as an indicator at national level is to smother with a blanket of unexamined ideology a vital professional debate about the conditions for learning and the complexities of teaching. Further, it is then very difficult for teachers to do other than attempt to enact the nostrum, or to pretend that they are doing so. In this way, just as ‘survival rate to grade 5’ is a dubious proxy for quality at the level of policy, so ‘group work’ or ‘use of TLMs’ become no less dubious proxies for child-centredness at the level of classroom practice. Teaching which is truly child-centred is indicated not by materials or grouping procedures but at a much more fundamental level in a consistent pattern

of relationships between teacher and taught, and by a deep and sympathetic engagement with the way children think, feel and act which informs every single aspect of the teacher's work, from task preparation to interaction and assessment. Child-centredness is a pervasive attribute of teaching, not a specific teaching method.

Five requirements follow from this, four about mechanisms and one about equity:

- First, the debate about quality needs to attend closely to the question of who needs to know what at each level of the education system.
- Second and commensurately, at each level people need to ask what, within their zones of power and responsibility, they can do or provide in order to help those working at school and classroom levels to secure high and consistent standards of teaching.
- Third, if responsibilities for the quality of pedagogy are shared at the different levels, it cannot be sufficient for indicators of quality to be confined to the vagaries of input, outcome or proxy process indicators at the topmost level of the system.
- Fourth, if a multi-level approach is taken, then the indicators at each level should focus not on the school and classroom, which is what tends to happen, but on the work of those at that level itself. Otherwise concern for quality is deflected downwards. So, for example – and a crucial example – at the level above schools it is as important to define and assess quality in teacher training as in teaching.
- Fifth, although pedagogy and pedagogic quality are manifested in the decisions and interactions of teachers and learners, the very fact that others at different levels are interested in it signals that quality depends on much more than the teacher alone. If responsibility is shared, culpability should be shared too.

This discussion adds a further test to those of validity, reliability and impact in relation to quality indicators: appositeness to the level to which the indicators refer.

3. Sources for Accounts of Pedagogical Quality

I identified five criteria by which the NCERT QMTs, and by extension all such frameworks, might be assessed – comprehensiveness, appropriateness, consistency, manageability and conceptual/empirical justifiability – and related the first four to the more familiar methodological tests of validity, reliability and impact. To the latter I appended appositeness in relation to the responsibilities of those at the respective levels of the system. We must now address the no less fundamental question of where the categories and items used in indicator frameworks and quality monitoring instruments actually come from, and how far they can be defended conceptually and empirically. This is all the more important because some such frameworks are notably short on explanation and justification.

Among the many possible sources, four are pre-eminent and will be considered in turn:

- national educational policy and its cultural context
- national research on pedagogy
- the international quality indicators literature
- international research on pedagogy.

3.1 National educational policy and its cultural context

Given that educational policy responds to what are identified as systemic and/or local visions, needs and problems, it is inevitable that national accounts of educational quality will reflect these. Versions of quality which are grounded in national statutory frameworks of educational aims and curriculum content will reflect agreed national priorities, while those which respond to identified problems will carry a remedial weighting as the policies attempt to correct adverse historical trends, for example (in India) teacher absenteeism, the dominance of rote learning or the uneven availability of textbooks and TLMs. In terms of what is needed in order to meet preferred goals and secure desired educational improvements, such emphases are proper and necessary.

There is a risk that the remedial approach will distort the overall quality profile. For example, there is some evidence, from the successive joint review missions (JRM)s mounted by the Government of India and its international partners, that TLMs may have become too exclusive a focus of monitoring procedures in DPEP and SSA at some levels of the system, to the detriment of other important aspects of teaching. Similarly, in England the OFSTED inspectors' concern to sharpen the pace in teaching has led to the pursuit of pace at all costs, regardless of the fact that pace without attention to students' understanding leaves all but the fastest learners stranded.

The second sense in which the national necessarily impacts on accounts and indicators of pedagogical quality is through force of circumstance. Thus, for example, in urban settings primary schools – universally, not just in India – tend to be larger and their teaching is typically monograde. In rural settings primary schools are much smaller, with no more than a handful of children in each year or grade, so teaching, perforce, is more likely to be multigrade. The indicators of pedagogical quality in such settings must therefore focus explicitly on strategies and practices which are able to promote learning where there is a wide

range of student age, need and outcome, and where because of this the teacher cannot treat the class as the same kind of educational unit as in monograde settings.

Thus far, we have a case for national accounts of quality to have a distinctively national and indeed local slant, and in the context of SSA this provides justification for the efforts of organisations like NCERT and the many state and district initiatives in pursuit of quality.

At a more subliminal yet profound level, national considerations impact on definitions of pedagogical quality in the realm of culture and values. There is now a substantial literature which contrasts supposedly 'Asian' and 'Western' models of teaching on the basis of their differing accounts of the relative importance of ability and effort, or differing their varying commitments to individualism/egocentrism and holism/sociocentrism (Shweder, 1991; Stigler and Hiebert, 1999). Though some believe that to corral all the countries and cultures of Asia, let alone all of those of Europe, North America and Australasia, into the opposing camps of 'Asian' and 'Western' is a generalisation much too far (Alexander 2006a, 2008), it is undoubtedly the case that pedagogy is shaped by culture. My preferred basis for analysis is national systems, and even then one must pay as close attention to intra-national as to international patterns, similarities and differences – not least in countries as culturally diverse as India. Further, history ensures that national systems are cultural hybrids, bearing the stamp not just of the most recent policy, but of the sedimented influences to which it has been subject over generations and centuries.

Thus, although an offspring of revolution, French public education retains features which recall its pre-revolutionary and ecclesiastical origins (Sharpe, 1997), and the conjunction of institutional secularism and individual liberty is not without its tensions. The more obvious Soviet trappings of Russian education have been shed, but the abiding commitment to *vospitanie*, and the emphasis in schools and classrooms on collective action and responsibility allied to unambiguous teacher authority, not to mention the methods of teaching, show all the more clearly that the continuities here are Tsarist as well as Soviet. The continuities in India reach back even further, and we can identify at least four traditions – two of them indigenous (Brahmanic and post-Independence) and two imposed (colonialist and missionary) combining to shape contemporary classroom practice in that vast and complex country (Kumar, 1991; Alexander, 2001). In England, the twin legacies of elementary school minimalism and progressive idealism offset government attempts at root-and-branch modernisation. The one still shapes school structures and curriculum priorities (and government is as much in its thrall as are teachers), while the other continues to influence professional consciousness.

This is not the place for a detailed discussion of the relationship between pedagogy and culture. The general point to be made by drawing attention to this issue is that culture is so pervasive a shaper of education and educational realities that it cannot possibly be ignored. It gives rise to varying and often competing accounts of knowledge, of learning and of the relationship between teacher and taught, in other words the very stuff of pedagogy; and beyond these it reflects differing versions of human relations and the proper basis for social cohesion. Out of such primordial values come contrasting valuations placed on individual, communal and collective action in society, and on individualised learning, group work and whole class activity as the proper foundations of effective teaching.

In other words, pedagogy is not just a matter of disembodied technique. It reflects and manifests values. In turn these are not merely the personal predilections of individual teachers, but the shared and/or disputed values of the wider culture.

All this is a long way indeed from the simplistic array of ‘inputs’, ‘outcomes’ and indeed ‘processes’ of which the typical portmanteau of pedagogical quality indicators is constituted, to the extent that it must reinforce what I trust is a growing sense through this paper’s discussion that to define pedagogical quality through conventional indicators is probably to miss the larger part of what pedagogy actually entails.

3.2 National research on pedagogy

As with using national policy as the starting point for defining and indicating quality, the advantage of grounding an approach to quality in national research is that it is likely to be alert to local circumstance, need and culture. Where educational governance and educational research are closely aligned – as exemplified in India’s concurrent system by the relationship between MHRD and NCERT, and between state departments of education and the state councils of educational research and training (SCERTs), research may serve as an explicit instrument of policy.

This situation, too, affords risks as well as benefits. Among the benefits, assuming an appropriate level of research capacity and a political climate in which it is possible for researchers to ‘speak truth to power’ without the relationship being compromised by patronage for those who conform and reprisal for those who do not, the alignment of policy and research can ensure that policy is firmly grounded in evidence.

If that climate does not exist, then the risk is that researchers become the hired hands of policy-makers, feeding them with ‘evidence’ to legitimate policies which may not be able to withstand proper evidential scrutiny; or that researchers will merely tell policy-makers what the latter expect to hear; or that what policy-makers do not wish to hear they will ignore or suppress. Either way, quality will become the second casualty (truth will be the first). The other risk is that the very emphases which mark a proper response to national circumstance and need may represent distortions in relation to the wider spectrum of educational and pedagogical possibilities, or that together research and policy will reinforce weakness as well as – and even at the expense of – strength.

Both risks are sufficiently real and serious for it to be sensible to combine international and national research sources in order to achieve breadth and balance in quality frameworks, alongside local relevance.

That noted, there is a growing indigenous research literature on quality issues relating to SSA, the earlier experience of DPEP, and generic matters like multigrade teaching. However, as yet this literature tends to be more descriptive or celebratory than evaluative, and to concentrate on innovations which are in line with established DPEP/SSA thinking (e.g. Chand and Amin-Choudhury, 2006). While such material serves the important purpose of expanding local perceptions of what is possible, there is a danger that for those who are habituated to comply rather than question it will be assumed to have normative intent.

There is also useful material from the non-governmental sector, including both major campaigning and enabling organisations like Pratham² (which is now operating in 21 states)

² Pratham is a major Indian NGO with missions in education and literacy.

and those such as Rishi Valley³ which have a more specific pedagogical focus and whose approach has been occasionally imported or replicated in DPEP and SSA. However, the material so far available from or on bodies such as these does not include research on process quality. Pratham's valuable Annual Status of Education Reports series (e.g. Pratham, 2007) uses input and outcome indicators which partly overlap and partly complement those available from official sources, while the Rishi Valley literature concentrates on spreading the message about its School in a Box multigrade teaching and learning materials (Rao and Rao, undated). Of the latter, despite their high profile, there appears as yet to be no formal evaluation.

3.3 The international quality indicators literature

In the first instance this takes us back to the international quality indicators literature, many of whose problems have already been identified and do not need to be repeated. Yet, all things being equal, this literature should be an ideal starting point. A major series like UNESCO's EFA Monitoring Reports includes evidence from most of the world's education systems and is informed by principles which are not only driven by commitment to justice, equity and human advancement – and which therefore rise above the transient preoccupations of national governments and ministers – but to which all participating nations claim to subscribe.

Further, because these indicators cover both inputs and outcomes it is possible to assess which of them are relatively 'safe' (in the parlance of the 2005 EFA report) if not valid; and because they include demographic and financial data it is also possible to assess which interventions, on cost-effect grounds, are 'promising avenues' and 'blind alleys' (Lockheed and Verspoor, 1991: 87).

The dangers of the latter use of indicators must be immediately underlined, obvious though they are. Not all 'interventions' can be costed with anything approaching the precision required, and the approach can result in spectacular errors. For example, Lockheed and Verspoor rate midday meals, as an intervention to improve learning achievement, a 'blind alley' not worth contemplating (ibid). Teachers, parents, children and policy-makers in India know otherwise.

But apart from the problems of the international indicators literature which this paper has already rehearsed, perhaps the most serious is the assumption – a necessary assumption given the extent of statistical analysis to which they are subject – that each of the indicators is stable and constant across (in the case of the UNESCO reports) all 203 of the countries to whose education systems they are applied.

Two examples suffice to illustrate the frailty of this position. The UNESCO EFA Monitoring Reports for 2003 and 2004 provide data on youth and adult literacy and illiteracy – arguably an outcome indicator of supreme importance (UNESCO, 2003; UNESCO, 2004). In the statistical tables the literacy/illiteracy rates are presented as percentages, country by country. They are preceded by a discussion of the difficulties of defining literacy which dismisses the older 'self-declaration' method in favour of direct assessment. But since in many of the

³ Rishi Valley, near Madanapalle in Andhra Pradesh, is the home of the Rishi Valley Education Centre. It is run by the Krishnamurti Foundation and comprises schools, environmental and health programmes organised according to Krishnamurti's principles. The Rishi Valley 'school in a box' materials for multigrade settings (Rao and Rao, undated) are widely used in government schools in India as well as by NGOs, and indeed have been adapted for use in some African settings.

developing countries literacy is either not defined or is based on self-declaration rather than objective assessment, and since even when objective measures are used they vary considerably between countries, the literacy data must not only be used, as the 2005 report urges, 'with caution'; such data are almost certainly too suspect to give other than the roughest of rough and ready assessments.

The other example is the EFA reports' use of 'trained teachers as percent of total' as an indicator of the educational quality goal. Training is defined by International Standard Classification of Education (ISCED) level of education achieved, but training itself is not defined. Yet national systems of teacher training vary so markedly in entry requirements, length, content, qualification, impact and – however defined and indicated – quality, that this indicator, too, is extremely suspect. And so one could go on.

But of course my most serious reservation about the international indicators literature, including that on educational quality, is that it says so little about pedagogy and what little it says has such limited value.

3.4 International research on pedagogy

To commend the application of international pedagogical research to the challenges of defining and monitoring educational quality in specific countries is to risk, in our post-colonialist, post-orientalist times, suspicion of cultural hegemony; the more so now that globalisation means, for advocates and opponents alike, westernisation; and now that the growing international dominance of the English language makes Anglo-American research so much more readily accessible than that from non-Anglophone countries and cultures. Accessibility, of course, is no guarantee of quality.

In fact, the EFA literature sometimes displays a certain insensitivity not just to the risks of cultural colonialism but to culture generally. This emerges not so much in the discourse which nowadays is generally aware of cultural nuance and sensitivity, as in the way that data are handled. Thus, country statistics are assembled, and indicators are listed, aggregated, disaggregated and correlated, all on the apparent assumption that they are universally valid and meaningful and that they are not weighted for or against particular cultural contexts. Identical outcome indicators are applied across all countries, regardless of differences in national educational goals, the scope and balance of national curricula or the very different social and economic circumstances which national school systems seek to address. Indicators may be rejected, but on the grounds that they are statistically unsafe rather than because they are culturally inappropriate, and the latter test is rarely applied.

If, as I have asserted, pedagogy must be understood as a cultural artefact which manifests the sedimented values and habits of a nation's history, should we even enter this territory? I believe that there are four reasons why we should, though cautiously:

- Pedagogy is so palpably the missing ingredient in the international debate about educational quality, and it is so obviously vital to student retention and progress and to learning outcomes, that we have no alternative but to find ways of remedying the deficiency.

- Anyone who criticises existing accounts of pedagogy in the international literature on quality as crude or ill-conceived – as I have – has an obligation to come up with something better.
- Comparative research shows that in practice pedagogy combines the culturally or nationally unique with the universal. So, for example, the ‘basics’ of literacy and numeracy are almost everywhere prioritised (Benavot et al, 1991), and both rote and the closed initiation-response-feedback (IRF) structure of recitation, regardless of local pedagogic tradition, have widespread currency as the default mode of teaching in public education systems (Alexander, 2001 and 2006b). However, the apparent universality of literacy disguises the fact that ‘literacy’ can carry very different meanings, not just – as we have seen – at the level of outcomes, but also in its very conception. For example, in the continental European tradition oracy and literacy are contingent and inseparable, but in the Anglo-Saxon tradition they are handled separately and literacy is defined in relation to the written word only. There is virtue, therefore, in disentangling and carefully differentiating the local and the international in pedagogy.
- Culturally, humanity is characterised by variety and difference. But humanity is also a biological species whose individuals and groups have much in common. If pedagogy is defined as the act of teaching together with its attendant discourses, theories and beliefs about human development, learning, curriculum and so on, the observable practice of pedagogy may vary but much of its psychological and developmental grounding could well be constant. That, at least, is an assumption which deserves to be tested.

3.5 Problems with the pedagogical research cited

The quality indicators literature is far from even-handed, let alone comprehensive, in its use of published pedagogical research. It attends very little to research on learning, and the cited research on teaching comes almost exclusively from just one tradition, that of school effectiveness research. The 2005 UNESCO EFA monitoring report on quality is the most prominent recent example of this empirical myopia (see ‘The importance of good quality: what research tells us’, in UNESCO, 2004: 60-78).

The twofold attractiveness of school effectiveness research is that it maps readily onto the dominant input-outcome paradigm of quality indicators, and it conveniently translates quality into quantity. Its near-exclusive hold on the international quality indicators community, such as it is, demands comment. In the particular context of SSA, it should also be noted that school effectiveness research has proved highly influential at NCERT, India’s apex educational research institution (NCERT, 1996, 1997, 2003), and this influence can be discerned in the Quality Monitoring Tools (QMT) discussed above.

School effectiveness research is an offshoot of Anglo-American process-product research, but stands well apart from the current research mainstream, and indeed from the critique which process-product research has generated over the past five decades. The first wave of school effectiveness research, during the late 1980s and early 1990s, was largely non-empirical. It consisted of territory demarcation and the collating of those few empirical studies which, as defined by school effectiveness researchers themselves, were deemed relevant to the endeavour (Reynolds et al, 1994). ‘Effectiveness’ was defined very simply, as a statistical calculation of the gain in output over input, and – another boon for the EFA quality indicators community – the calculation included measures of equity as well as quality:

We define effectiveness in two dimensions [graph shows axes of input and output] The 'quality' dimension is modelled as the average score of each school on output (corrected for input) and is represented by the intercept (each school has a different intercept). The 'equity' dimension encompasses the compensatory power or selective quality of schools. Some schools can better compensate for input characteristics than others. This dimension is represented by the slopes of the within school regression of input on output. (Creemers, 1994: 10-11)

Those studies which conformed to this statistical paradigm were extensively reviewed in the publications of the school effectiveness group which established itself in the UK and USA and then networked across several other countries. In a parallel venture, the English schools inspectorate, OFSTED, commissioned an extrapolation of the 'key characteristics of effective schools' from school effectiveness research from a group at the University of London's Institute of Education. This came up with eleven factors in the effective school:

- Professional leadership (of head)
- Shared vision and goals
- A learning environment
- Concentration on teaching and learning
- Purposeful teaching
- High expectations
- Positive reinforcement
- Monitoring progress
- Pupil rights and responsibilities
- Home-school partnership
- A learning organisation

(Sammons et al, 1995)

Each of these was subdivided. Thus 'professional leadership' included 'firm and purposeful', 'a participative approach' and 'the leading professional', while 'purposeful teaching' was explicated as 'efficient organisation', 'clarity of purpose', 'structured lessons' and 'adaptive practice'.

Hamilton's critique of this exercise sees it as predicated on a pathological view of schools as sick institutions in need of clear policy prescriptions presented as 'magic bullets or smart missiles'; he faults the methodology of aggregating findings from studies conducted by different methods, at different times and in different countries; and rejects 'the suppositions and conclusions of such research ... as an ethnocentric pseudo-science that serves merely to mystify anxious administrators and marginalise classroom practitioners.' (Hamilton, 1995).

In my view, the aggregation is not only indefensible (it yields, for example, a model of an all-powerful school head which, whatever its currency in the UK or USA where most of the reviewed studies were undertaken, makes no sense in those countries, like France or India, where primary school heads have limited jurisdiction); it is also reductionist and banal. Not one of the listed factors from the OFSTED study takes us beyond what the commonsense of a layperson would have predicted.

The most prominent school effectiveness group then put together an International School Effectiveness Research Project (ISERP) (Reynolds et al, 2002) whose statistical paradigm

was compromised from the outset by sampling problems – between five and twelve schools, serendipitously identified, were deemed representative of each of the nine countries involved. On the other hand, the project at least confirmed some of the factors in effective teaching which emerged from other classroom research, notably the importance of organising classroom time and space as economically as possible, maximising children's opportunity to learn, and generating challenging and focused pupil-teacher interaction. Meanwhile, I would add the following further reservations to Hamilton's:

- First, and perhaps most important given its international and comparative claims and my earlier arguments on this score, school effectiveness research does not deal more than cursorily with culture. Culture, indeed, is treated as no more than another variable, having significance no greater than, say, time on task or opportunity to learn.
- Second, because it focuses exclusively on behaviours, school effectiveness research is technically unable to engage with the purposes, meanings and messages which elevate pedagogy from mindless technique to considered educational act. Teaching is therefore presented as value-neutral, content-free and entirely devoid of the dilemmas of ideal and circumstance which confront real teachers daily.
- Third, there is a degree of arbitrariness in the variables which the paradigm includes and excludes, as can be seen in Creemers' frequently-cited model (Creemers, 1994 and 1997). In fact, most are derived from literature searches, so the model – being merely a representation of what other have chosen to write about or investigate – is by no means as comprehensive as it appears or claims.
- Fourth, there are obvious technical questions to be addressed in ISERP and related studies: sampling, the use of questionnaires rather than observation as the basis for identifying effectiveness factors, the highly mechanistic approach to classroom observation.
- Fifth, there is a spurious absolutism to the terminology of school effectiveness – 'success', 'failure', 'improvement', and of course 'effective' itself – which conceals the technical deficiencies of the research and implies a degree of homogeneity in schools, classrooms and lessons which cannot be sustained empirically.
- Sixth, school effectiveness research is unacceptably exclusive and tacitly rejects the principle of cumulation which is vital, in any discipline, to the advancement of knowledge. Its methodological discussions and bibliographies make little or no reference to the much longer and more substantial tradition of pedagogic research which has attempted to address the same question – what teaching makes the most difference – but by different means.

Despite these problems, which are fundamental, this particular strand of process-product research has wielded considerable influence in policy circles and, as noted, in the international quality indicators literature, where OFSTED's 'eleven factors for effective schools' and the 48-factor 'comprehensive model of educational effectiveness' constructed by Creemers and his colleagues have become standard points of reference in the EFA context.

The OFSTED list extracts from its review of mainly American and British research classroom-level factors such as 'orderly atmosphere', 'attractive working environment',

‘maximisation of learning time’, ‘academic emphasis’, ‘focus on achievement’, ‘efficient organisation’, ‘clarity of purpose’, ‘structured lessons’, ‘adaptive practice’, ‘high expectations’, ‘intellectual challenge’, ‘clear and fair discipline’, ‘monitoring pupil performance’ and ‘feedback’. Creemers (1997) provides a fuller list upon which, in an attempt to create a coherent model of educational effectiveness, he superimposes four ‘levels’ – context, school, classroom and student – and three recurrent or permeative principles: ‘quality’, ‘time’ and ‘opportunity’. Creemers gives most attention to his principle of quality at the classroom level, identifying 25 ‘quality of instruction’ factors under the headings of ‘curriculum’, ‘grouping procedures’ and ‘teacher behaviour’, and setting these alongside ‘time for learning’ and ‘opportunity to learn’.

Long ago, Philip Jackson (1962) warned that some correlational effectiveness studies are ‘so low in intellectual food value that it is almost embarrassing to discuss them’, and there are obvious dangers in aggregating the findings of research projects which have been conducted in different countries and school contexts, using not necessarily compatible measures and yielding correlations of unspecified but no doubt varying orders of magnitude. Even OECD, which in its pursuit of indicators of educational and economic success is generally eager to quantify every aspect of education it can identify, warns policy-makers against excessive aggregation (OECD, 1995b).

If effectiveness studies can be sustained in the face of this kind of criticism then we should take seriously their claim to provide a first-order list of those teaching elements which most merit attention in empirical study, professional training and strategies for school improvement. A research-based list of this sort would include the following: how curriculum content is structured and presented; how pupils are grouped and the extent to which such grouping allows for cooperative learning; how efficiently classroom events are managed; the way time is used so as to maximise both pupils’ opportunity to learn and the time they spend on task; the range and clarity of the teacher’s lesson goals; the degree to which the teacher maintains a consistent emphasis on basic skills, cognitive learning and learning transfer; the quality of direct instruction procedures such as questioning and explaining; the timing and precision of feedback to pupils on work undertaken; and the transition from evaluation and feedback to the immediate rectification of pupils’ mistakes and misunderstandings (Creemers, 1997).

Many of these factors have common-sense validity. However, the elements of teaching which feature in frameworks such as those cited above are rather more arbitrary than they may seem. For the items in Creemers’ framework are an undifferentiated mixture of what he sees as desirable in theory and what process-product research has demonstrated empirically. While the latter, or some of them, can claim the status of factors in educational effectiveness, the theoretical options are no more or less convincing and complete than whatever theory of teaching has spawned them. Moreover, the factors which can lay claim to inclusion in the model on empirical rather than theoretical grounds may be there for no reason other than that they are technically amenable to statistical treatment. How else can one explain, in a World Bank document grandly and synoptically entitled ‘What do we know about school effectiveness and school improvement?’ the reduction of ‘teaching/learning process’ to just four factors: ‘high learning time’, ‘variety in teaching strategies’, ‘frequent homework’ and ‘frequent student assessment and feedback’ (Saunders, 2000)? Or perhaps the number of factors is really two, since time is not a process and homework, frequent or otherwise, does not take place in the classroom.

The ‘comprehensive model’ which informs school effectiveness research, then, is neither comprehensive nor a model. It is more a list whose purpose teeters between description and prescription when it needs to provide, unambiguously, one or the other. While, of course, no model can include everything that needs to be contained within a meaningful account of effective education, this one ends up including surprisingly little. It provides no indication of how schooling and teaching actually work, with the result that the path from effectiveness factor to educational action remains obscure. Perhaps given the inevitable limits to any attempts to represent the complexities of practice in a theoretical framework, the word ‘comprehensive’ should be used more circumspectly.

In fact, if we return to the American roots of this research we find that syntheses such as those provided by Dunkin and Biddle over three decades ago are rather clearer on this score. They propose a model in which presage variables (‘teacher formative experiences’, ‘teacher training experiences’, ‘teacher properties’) and context variables (‘pupil formative experiences’, ‘pupil properties’, ‘school and community contexts’, ‘classroom contexts’) interact with classroom variables (‘teacher classroom behaviour’, ‘pupil classroom behaviour’, ‘observable changes in pupil behaviour’) to yield product variables (‘immediate pupil growth’, ‘long-term pupil effects’). However, they make it clear that theirs is a model for the study of teaching, not a model of teaching itself (Dunkin and Biddle, 1974).

Process-product research, of which school effectiveness research is an normative offshoot, offers a *prima facie* way forward in that it has a rationale for the selection which it makes among the many possibilities, namely that the chosen elements correlate with gains in pupil learning. But as a path to a broader understanding of the character and efficacy of teaching as a particular kind of human activity it takes us only so far, as do all frameworks which merely differentiate categories of variables, whether presage, context, process or outcome, but leave unexamined the question of how these are reconstituted as real-time teaching.

3.6 Pedagogical research and multigrade reality

These considerable empirical and conceptual difficulties apart, the enthusiasm for school effectiveness research in the contexts of DPEP and SSA becomes downright baffling when one considers that most of it has been undertaken in circumstances utterly different from those which obtain in India. The countries which featured in the ISERP research programme referred to above (Reynolds et al, 2002) were Australia, Britain, Canada, China (Hong Kong), Ireland, the Netherlands, Norway, Taiwan and the United States. The dominant setting, therefore, was one of high GDP, high parental literacy and – especially significant in the context of pedagogy – monograde teaching in large urban schools. As the most recent DISE data show, the contrasting Indian reality, for approximately 78 per cent of primary schools, is three or fewer teachers to the five primary grade levels, making multigrade teaching inevitable. Indeed, 58 per cent of primary schools have fewer than 100 students (Mehta, 2006). In terms of the model of teaching which is outlined in the next section, this impacts directly and decisively on four key aspects of teacher decision-making:

- curriculum (common vs. differentiated);
- student organisation (e.g. the balance of whole class teaching, collective group work and collaborative group work; the basis for student grouping);
- task (e.g. common tasks vs. tasks differentiated by age/grade and/or perceived ability; the use of task-related materials which require regular teacher intervention and feedback vs. those which are structured for self-study);

- interaction (e.g. the balance of teacher-led and peer-led interaction; the character of the teacher talk and pupil talk in these two contexts).

This can be compared with the multigrade typology of Little et al (2006) which focuses on curriculum (common/differentiated), grouping and materials. The most striking dissonance between the contexts of school effectiveness research and multigrade teaching, then, is that the default organisational and interactive mode for the former is whole class teaching. In contrast, research on multigrade teaching exposes the limitations of whole class teaching and explores the learning potential for students of working in groups. In the one context, then, the progress of learning centres on interaction between student and teacher; in the other, on students' engagement with materials and on their interaction with peers.

Multigrade teaching has generated a significant literature. Much of it is of the 'how to do it' variety, with handbooks and materials, often generated by NGOs. But there are also studies which attempt to assess the impact and outcomes of the various versions of multigrade teaching. Some of these (for example Gupta et al, 1996) are indigenous, while there is a much larger body of international material (e.g. Little, 2006). Indeed the international reach of this literature is considerably more extensive than that of school effectiveness research, and it includes medium and low GDP contexts as well as high. DfID itself has supported developments in this field, including the work of the group at the Institute of Education, University of London headed by Little and Pridmore.

It is also important to note that in some quarters (for example, among defenders of small rural schools in Britain) multigrade teaching is viewed as educationally and socially desirable rather than as an awkward necessity born of geography and/or scarce resources. In other words, if a multigrade pedagogy can be developed which bears comparison in terms of process rigour and student learning outcomes with the best of monograde teaching – a big 'if' perhaps – then this might make squaring the quality circle in EFA a more viable proposition. Nor should the motivational and communal potential of mixed-age teaching be underplayed. Such are the issues which, for example, the Rishi Valley 'School in a Box' system claims successfully to have addressed (Rishi Valley, 2000).

Yet the situation is far from straightforward. On the one hand, the Indian national curriculum framework and teacher education programmes are premised on monograde staffing and teaching; on the other, multigrade classes persist in the majority of the country's primary schools and are likely to do so for a long time yet, because so many of the schools are too small to make monograde staffing a realistic economic proposition. Most critically, multigrade aligns with geography and poverty, and multigrade teachers therefore contend not only with a generic pedagogical challenge for which they have not been trained – that of teaching a complete curriculum to students of widely-varying ages at the same time – but also with the contexts of professional isolation and rural poverty which make teaching difficult in any circumstances.

In Britain, in contrast, campaigns for the preservation of small rural primary schools with multiage classes are generally led by affluent parents in the shire counties, and what schools cannot provide, such parents will and do. In any case, in Britain the unit cost of education in rural schools is usually higher than urban. A romanticised approach to the multigrade question must at all costs be avoided, therefore.

3.7 Alternative research sources, definitions and frameworks

Once one escapes from the confines of school effectiveness research and registers the countervailing challenge and developing response to multigrade teaching, the alternative literature on pedagogy is truly vast. On what basis does one select from this literature to take forward the debate about pedagogy and pedagogical quality?

What should now be clear is that we can go no further with this matter without a definition of pedagogy and a delineation of its scope. Otherwise our selection from the literature will be no more than random.

‘Pedagogy’ carries definitions ranging from the narrow – as in the Anglo-American tradition where it tends to be equated with the techniques of teaching – to the broad and indeed comprehensive, as in that continental European tradition whose roots can be traced back to the *Didactica Magna* of Jan Kamenski (Comenius). Because teaching is, or ought to be, a purposeful activity whose decisions are, or ought to be, grounded in professional knowledge, to equate pedagogy with the observable acts of teaching alone is unacceptably restrictive, and indeed positively encourages the view that classroom decisions are never more than instinctive or pragmatic. This gives credence to the view that the quality indicators which appear in the OECD, UNESCO and other examples cited are neutral, stable and therefore amenable to measurement. They are not.

My preferred definition, therefore, is broader:

Pedagogy is the observable act of teaching together with its attendant discourse of educational theories, values, evidence and justifications. It is what one needs to know, and the skills one needs to command, in order to make and justify the many different kinds of decisions of which teaching is constituted.⁴

This definition requires two subsidiary and complementary frameworks, one dealing with the ‘observable act’ of teaching, and the other with the ‘knowledge, values, beliefs and justifications’ which inform it.

3.7.1 Pedagogy as ideas

Let us consider, first, the ideas which inform and justify the act of teaching. These can be grouped into three domains, as shown in Figure 1. Here we see that pedagogy has at its core ideas about learners, learning and teaching, and these are shaped and modified by context, policy and culture. Where the first domain enables teaching and the second formalises and legitimates it by reference to policy and infrastructure, the third domain locates it – and children themselves – in time, place and the social world, and anchors it firmly to the questions of human identity and social purpose without which teaching makes little sense. Such ideas mark the transition from teaching to education. That is why the omission of culture and ideas in school effectiveness research is so demeaning of what pedagogy actually entails.

⁴ See Alexander (2008) for a detailed exploration of pedagogy in its various dimensions – as concept, cultural artefact, policy, practice and discourse – and of the benefits to the latter of historical and comparative analysis.

Figure 1: Pedagogy as ideas (theories, values, evidence and justifications)

Classroom level: ideas which enable teaching

- *Students* characteristics, development, motivation, needs, differences.
- *Learning* nature, facilitation, achievement and assessment.
- *Teaching* nature, scope, planning, execution and evaluation.
- *Curriculum* ways of knowing, doing, creating, investigating and making sense.

System / policy level: ideas which formalise and legitimate teaching

- *School* e.g. infrastructure, staffing, training.
- *Curriculum* e.g. aims, content
- *Assessment* e.g. formal tests, qualifications, entry requirements
- *Other policies* e.g. teacher recruitment and training, equity and inclusion

Cultural / societal level: ideas which locate teaching

- *Community* the familial and local attitudes, expectations and mores which shape learners' outlooks
- *Culture* the collective ideas, values, customs and relationships which shape a society's view of itself, of the world and of education
- *Self* what it is to be a person; how identity is acquired.

Based on Alexander (2004: 11-12)

3.7.2 Pedagogy as practice

Let us move to the other part of the definition, pedagogy as the observable practice of teaching. It, too, can be conceptually elaborated in several different ways. In my own comparative analysis of international classroom data, for which I needed a descriptive model which was as inclusive as such models ever can be yet able to frame analysis of data from five very different national education systems, I started with this irreducible proposition:

Teaching, in any setting, is the act of using method x to enable students to learn y.

In so skeletal a form the proposition is difficult to contest. If this is so we can extract from it two no less basic questions to steer empirical enquiry:

- What are students expected to learn? (from '...to enable students to learn y')
- What method does the teacher use to ensure that they do so? (from '... the act of using method x')

3.7.3 Act, form and frame

'Method' needs to be unpacked if it is to be useful as an analytical category which is able to cross the boundaries of space and time. Any teaching method combines tasks, activities, interactions and judgements. Their function is represented by four further questions:

- In a given teaching session or unit what learning tasks do students encounter?
- What activities do they undertake in order to address these learning tasks?
- Through what interactions does the teacher present, organise and sustain the learning tasks and activities?
- By what means, and on the basis of what criteria, does the teacher reach judgements about the nature and level of the tasks and activities which each student shall undertake (differentiation), and the kinds of learning which pupils achieve (assessment)?

Task, activity, interaction and assessment are the building-blocks of teaching, the constituents of teaching as act. However, as they stand they lack the wherewithal for coherence and meaning. To our first proposition, therefore, we must add a second, and this unpacks ‘in any setting’, the other question-begging phrase in our definition:

Teaching has structure and form; it is situated in, and governed by, space, time and patterns of pupil organisation; and it is undertaken for a purpose.

Structure and form in teaching are most clearly and distinctively manifested in the lesson. Lessons and their constituent teaching acts are framed and governed by time, by space (the way the classroom is disposed, organised and resourced) and by the chosen forms of pupil organisation (whole class, small group or individual).

But teaching is framed conceptually and ethically, as well as temporally and spatially. A lesson is part of a larger curriculum which may include both established subjects and domains of understanding which are not subject-specific. Curriculum embodies purposes and values, and reflects assumptions about what knowledge and understanding are of most worth to the individual and to society. This is part of the force of ‘teaching..... is undertaken for a purpose’.

There is one more element to put in place. Teaching in classrooms is not a series of one-off encounters. Teachers develop procedures for regulating the complex dynamics of pupil-pupil relationships, the equivalent of law, custom, convention and public morality in civil society. Further, teachers and teaching convey messages and values which may reach well beyond those of the particular learning tasks which give a lesson its formal focus. This element we can define as routine, rule and ritual.

The complete framework is shown in Figure 2. The components of the act of teaching (task, activity, interaction and judgement) are framed by classroom organisation (space, student organisation, time and curriculum and by classroom routines, rules and rituals). They are given form by the lesson or teaching session.

Figure 2: Pedagogy as practice		
Frame	Form	Act
Space		Task
Student organisation		Activity
Time	Lesson	
Curriculum		Interaction
Routine, rule and ritual		Judgement

Source: Alexander (2001: 325)

An ingenious modeller would construct a composite from Figures 1 and 2, showing not just how ideas inform practice (which they do sometimes but not always, and even then in unpredictable ways, though that's another story); but also how the practice in turn shapes the ideas as accumulated craft knowledge. Though idealists talk of the need to 'relate theory to practice' others who are longer in the tooth know that matters can never be that simple, and certainly not as tidily linear. This level of representation is beyond me, so I make the point verbally: the two frameworks have an intimate, necessary but highly complex relationship.

In the context of this paper's discussion, our two-stage elaboration of the notion of pedagogy serves two purposes. First, it provides pointers to sources of pedagogical research other than that on school effectiveness. Second, out of this model can be constructed a far more extensive and sustainable framework of quality indicators than most of those currently in use.

This is not the place to venture a comprehensive review of the vast alternative field of research intimated by the model. But using its categories we can note the following:

Figure 2 exposes the conceptual and empirical impoverishment of many of the quality indicators frameworks discussed earlier. Taking all those illustrated in the first part of this paper other than the NCERT QMT, we can locate them within the framework as follows:

Lesson structure:	-
Space and resources:	the learning environment
Pupil organisation:	class size
Time:	teaching time
	total intended instruction time
	student absenteeism
	learning time
Curriculum	students' attitudes to science
	students' beliefs about mathematics
Task	-
Activity	-
Interaction	-
Judgement	assessment/feedback/incentives
	learning outcomes in literacy, numeracy and science

In the various indicators frameworks reviewed, 'time' is confined to intended time rather than time as used, other factors are perverse in their arbitrariness, while the core components of task, activity and interaction do not feature at all.

At first sight, the NCERT QMT framework fares rather better:

Lesson structure:	lesson/topic introductions recapitulation/evaluation at end of lesson
Space and resources:	school and classroom environment (physical) teaching resources availability of textbooks availability of supplementary materials use of blackboard
Pupil organisation:	school and classroom environment (social) pupil grouping pupil-teacher ratio seating facilities
Time:	opportunity time
Curriculum:	curriculum and TLMs
Task:	-
Activity:	use of TLMs individual/large group/small group activities
Interaction:	teacher dominated/child-centred type of questions asked by teacher whether questions are asked by students
Judgement:	learners' assessment, monitoring and supervision mode and frequency of assessment reporting procedures

However, this too is selective and there are significant gaps, though we have noted that the selectivity is in part a compensatory response to local circumstance and need.

If, however, we combine the two parts of the framework, pedagogy as ideas and pedagogy as practice, we can test quality indicator frameworks not just in terms of which aspects of observable practice they include or, more commonly, exclude (which is a pretty low-level test and a not particularly productive one since it is already abundantly clear that judged in this way many such frameworks are woefully inadequate), but how far they register what the research tells us is most central to learning and teaching and therefore ought to be included in an account of quality.

To take just one component from Figure 2, interaction. Pedagogical, psychological and neuroscientific evidence converge on language, and especially high quality talk, as a key ingredient in young children's development and learning between birth and adulthood, especially in the pre-adolescent years. We refer here not just to the established relationship between language and thought, but to also the more recent evidence on synaptogenesis and neural development (Vygotsky, 1962; Wells, 1992; Bruner and Haste, 1987; Wells, 1999; Mercer, 2000; Cazden, 2001; Wood, 2004; Johnson, 2005, Goswami and Bryant, 2007).

Judged against this developmental imperative, classroom observational research shows us that the interaction which many children experience in classrooms is far from the kind that will maximise cognitive engagement and growth (Galton et al, 1980 and 1999; Edwards and Westgate, 1994; Nystrand et al, 1997; Moyles et al, 2003; Smith et al, 2004; Alexander, 1995, 2001 and 2006b), and in my own international research I have charted the recurrent use of three kinds of teaching talk (Alexander, 2001: 526-7):

- *rote*, or the drilling of facts, ideas and routines through constant repetition;
- *recitation*, or the accumulation of knowledge and understanding through questions designed to test or stimulate recall of what has previously been encountered, or to cue students to work out answers from clues provided in the question; and
- *expository instruction*, or imparting information and/or explaining facts, principles or procedures.

Drilling, questioning and telling are used in some form worldwide, and they certainly have their place. But they remain one-sided. American researchers from the 1960s onwards have documented the dominance of recitation. This endlessly and remorselessly repeats the initiation-response-feedback (IRF) sequence which centres on what Nystrand calls ‘test’ questions to which there is only one possible answer, which the teacher knows and the student must correctly remember, work out or guess. These test questions are contrasted with ‘authentic’ questions which encourage students to think for themselves and which on the basis of his large pre-test/post-test study Nystrand showed were much more likely to lead to successful learning and genuine understanding (Nystrand et al, 1997; Cazden, 2001). The tendency is no less common in Britain.

Albeit considerably less common than rote, recitation and exposition, classroom research uncovers two other forms of pedagogical interaction which have greater power to provoke cognitive engagement and understanding (Barnes and Tood, 1995; Mercer, 2000; Alexander, 2001 and 2006b):

- *discussion*, or open exchanges between teacher and student, or student and student, with a view to sharing information, exploring ideas or solving problems;
- *dialogue*, or using authentic questioning, discussion and exposition to guide and prompt, minimise risk and error, and expedite the ‘uptake’ or ‘handover’ of concepts and principles.

The implications of this brief reference to research relating to just one domain of pedagogical quality are these:

- Interaction needs to be central to indicators of quality.
- It needs to focus on a much fuller spectrum of interaction than, say, what questions teachers ask and whether children ask questions. The research indicates that children’s answers, and especially what teachers do with them, are at least as important.
- A research-responsive set of quality indicators for monitoring classroom interaction would focus in a systematic way on a wide range of interactive features covering the context and organisation of talk, its characteristics in whole class, group, and one-to-one settings, its distribution across all the students in a given class, the character and content of questions, answers and feedback, and so on.

As an illustration of what is possible, one published framework for the interactive component of pedagogy alone has 61 indicators (Alexander, 2006b: 37-43).

We could undertake a similar elaboration, in light of published research, of other components of the framework. Necessarily informing every aspect of teaching is a research literature on learning (e.g. Goswami, 2002; Kuhn and Siegler, 2006, Goswami and Bryant 2007). The literature on the use of time in teaching reaches further than opportunity time and time on

task. Classroom assessment is a subtler, pervasive and more problematic component of pedagogy than is captured by the usual references to test procedures and outcomes. Teachers' theories and beliefs about children do not just inform their classroom decisions; they subliminally shape the climate of expectations, for better or worse, in ways which it is extremely difficult for children to resist. And if we investigate the research on children which attends to what they think and say about education, as opposed to what adults say about them, we acquire a different perspective again. It would be a novel set of quality indicators, but an entirely legitimate one, which replaced the standard effectiveness factors ('maximisation of learning time', 'academic emphasis', 'focus on achievement', 'efficient organisation', 'clarity of purpose', 'high expectations' and so on) by this:

- Creation of a relaxed and enjoyable atmosphere
- Retention of control in the classroom
- Presentation of work in a way which interests and motivates students
- Providing conditions so that students understand the work
- Making clear what students are to do and achieve
- Judging what can be expected of a student
- Helping students with difficulties
- Encouraging students to raise their expectations of themselves
- Development of personal and mature relationships with students

(Brown and McIntyre, 1993: 28-9)

The list is derived from what secondary school students in one research study said that their teachers did well: indicators of quality, in other words, but from the perspective of the learner.

4. The Sting in the Tail: Indicators and Measures

By and large, the EFA literature treats the words ‘indicators’ and ‘measures’ as interchangeable, and slides from one to the other without discrimination or comment. Is this legitimate? Does it matter?

A measure is a device or unit for measuring and is irrevocably tied to quantity. An indicator is an altogether more complex and variable kind of clue about whether something is happening and if so to what extent. Indicators may be gently suggestive of the presence or absence of a tendency or phenomenon, in the way that gathering clouds indicate the imminence of rain, or like the cloudburst they may signal that presence with certainty and precision. ‘Indicator’, therefore, is a much looser and more flexible term than ‘measure’, and may call for a high degree of inference.

Armed with this elementary distinction we immediately perceive additional frailties in our field of enquiry. Most obviously, what are presented as indicators may in fact be measures. For example, the UNESCO ‘indicators’ for Dakar Goal 6 (quality of education) – pupil/teacher ratio, female teachers as per cent of total, educational expenditure as per cent of GDP – are all clear-cut measures, and their measurability can be demonstrated. The problem with confusing them with indicators is that the vital intermediate question – ‘What in the pursuit of quality really matters?’ – has been ignored.

Conversely, what are presented as indicators may be so vague as to serve not even that purpose. For example, the OECD teacher quality ‘indicators’ (OECD, 1994) on which we commented earlier – ‘content knowledge’, ‘reflection’ – don’t actually indicate anything, let alone teacher quality, because they remain undefined and unqualified. Here, the use of the term ‘indicators’ accords the isolated elements a degree of objectivity and precision which they simply do not have. In fact, in this and comparable examples, such lists of supposed qualitative indicators, being neither meaningful nor grounded in discernible evidence, are virtually useless.

More insidiously, indicators and measures may be mixed together within a given framework and may thereby imply to the unwary a consistency of treatment of the phenomena in question which cannot be sustained. Thus, as have seen, the EFA Development Index (UNESCO, 2004) proposes net primary education enrolment ratio as an indicator of universal primary education, literacy rate of 15 year olds as an indicator of adult literacy, and survival rate to grade 5 as an indicator of educational quality. The term ‘indicator’ is used in each case. In fact, net enrolment ratio is a measure of UPE, literacy rate at age 15 is an indicator of adult literacy, and grade 5 survival rate is, as elaborated in the report, a measure of quality. The difference is in the matter of measurement. Measures are proposed for net enrolment ratio (NER) and grade 5 survival, but for adult literacy the measure remains undefined. Whether the two measures in this example are valid or reliable is another matter entirely; the problem of proxies apart, I’ve already suggested that they are not.

The problem of proxies generally arises when – whatever it is called – there is an attempt to make the indicator measurable. This is most conspicuously the case in the debate about indicators of process quality, where the proxies may be so far removed from process reality, or constitute such a minimalist representation of it, as to render them invalid. I cited earlier many examples of this problem and so do not need to repeat them here. The tendency

suggests nothing so strongly as that we are in the realm of the non-measurable, yet those who have used proxies to bypass this inconvenient truth appear to operate in a professional culture where this may not be admitted.

Three propositions or recommendations follow:

- First, it is desirable to maintain the distinction between indicator and measure both to help in our evaluation of frameworks like those from NCERT, and to support the refinement of this and other frameworks. Thus, it is a fairly straightforward task to separate NCERT's indicators from its measures and then assess each in their own terms: the indicators for comprehensiveness and validity in relation to what is described; the measures for reliability and impact in evaluative use.
- Second, if we admit the indicator/measure distinction we might see the explication of valid indicators as an essential first step towards the generation of reliable measures. This would be an important corrective not just to the tendency to muddy the waters generally by confusing indicators with measures but also to the concomitant tendency to proceed straight from a generalised goal (for example, the fostering of good teaching) to its measurement, without the vital intermediate stage of investigation, analysis and description. But, as I have tried to show with my discussion of descriptive frameworks for the analysis of pedagogy, one cannot measure something without first being clear about its nature.
- Third, if we follow this path we might at last be forced to acknowledge two truths which it seems to me are fundamental to this whole debate but are not admitted by many who engage in it: (i) that in the realm of education in general, and in the defining and pursuit of quality in teaching in particular, many of the necessary elements may be describable as indicators but may not be translatable into measures; (ii) that the solution is not to follow the usual path of excluding them as indicators for that reason, reaching instead for the list of proxies and therefore producing an account of teaching which is banal and distorted, but leaving them in place as indicators and looking for some other way to do them justice in contexts such as training, monitoring and evaluation. We leave them in place not to be perverse, but because such inconveniently unmeasurable indicators may well be about what really matters in learning and teaching.

This last, I accept, is a radical proposal, and I can already hear the riposte: 'An indicator for which there is no measure has no useful purpose: away with it!' But is that really so? And is it right that our attempts to understand and evaluate teaching should be subverted by misapplied scientific zeal and/or an imperfect grasp of language? Or that our account of what matters in the pursuit of educational quality should be so seriously distorted by the application of vocabularies devised for contexts a long way removed from the classroom?

In fact, in teaching as in life an indicator for which there is no convenient measure has all kinds of purposes, especially if it is understanding that we are after. In pursuit of EFA we should be prepared to harness the paradigms and insights of any mode of enquiry which can help us, just as those involved in the study of education generally welcome a modern disciplinary eclecticism which is a long way removed from the psychometrics and econometrics which between them cornered the market half a century ago. In qualitative enquiry 'validity' is no less valid for reaching towards authenticity by other than statistical means, and indeed validity in this context is 'an incitement to discourse' (Lather, 2001) of just

the kind that discussion of pedagogy in the EFA context has so sorely lacked.

Few who are familiar with the literature on teaching would now contest the claim that we only began to understand the complexities of pedagogy once we started combining the perspectives and procedures of psychology, neuroscience, sociology, anthropology, philosophy and applied linguistics. Comparing the scope of the first and current editions of the immensely authoritative AERA Handbook of Research on Teaching makes this abundantly clear (Gage, 1963; Richardson, 2001). Yet where in the EFA literature are the disciplines I have cited and the insights they have yielded?

5. Summary

5.1 Quality in EFA discourse

We have charted the shift, in the context of EFA, from an almost exclusive preoccupation with access, enrolment and retention, to a greater interest first in outcomes and more recently in quality. This shift has followed the growing interest (i) in primary education as an end in itself rather than merely a device for selecting the minority who, on the basis of selective tests at age 11 or thereabouts, are deemed worthy of secondary education, (ii) in bringing post-primary education into the definitional frame (so that, for example, India's Sarva Shiksha Abhiyan, a 6-14 elementary education initiative is defined as EFA). We have noted that quality and equity are now recognised as linked rather than separate. We have recognised the beginnings of an acceptance that quality cannot be defined by reference to inputs and outcomes alone and that pedagogical process must be engaged with.

5.2 Indicators of quality

We have shown how mainstream indicators frameworks proposed by agencies such as OECD, the EC, DfID and UNESCO are all highly problematic in their handling of quality generally and pedagogical quality in particular. From being ignored altogether, pedagogy is now included, but in an arbitrary and selective fashion. Indicators frequently display high levels of ambiguity and are capable of such varying interpretation that they are of doubtful practical value. Outcomes and input are frequently smuggled in under the guise of process, or are explicitly offered as proxies. Process itself remains largely invisible.

More generally, there is confusion in the use of the word 'quality' itself. Its usage shifts between noun (quality as attribute) and adjective (quality as aspiration or achievement), or between description and prescription. Most accounts of quality in the EFA literature use the term prescriptively ('quality education' / 'the quality imperative'). This leaves open the question of the descriptive attributes of education to which in the pursuit of quality in the normative sense we should particularly aspire, and allows those who frame indicators of quality to continue to operate in a highly arbitrary way, without reference either to a reasoned pedagogical framework or to evidence about which aspects of pedagogy are most critical to the pursuit of learning. Instead, what is prioritised is what is most readily measurable, regardless of whether it is educationally significant, or is a response to the shifting sands and polarised discourse of educational ideology ('child-friendly' / 'teacher-centred').

The NCERT Quality Monitoring Tools are, in certain respects, an exception to these tendencies, and have two major advantages over the generalised frameworks used by the international agencies. First, they attend closely and knowledgeably to local conditions. Second, they contain not merely one generalised framework but several, on the assumption that at different levels of an education system different kinds of information will be required, and pedagogy is engaged with in progressively greater detail as the QMT move from national to classroom level.

5.3 Criteria for assessing quality frameworks

Nevertheless, reservations can be voiced about the QMT too, and in any event the analysis of the QMT enabled us to crystallise criteria for assessing the adequacy of all such frameworks. The criteria we considered were:

- comprehensiveness
- appropriateness
- consistency
- conceptual/empirical justifiability
- manageability.

These overlapped the more familiar tests of:

- validity
- reliability
- impact.

Each of these we unpacked further. We differentiated descriptive validity from prescriptive, and impact in use from impact in subsequent reporting. To these tests, having raised the question of who at each level of an education system needs to know what, we added

- appositeness of the indicators to what can and needs to be investigated at the levels in question.

Few quality indicator frameworks that we have encountered come close to meeting criteria such as these.

5.4 Sources for defining quality and quality indicators

Since the sources of definitions and indicators of quality in the domain of process are rarely made explicit, we considered next the main sources by which they might be informed:

- national educational policy and the history and culture in which it is embedded
- national pedagogical research
- the international quality indicators literature
- international pedagogical research.

Where quality definitions and frameworks are grounded in national circumstances and research they are able to address local circumstances and needs. They are also likely to be responsive to the nuances of culture which are such an important ingredient of pedagogy and which, in the international EFA literature, are largely ignored. (In the 2005 EFA report there is a brief discussion of the need to respect ‘indigenous’ views of quality – UNESCO, 2004: 34 – but this is instantly undermined by the advocacy of school effectiveness research and the EDI). But indigenous quality frameworks also risk being compromised by the not always easy relationship between research and politics, and the absence of extra-national detachment may cause problems and gaps to be reinforced.

The international EFA literature ought to be able to provide a corrective to the latter tendency but, in the realm of pedagogy at least, it does not. The first part of the paper demonstrated why.

Although applying international research to national situations carries risks of ethnocentrism, cultural colonialism or indeed irrelevance, and although some research is sadly insensitive to local culture, the international research literature on pedagogy is too rich to be ignored.

However, the EFA literature fails to exploit this richness. It attends very little to research on learning, and the cited research on teaching comes almost exclusively from just one tradition, that of school effectiveness research.

We reviewed the main conceptual and empirical flaws in school effectiveness research. They are transparent and serious. Culture and meaning are ignored. The selection of variables is arbitrary. Sampling is weak. The ostensibly firm and objective judgements of effectiveness, success and failure cannot be sustained. The ‘findings’ are obvious and banal rather than insightful. The research itself is exclusive, self-referenced and self-sealing, and immune to the positive influence of the much wider and longer-established body of research on learning and teaching which provides a rich array of alternative sources for the debate about quality.

We also noted that school effectiveness research derives from circumstances very different from those obtaining in India and that in particular they are incongruent with the multigrade setting which remains the default context for the majority of Indian primary schools. Research on multigrade teaching is a necessary additional resource, and it has generated a literature which is both India-specific and international.

However, it might be suggested that those operating in the multigrade context are confronted by a quadruple whammy, in that policy, curriculum, teacher training and the approved models of school effectiveness which inform indicators of quality are all premised on monograde teaching.

5.5 Pedagogy: definitions and frameworks

As a way of signposting the alternative sources, and indeed the framing of accounts and indicators of quality more generally, we proposed a working definition of pedagogy and a comprehensive descriptive framework encompassing both pedagogy as practice and the realms of ideas and evidence which arguably inform that practice. This, finally, responded to our earlier objection that the EFA quality literature is strong on prescription (the adjectival use of ‘quality’) but weak on description of what pedagogy actually entails. This descriptive deficit makes the prescriptions seem all the more arbitrary.

The framework for pedagogy as ideas included ‘enabling’ ideas (on students, learning, teaching and curriculum), ‘formalising’ ideas (on policy and schooling) and ‘locating’ ideas (on culture, self, and identity). The framework for pedagogy as practice included three dimensions: the teaching act itself (comprising task, activity, interaction, judgement), the form that teaching typically takes (lesson), and the contextual and policy frame (space and resources, student organisation, time, curriculum, routine, rule and ritual) within which the act of teaching is set.

Invoking such a framework reinforced our sense of the extreme selectivity and incompleteness of most accounts and indicators of quality in the EFA context, and we illustrated this deficiency by reference to what by common consent – outside the EFA community if not within it – is a central component of learning and teaching, classroom interaction (from the ‘act’ dimension of the second framework). This example hinted at what a properly-conceived account of pedagogical quality might include, and we also glimpsed the radical possibilities of defining quality by starting with what students – rather than teachers, administrators, policymakers and researchers – look for in their teaching.

5.6 Indicators and measures

Finally, we explored the failure in EFA discourse to distinguish indicators from measures, and the damaging consequences of this failure both for our approach to educational monitoring and, more fundamentally, for our understanding of pedagogy. By ruling out of order non-measurable indicators we dismiss from the fields of teacher training and educational evaluation much of what is most important for students, for teachers, and for those who frame the policies within which they work.

Instead, I argued, we should treat the identification of indicators and the development of measures as contingent but distinct, and allow all preconditions for learning, measurable or otherwise, to remain within the frame.

6. Where Do We Go From Here?

Quality is a recent arrival in EFA discourse, and pedagogy is only just beginning to be recognised as central to a proper account of what educational quality entails. The problems which this paper has identified in the way quality is handled, and especially in the way that pedagogy is treated, are real and serious. That the problems are real can be shown by comparing the cursory and skewed characterisation of pedagogy in the international EFA literature with its treatment in mainstream educational research; and indeed with the way that teaching happens in real life, especially in the multigrade settings which are the international default for rural education. That the problems are serious can be confirmed when one considers that if those who target policy, resources and training are given ill-conceived accounts of pedagogy and pedagogical quality to work with, then the resources, policy and training may be misdirected, as may be the work of teachers themselves, and efforts to improve quality and equity may be seriously frustrated.

In attempting to remedy the situation, and give pedagogy its rightful place in the quality debate, I suggest that we (i) sign up to some principles of procedure and (ii) identify priorities for action. Here are my proposals in respect of each.

6.1 Principles of procedure for the handling of matters pertaining to quality in pedagogy

- Clearly separate, conceptually and procedurally, the defining of pedagogical indicators from the development of measures. Do not confine the indicators, and hence everyone's apprehension of what teaching and learning are about, to what can be measured, but instead exploit the indicator/measure distinction to reach a fuller understanding of pedagogy and pedagogical quality. Treat this as the essential first step to devising measures of quality.
- Do not rely exclusively on school effectiveness research. Instead, plug the EFA debate about pedagogy and pedagogical quality into the richer and more extensive mainstream of international research on learning and teaching, and use this as the basis for defining indicators.
- Note that although the greatest need is to fill out the process dimension of quality, the outcome dimension needs attention too. Using test scores in literacy and numeracy alone is not acceptable. Important though these are, they cannot cover for all curriculum domains, especially in the creative and affective spheres.
- Do not try to plug the gap by the over-use of proxies. This, again, arises from too exclusive a concern with measures. If an aspect of pedagogy is important, then it should register in its own right. If it cannot be measured, that is because it is too complex, and alternative ways of monitoring it must be sought. The last thing we should do is ignore vital aspects of learning and teaching simply because they are not readily translatable into measures.
- Heed the example of the NCERT QMT and consider the different perspectives on quality, and the different kinds of indicator and information, which are needed for quality monitoring at different *levels* of the system, from national policy to classroom practice.

- In respect of what purport to be *accounts* of pedagogy and pedagogical quality, apply the test of comprehensiveness.
- In respect of purported *indicators* of quality, apply the tests of validity to what is indicated, and appositeness to the administrative level at which the indicators are to be used.
- In respect of purported *measures* of quality, apply the tests of reliability, manageability and impact.
- Be realistic about the limits to measurement. In the educational sphere few measures are more than rough and ready.
- As a final test, ask whether accounts, measures and indicators of quality focus on what really matters in learning and teaching (a) by reference to a coherent descriptive account of pedagogy, (b) by reference to national/local circumstances, culture and need (including multigrade teaching), (c) by reference to what we know from international research about the conditions for effective learning and teaching.
- In considering definitions, indicators and measures of quality for classroom use, be aware of the dangers of micro-management and the stifling of teacher development and initiative. The intention may be to improve the quality of teaching, but the more teachers are expected to comply with handed-down models and procedures, the less able they will be to handle unpredicted eventualities. No model or procedural framework can possibly meet every eventuality. In the management of pedagogy the law of diminishing returns may apply. Note again the earlier point about ill-conceived frameworks leading to misdirected policies.
- Consider ways of involving teachers in the exploration of pedagogy and pedagogical quality. The account will be even more useful if students, too, are involved. In fact, both the debate about quality and its pursuit in the classroom would be immeasurably enhanced if teachers and students were empowered to participate in it rather than merely enact versions of 'quality' handed down from above. I acknowledge that since that empowerment will come in the first instance from transforming the cultures of teacher training and teaching, a measure of top-down initiative and support is needed, especially in respect of infrastructure, resources and training. Yet the longer-term question we should be asking, if we want 'quality' to mean more than change which is merely temporary or cosmetic, is not 'How can we [i.e. governments, national institutions, international partners, researchers, inspectors, administrators] make teaching better?', but 'How can we so empower teachers and students that together they come to understand the nature and possibilities of learning and teaching and strive to maximise their potential?' The undoubted blessing of having governments which genuinely care about what happens in classrooms can also be a curse if they presume that the only way to make a difference is to take control. The English experience since 1997 illustrates this only too well.

6.2 Priorities for action

Whatever action is taken, I suggest that the priorities should be:

- to place pedagogy, and its training implications, centre-stage;
- to encourage a reappraisal of existing quality monitoring needs and options at each level of the system;
- to introduce a feedback loop into the system whereby we don't just monitor quality but also appraise and refine our procedures for doing so, making habitual the application of tests such as validity, reliability and impact;
- to foreground the continuing reality of multigrade teaching;
- to encourage the appropriate use of the best available evidence – local, national and international;
- to democratise the quality debate, thereby invigorating and empowering those on whom quality at the point of delivery most depends.

References

- Alexander, R.J. (1995) *Versions of Primary Education*. London: Routledge.
- Alexander, R.J. (2001) *Culture and Pedagogy: International Comparisons in Primary Education*. Oxford: Blackwell.
- Alexander, R.J. (2004) 'Still no pedagogy? Principle, pragmatism and compliance in primary education'. *Cambridge Journal of Education*, 34(1), pp. 7-34.
- Alexander, R.J. (2006a) 'Dichotomous pedagogies and the promise of cross-cultural comparison'. In Lauder, H., Brown, P., Dillabough, J. and Halsey, A.H. (eds.) *Education, Globalization and Social Change*. Oxford: Oxford University Press.
- Alexander, R.J. (2006b) *Towards Dialogic Teaching: Rethinking Classroom Talk*, 3rd edition. York: Dialogos. (4th edition 2008).
- Alexander, R.J. (2008) *Essays on Pedagogy*. London: Routledge.
- Barnes, D. and Todd, F. (1995) *Communication and Learning Revisited*. London: Heinemann.
- Benavot, A., Cha, Y-K., Kames, D., Meyer, J.W., and Wong, S-Y., (1991) 'Knowledge for the masses: World models and national curricula, 1920-86'. *American Sociological Review*, 56.
- Brown, S. and McIntyre, D.I. (1993) *Making Sense of Teaching*. Buckingham: Open University Press.
- Bruner, J.S. and Haste, H.E. (eds.) (1987) *Making Sense: The Child's Construction of the World*. London: Routledge.
- Cazden, C.B. (2001) *Classroom Discourse: The Language of Teaching and Learning*. Portsmouth: Heinemann.
- Chand, V.S. and Amin-Choudhury, G. (2006) *Shiksha Sangam: Innovations under the Sarva Shiksha Abhiyan*. Ahmedabad: Indian Institute of Management.
- Creemers, B. (1994) 'The history, value and purpose of school effectiveness studies'. In Reynolds, D., Creemers, B., Nesselrodt, P., Schaffer, E., Stringfield, S., and Teddlie, C. *Advances in School Effectiveness Research and Practice*. Oxford: Pergamon.
- Creemers, B. (1997) *Effective Schools and Effective Teachers*. Warwick: Warwick University Centre for Research in Elementary and Primary Education.
- Department for Education and International Development (2000) *Education for All: The Challenge of Universal Primary Education*. London: DfID.

- Dunkin, M.J. and Biddle, B.J. (1974) *The Study of Teaching*. New York: Holt Rinehart and Winston.
- European Commission (2000) *Sixteen Quality Indicators: European report on quality of school education*. Brussels: European Commission Directorate-General for Education and Culture.
- Edwards, A.D. and Westgate, D.P.G. (1994) *Investigating Classroom Talk*. London: Falmer.
- Galton, M. and Simon, B. (eds.) (1980) *Progress and Performance in the Primary Classroom*. London: Routledge.
- Galton, M., Hargreaves, L., Comber, C., Wall, D. and Pell, A. (1999) *Inside the Primary Classroom: 20 Years On*. London: Routledge.
- Gage, N. (ed.) (1963) *Handbook of Research on Teaching* (first edition). Washington, DC: American Educational Research Association.
- Gage, N. (1978) *The Scientific Basis of the Art of Teaching*. New York: Teachers College Press.
- Goswami, U. (2002) *Handbook of Childhood Cognitive Development*. Oxford: Blackwell.
- Goswami, U. and Bryant, P. (2007) *Children's Cognitive Development and Learning* (Primary Review Research Survey 2/1a), Cambridge: University of Cambridge, Faculty of Education.
- Graff, H.J. (1991) *The Legacies of Literacy: Continuities and Contradictions in Western Culture and Society*. Bloomington: Indiana University Press.
- Gupta, D., Jain, M., Bala, N. (1996) *Multigrade Teaching: Status and implications*. Delhi: NCERT.
- Hamilton, D. (1995) 'Peddling feel-good fictions: Reflections on "Key characteristics of effective schools"'. Liverpool: University of Liverpool School of Education.
- Harlen, W. (2007) *The Quality of Learning: Assessment Alternatives for Primary Education*. Primary Review Research Survey 3/4. Cambridge: University of Cambridge Faculty of Education.
- Jackson, P.W. (1962) 'The way teaching is'. *NEA Journal* 54, pp. 10-13.
- Johnson, M. (2005) *Developmental Cognitive Neuroscience*. Oxford: Blackwell.
- Kuhn, D. and Siegler, R. (2006) *Handbook of Child Psychology*. New York: Wiley.
- Kumar, K. (1991) *Political Agenda of Education: A Study of Colonialist and Nationalist Ideas*. Delhi: Sage.

- Lather, P. (2001) 'Validity as Incitement to Discourse: Qualitative research and the crisis of legitimation'. In Richardson, V. (ed.) *Handbook of Research on Teaching* (fourth edition). Washington DC: American Educational Research Association.
- Little, A.W. (ed.) (2006) *Education for All and Multigrade Teaching: Challenges and Opportunities*. Dordrecht, Netherlands: Springer.
- Little, A.W., Pridmore, P., Bajracharya, H., and Vithanapathirana, M. (2006) *Learning and Teaching in Multigrade Settings: A final report to DfID*. London: Institute of Education, University of London.
- Lockheed, M. and Verspoor, A.M. (1991) *Improving Primary Education in Developing Countries*. Washington DC: World Bank/Oxford University Press.
- Mehta, A.C. (2006) *Elementary Education in India: Progress towards UEE, Analytical report 2005-6*. Delhi: National University of Educational Planning and Administration / Department of Elementary Education and Literacy.
- Mercer, N. (2000) *Words and Minds: How We Use Language to Think Together*. London: Routledge.
- Moyles, J., Hargeaves, L., Merry, R., Paterson, F. and Esarte-Sarries, V. (2003) *Interactive Teaching in the Primary School*. Maidenhead: Open University Press.
- National Council of Educational Research and Training [NCERT] (1996) *School Effectiveness and Learning Achievement at the Primary Stage: International Perspectives*. Delhi: NCERT.
- National Council of Educational Research and Training (1997) *Studies on Classroom Processes and School Effectiveness at the Primary Stage*. Delhi: NCERT.
- National Council of Educational Research and Training (2003) *Indicators of Quality Education at the Elementary Stage*. Delhi: NCERT.
- National Council of Educational Research and Training (2006) *Monitoring Formats for Quality Dimensions Under SSA: Tools for Monitoring*. Delhi: NCERT.
- Nystrand, M., Gamoran, A., Kachur, R. and Prendergast, C. (1997) *Opening Dialogue: Understanding the Dynamics of Language and Learning in the English Classroom*. New York: Teachers College Press.
- OECD (1994) *Quality in Teaching*. Paris: OECD.
- OECD (1995a) *Literacy, Economy and Society: Results of the First International Literacy Survey*. Paris: OECD.
- OECD (1995b) *Measuring the Quality of Schools*. Paris: OECD.
- OECD (2000) *Education at a Glance: OECD Indicators*. Paris: OECD.

- Pratham India Education Initiative (2007) *Annual Status of Education Report (Rural) 2006*. Mumbai: Pratham Resource Centre.
- Rao, Y.A.P. and Rao, R.P. (undated) *School in a Box*. Madanapalle: Rishi Valley Institute for Educational Resources.
- Reynolds, D., Creemers, B., Nesselrodt, P., Schaffer, E., Stringfield, S. and Teddlie, C. (1994) *Advances in School Effectiveness Research and Practice*. Oxford: Pergamon.
- Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C. and Schaffer, G. (2002) *World Class Schools: International Perspectives on School Effectiveness*. London: Routledge.
- Richardson, V. (ed.) (2001) *Handbook of Research on Teaching* (fourth edition). Washington DC: American Educational Research Association.
- Rishi Valley Education Centre (2000) *Report of Second National Workshop on Planning and Implementing of Multigrade Programmes*. Madanapalle: Krishnamurti Foundation.
- Rudduck, J., Brown, N. and Hendy, L. (2006) *Personalised Learning and Pupil Voice*. London: DfES.
- Sammons, P., Hillman, J. and Mortimore, P. (1995) *Key Characteristics of Effective Schools: A Review of School Effectiveness Research*. London: OFSTED.
- Saunders, L. (2000) *What Do We Know About School Effectiveness and School Improvement?* Washington: World Bank.
- Sharpe, K. (1997) 'The Protestant Ethic and the Spirit of Catholicism: Ideological and instructional constraints on system change in English and French primary schooling' *Comparative Education*. 33(3), pp. 329-48.
- Shweder, R.A. (1991) *Thinking Through Cultures*. Cambridge, MA: Harvard University Press.
- Smith, F., Hardman, F., Wall, K. and Mroz, M. (2004) 'Interactive Whole Class Teaching in the National Literacy and Numeracy Strategies'. *British Educational Research Journal* 30(3), pp. 395-412.
- Stigler, J.W. and Hiebert, J. (1999) *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*. New York: The Free Press.
- UNESCO (2000) *The Dakar Framework for Action: Education for All – Meeting our Collective Commitments*. Paris: UNESCO.
- UNESCO (2003) *Gender and Education for All: The Leap to Equality (EFA Global Monitoring Report 2003/4)*. Paris: UNESCO.
- UNESCO (2004) *Education for All: The Quality Imperative (EFA Global Monitoring Report 2005)*. Paris: UNESCO.

UNESCO (2006) *Strong Foundations: Early Childhood Care and Education (EFA Global Monitoring Report 2007)*. Paris: UNESCO.

UNESCO (2007) Education For All Development Index (EDI). Available online: at <http://portal.unesco.org/education>

Vygotsky, L.S. (1962) *Thought and Language*. Cambridge, MA: MIT Press.

Wells, G. (1992) 'The Centrality of Talk in Education'. In Norman, K. (ed.) *Thinking Voices: The Work of the National Oracy Project*. London: Hodder.

Wells, G. (1999) *Dialogic Enquiry: Towards a Sociocultural Practice and Theory of Education*. Cambridge: Cambridge University Press.

Wood, D. (1998) *How Children Think and Learn*. Oxford: Blackwell.



Consortium for Research on
Educational Access, Transitions & Equity
Funded by DFID

Report summary:

This monograph critically examines the emerging discourse on quality associated with Education for All (EFA). It contends that EFA discourse has moved from a welcome and vital commitment to quality to its measurement without adequate consideration of what 'quality' entails, particularly in the vital domain of pedagogy. Meanwhile, the demand for quality indicators by governments and international agencies has left important methodological questions unanswered. Citing various international examples, the paper notes a concern with input and context at the expense of the proper elucidation of educational process and outcome, arbitrariness in what is focused upon, excessive use of proxies, overly selective use of international research on learning and teaching, and confusions about the key terms 'quality', 'indicators' and 'measures'. The paper proposes criteria for assessing frameworks for evaluating the quality of classroom provision, central to which are evidential breadth, conceptual comprehensiveness, validity, reliability, impact, manageability and appositeness to level and context of use. Arguing that at root the problem is as much conceptual as empirical and procedural, the paper proposes a map of the territory of pedagogy at the levels of ideas and action, together with principles of procedure to guide future work on indicators and measures of quality in the EFA context. Originally prepared for use in India, the paper uses examples from that country by way of illustration.

Author notes:

Robin Alexander is Fellow of Wolfson College at the University of Cambridge, Emeritus Professor of Education at the University of Warwick, and former Professor of Education at the University of Leeds. He has served on UK government advisory bodies and enquiries, and has undertaken research, evaluation and consultancy in many other countries, including a long association with the reforms in India which provide the context for this monograph. His five-country comparative study *Culture and Pedagogy* won the 2002 American Educational Research Association Outstanding Book Award and he builds on this work in his latest book, *Essays on Pedagogy*. He currently directs the Primary Review, the most comprehensive enquiry into English primary education for forty years.

Author's Address:

Professor Robin Alexander
Faculty of Education
184 Hills Road,
Cambridge CB2 8PQ, UK
Email: rja40@cam.ac.uk
Website: www.robinalexander.org.uk

CREATE Address:

CREATE, Centre for International Education
Sussex School of Education,
University of Sussex
Brighton BN1 9QQ, UK.
Email: create@sussex.ac.uk
Website: www.create-rpc.org

