



CENTRE FOR
SOCIAL SCIENCE RESEARCH

**HOW DO WE KNOW WHAT TO DO?
THE IMPORTANCE OF
EXPERIMENTATION IN DESIGNING
PROGRAMMES AND POLICIES TO
IMPROVE SOUTH AFRICAN
SCHOOLING**

Jeremy Seekings

CSSR Working Paper No. 30

Published by the Centre for Social Science Research
University of Cape Town
2003

Copies of this publication may be obtained from:

The Administrative Officer
Centre for Social Science Research
University of Cape Town
Private Bag
Rondebosch, 7701
Tel: (021) 650 4656
Fax: (021) 650 4657
Email: kforbes@cssr.uct.ac.za

Price in Southern Africa (incl. VAT and postage): R 15.00

or it can be downloaded from our website
<http://www.uct.ac.za/depts/cssr/pubs.html>

ISBN 0 7992 2170 8

© Centre for Social Science Research, UCT, 2003

CENTRE FOR SOCIAL SCIENCE
RESEARCH

Social Surveys Unit

**HOW DO WE KNOW WHAT TO DO?
THE IMPORTANCE OF
EXPERIMENTATION IN DESIGNING
PROGRAMMES AND POLICIES TO
IMPROVE SOUTH AFRICAN
SCHOOLING**

Jeremy Seekings

CSSR Working Paper No. 30

February 2003

Jeremy Seekings is Director of the Social Surveys Unit and Professor of Political Studies and Sociology at the University of Cape Town.

How Do We Know What to Do? The Importance of Experimentation in Designing Programmes and Policies to Improve South African Schooling

Abstract

*The quality of schooling in South Africa is a continuing cause for concern. There is broad consensus around the goals of reform, but not about how these goals might or should be achieved. This is because there is little evidence on the effects of alternative reforms allowing assessment of their efficacy and utility. This paper argues for much more experimentation in educational reform, i.e. running experimental reforms and monitoring closely their effects. Some evidence is available from projects and programmes funded by foreign donors, business donors or occasionally provincial education departments. This paper examines some of these, focusing especially on the design and evaluation methodology of the **Imbewu** programme run in the Eastern Cape funded through British aid. Unfortunately, little attempt has been made to standardise or collate research on the effects of these initiatives.*

How Do We Learn About Improving Educational Outcomes? ¹

There is a substantial body of work documenting the poor performance of South Africa's public schools, and a similarly substantial body of evidence on the main causes of this poor performance. Too many South African children receive a poor education, and hence poor opportunities in life, not so much because of a shortage of resources invested in schooling, but rather because the considerable resources that are invested are invested very inefficiently. There are too few textbooks in classrooms, teachers are poorly motivated and sometimes incompetent, and management is all too often deficient. Many children bring considerable disadvantages with them to school, on account of their family background and neighbourhood environment. The schooling they receive is

¹ This *Working Paper* comprises Chapter 6 of a Report commissioned by the Standing Committee on Public Accounts, in the South African Parliament, on *Making An Informed Investment: Improving the Value of Public Expenditure in Primary and Secondary Schooling in South Africa*. Research for the report was funded by the European Commission.

insufficient to overcome these disadvantages, ensuring that inequalities in society are reproduced. Disadvantaged children get the worst opportunities.

All of this is widely recognised within national and provincial education departments. The various education departments have identified what they aim to achieve. Now, as the Joint Education Trust has suggested, the question is not so much ‘what to do?’ as ‘how to do it?’

‘During the first five years of South Africa’s first democratic government the Department of Education has established a coherent policy framework for the reform of the public schooling system. The priority now is to implement these policies in order to improve the effectiveness of schools in delivering quality education. The problem is no longer what to do – this has been established by much consultation, research and parliamentary debate, and regulated and legislated in numerous policy statements, White Papers and Government Gazettes. The problem now is how to implement these policies.’ (JET *et al.*, 1999: 8)

The activities that attract most attention in the media (in large part because the politicians involved seek media coverage) are likely to be of secondary importance. The Minister of Education visits schools and admonishes teachers and students for negligence. Former President Nelson Mandela presides over the opening of new school buildings in rural areas, paid for by big business. For sure, both the culture of schooling and physical infrastructure are important, but it is far from clear what effect admonitions and new buildings have on what actually happens inside the classroom, and whether student performance is significantly improved.

Elected political leaders and appointed officials have to make decisions on how to allocate scarce resources, with the goal (we hope) of maximising the benefit to children in terms of the quality of education. They do so without adequate information on the costs and benefits of alternative uses of resources. There is a pressing need to generate such data. This working paper is concerned with developing the *procedures* required to identify the benefits and costs of different policies. How can we know which policies ‘work’, and which do not, in terms of building students’ skills? What can we do – and what should we be doing – so that we are confident that public resources are being put to the best possible use?

The suite of forms of assessment that are being developed by the national Department of Education might provide some information of what works and what does not work. This suite comprises the following (see further, Seekings, 2002):

- The existing matriculation examination: this provides an indication of the performance of the one-third to one-half of South African children who reach grade 12;
- The national Assessment Policy for students: this provides detailed information, through continuous assessment and a public examination (with some external moderation) at the end of grade 9, in order to assess the performance and progress of individual children;
- ‘Systemic evaluation’ of the schooling system: testing of samples of pupils in specific grades across the country in order to assess the performance of the education system as a whole;
- ‘Whole school evaluation’ of individual schools: a combination of external and self-evaluation, but without external assessment of pupils’ performance, in order to assess the performance of individual schools; this is an essentially qualitative evaluation.

In principle, if these were conducted in an exemplary manner, they would generate much of the information required to identify what ‘works’ and what does not. The results of the grade 9 and 12 examinations, together with the results of testing as part of systemic evaluation, would provide some indication of the performance of individual schools. The Whole School Evaluation would provide some information on what schools are actually doing, which might reveal the differences between schools that produce good results and those that produce poor results (however measured, and taking into account family and community factors). And the information on individual students would provide a very detailed ‘micro’ picture of what this means for individual children.

But the contribution of this suite of forms of assessment will be undermined by two major problems. First, there is insufficient regular, external testing assessment of pupils. Secondly, it will be no easy task to ensure that the information provided by Whole School Evaluation is both of high quality and usable. Both problems stem from the scale of the exercise: an estimated 2 500 specialist supervisors, across the country, visiting 27 000 schools at regular intervals and writing up their reports. The numbers of people involved is perhaps of less concern, since they will be playing a valuable developmental role, advising schools and teachers on how to do better. But the number of teams

of supervisors visiting schools means that there will be a flood of reports to provincial and national departments, and it is far from clear what bureaucratic procedures and structures will be capable of making sense of this.

In short, whilst the suite of forms of assessment might provide some data on performance and some information required for decentralised developmental efforts (i.e. for teachers to assist students whose progress is slow and supervisors to assist schools and teachers whose practice is deficient), it is unlikely to provide sufficient data on what ‘works’ and what does not. These forms of assessment are necessary but not sufficient. In assessing *how* to improve students’ performance, *experimental* programmes and projects are indispensable. Such programmes involve concentrating assessments on small numbers of schools that have been selected as the sites of focused interventions or reforms. The Department of Education’s suite of forms of quality assurance is a suite designed for an education system that is basically working well, and simply needs some tinkering at the edges; it is not designed to deal with an education system that is in crisis in many respects, and which probably requires radical interventions and reforms.

Careful planning, implementation and evaluation in experimental programmes should provide lessons that save time and money when broader reforms are attempted. Sometimes educational programmes are not designed as experiments, but nonetheless may serve this purpose. Conversely, projects may be set up as experiments but have no value due to careless design or shoddy evaluation. It must be emphasised that experimental programmes are not a *substitute* for the kinds of assessment that the various departments of education have set up or are planning; experimental programmes cannot provide the important information that other forms of assessment generate. Rather, they are a complement, providing different information for different purposes.

The use of experimentation is strongly advocated by Hanushek *et al.* in their study of education in the USA, *Making Schools Work: Improving Performance and Controlling Costs*. Hanushek *et al.* (1994) note that, in the USA (as in South Africa), there are grave concerns with the efficiency of public spending on education. In the USA, standards have been stagnant (or even fallen) whilst costs have risen. ‘The nation is spending more and more to achieve results that are no better, and perhaps worse’ (*ibid*: 25); America’s ‘schools are demonstrably inefficient’ (*ibid*: 7). Hanushek and his co-authors, who include noted left-wing economists, argue that both the costs and the benefits need to be taken into account in evaluating programmes to improve the quality of

education. Above all, Hanushek *et al*, argue, schools themselves need to be transformed into ‘learning institutions’. ‘How’, they ask (*ibid*: 125), ‘can schools hope to improve systematically without ascertaining what is and is not currently working?’ What is needed is an altogether new commitment to evaluation. Hanushek *et al* condemn existing approaches:

‘Schools invest remarkably little in information about their own operations. Compared with that on private industry, the amount of research and development on schools is minute. So also is the amount of resources and energy devoted to quality assurance in schools. More than that, school personnel are accustomed to think that evaluation of programs and activities is unimportant, more a nuisance than a necessity.’ (*ibid*: 148)

Evaluation is not the same thing as a slavish use of indicators, whether indicators of inputs or of outputs.

‘Some argue that improved student testing and assessment can lead reform by itself. With accurate assessments of student performance from tests, runs the argument, schools will automatically focus on improving test scores: teachers will adjust to improve student achievement, management decisions will become obvious, and reform will occur almost spontaneously. We know of no evidence to suggest this is a realistic expectation.’ (*ibid*: 126)

Improved assessment is a necessary but not a sufficient condition for successful educational reform. What is needed in addition is the development and initiation of ‘a systematic approach to experimentation and evaluation of school programs’ (*ibid*: 127).

Hanushek *et al*. advocate ‘randomised assignment experimentation’. In order to isolate the effect of any particular policy intervention, the effect of the intervention on one group of schools and students needs to be compared to a control group of schools and students which is identical to the target group in all ways except that the intervention was not extended to them.

‘Suppose, for example, that one wants to study the effect of smaller classes on student achievement. If researchers simply look at performance in the large and small classes found in a typical school district, they cannot be sure whether any observed differences in

student performance come from the effects of class size or from the specific students and teachers in the different classrooms. For example, if the school principal attempted systematically to provide small classes for the students performing at the lowest level, performance might appear to be lower in the smaller classes, even though that is actually not an effect of the smaller classes per se. To overcome such problems, students (and teachers) could be randomly assigned to classrooms with varying numbers of students, which would reduce the chance that pre-existing performance differences would bias measurements of performance by class size.’ (*ibid*: 143)

Clearly this is easier with some kinds of reform (i.e. experiment) than others, but the general point is valid.

The lessons of experimental projects are clear from international experience. A classic example concerns class size. Reduced class size is often advocated as a, if not *the*, main way of improving student performance. It is also one of the most expensive possible reforms, as smaller class sizes require more teachers, and salaries already account for the lion’s share of expenditure in most school systems. In 1985, the American state of Tennessee initiated an experimental programme to ascertain what difference was made by reduced class sizes. At the time most classes had between 22 and 25 students in them. The programme (named STAR, for ‘Student/Teacher Achievement Ratio’) covered 79 schools, including kindergarten and the first three grades of primary school, for a four-year period. Students in these schools were assigned randomly to either (1) standard sized classes, (2) standard sized classes with a full-time teaching assistant in addition to the teacher, or (3) smaller classes, with between 13 and 17 students. Each year, students were tested in reading, maths and other skills. The experiment found that reduced class size resulted in a small improvement in students’ performance in kindergarten and perhaps in first grade also, but made no difference in grades two and three. The presence of a teaching assistant made no difference in the regularly-sized classes. The experiment suggested that the huge increase in expenditure required to reduce class sizes in grades one to three would probably be wasted (Hanushek *et al.*, 1994: 144-5).

In South Africa, as we have already seen, huge resources have been spent on reducing class size in larger classes. Porteus (2000) is typical of many writers in suggesting that we should assume that class size makes a difference until we have powerful evidence to the contrary. Our evidence is still weak, and not entirely consistent. Case and Deaton (1999) suggest that class size makes a

considerable difference around the ages of thirteen and fourteen, but no difference at other ages. Crouch and Mabogoane (2001) found that it made some difference for some subjects. But, to the best of my knowledge, there is no *experimental* evidence on the effects of class-size in South Africa. In a country with nearly 27 000 public schools, it is bizarre that no experimental research has been done into the effects of class size on learner achievement. Of course, the need for equity required that class sizes in most schools were adjusted after 1994 to redress the racial discrimination of the apartheid years. The departments of education rightly made these adjustments rapidly. But now, faced with improving the quality of education across the board, it would be very useful to know precisely what effect there would be on pupils' performance if additional resources were invested in the number of teachers in schools (or, more specifically perhaps, what other conditions are required for such an investment to pay dividends). This information might be generated through a programme under which the national Department provided conditional grants to the provinces for the specific purpose of reducing class sizes in some schools or in some districts, and monitoring carefully the progress made by students in these schools compared to students in schools that were similar in all respects except that their class sizes were not reduced.

Hanushek *et al.*'s recommendations need to be applied to the South African case. We can study international experiences of policy reforms as much as we like, but the exercise will provide 'answers' to the question 'what works in other countries?', not 'what works in South Africa?' What is needed is the systematic analysis of the massive amount of experience that exists in South Africa as the inevitable consequence of investing more than R50 billion or so each year in primary and secondary schooling. This analysis requires urgently not only the continued development of data collection on both financial and other inputs and the output of student achievement, but also the design of educational investments along quasi-experimental lines.

The Range of Recent Experiments in South Africa

South Africa has spent huge sums on education, but very little of this has been spent in ways that can be considered experimental. Policies have rarely been designed or assessed in ways that provide evidence on their effects. National and provincial education departments have rarely designed and assessed their core reforms in these ways, although there are notable cases (such as the Gauteng

Education Action Zones, discussed below), where reforms have had experimental value. There are no doubt many reasons for this lack of well-designed and measured experimental reforms, but I imagine that they include: the inadequacy of systems of assessment and evaluation that the post-1994 government inherited from the apartheid state; the shortage of discretionary funds in most provinces; the prevalence of ‘inside-out’ philosophies of change in education (see Muller with Roberts, 2000), and a pervasive and related hostility to formal testing that is indispensable in the evaluation of experimental programmes.

Most of the experience with experimental programmes comes from partnerships between education departments and the private sector, in which funds from private or foreign sources have been used to supplement the standard provisions of the public sector. There have been, in South Africa, a large number of school improvement programmes that should generate important evidence on what works and what does not work in school improvement. As Taylor (2001) writes, ‘there is no shortage of effort in the field of school reform. Besides the government’s own programmes in this area, it is estimated that something in the order of 20 percent of the nation’s nearly 30 000 schools are involved in donor- and NGO-initiated development projects of one or other kind, with a total off-budget expenditure of around R200m annually.’ Other estimates of the value of donor funding for school reform programmes put the figure much higher. In 1999/2000, South African business is estimated to have invested R645 million in education projects.²

Just as there are reasons why education departments have been slow to adopt an experimental approach to reform, there are clear reasons why the private sector has done so. Most obviously, donors and NGOs are very aware that the funds they commit to education are paltry in comparison to the investment of tax revenues by the state. If the sole purpose of their expenditures was to improve the quality of schooling for the children directly affected, their efforts would be insignificant. But if their funds are invested wisely, they should have a much greater impact indirectly, through demonstrating what kinds of intervention work and what kinds do not. The experimental features of these programmes driven by the private sector are often explicit. The National Business Institute, for example, supported the Quality Learning Project (QLP, see below), on the basis that the programme’s design would enable lessons to be spread to other parts of the country and the successful features of the programme replicated.

² *Sunday Times*, 20 May 2001.

Curiously, neither the national nor the provincial departments of education appear to have compiled adequate lists of existing programmes, and are yet to conduct any rigorous analysis of the lessons of these programmes. In 1999 the national Department, through its National Centre for Curriculum Research and Development (NCCRD), initiated an audit of curriculum-related research and developmental projects. The audit revealed a total of 326 projects, ranging from the massive to the tiny, including projects that were not primarily curriculum-oriented as well as those that were. But the published report from this audit (Department of Education, 2000b) provides very little analysis or information. The Gauteng Education Department has also compiled a list of its partnerships with private organisations, as well as preparing a policy framework governing its partnerships. In the Western Cape, the provincial education department has taken the initiative in establishing a ‘Western Cape Education Trust’, one of whose first tasks has been to compile a list of programmes in the province. It does not appear that any such list exists for the Eastern Cape. None of these initiatives appears to be concerned with learning the lessons of education programmes. Their purpose seems to be promote co-ordination (to be generous) or departmental control (to be cynical).

The primary attempts to collate purposefully basic data on a number of major programmes and their evaluations, with the goal of systematically examining using the experience of programmes to identify what works, have been made by JET. In 1995-96 JET conducted an investigation into INSET programmes (see JET, 1996). More recently, JET embarked on a careful study of a selection of major programmes in a range of fields. Besides this specific line of inquiry, JET has amassed a collection of perhaps one hundred and fifty project evaluations – a valuable archive of research. Government departments might usefully build similar collections; if they have begun to do so, this is not apparent to senior officials within those same departments!

Critical assessments of programmes in the early- and mid-1990s are not encouraging. Writing about INSET programmes in early 1996, Jansen concluded:

‘Teacher development delivered through NGOs in the form of in-service education (INSET) does not produce significant learning gains in the classroom. One evaluation report after another has delivered the same finding: that while INSET provides important motivational benefits to practising teachers, and begins to influence the behaviours of participating teachers, such programmes simply do not translate

into learning gains for students. Yet millions of rands have been invested in teacher development programmes...' (Jansen, in JET, 1996: 13).

In a thoughtful reflection on recent, independently-initiated programmes with a wider ambit than in-service teacher training alone, Taylor acknowledges that 'the effects of these efforts have been hard to discern to date', approaches are sometimes inappropriate, and the relationships between NGOs and state are fraught with tensions (2001: 1). As we shall see below, further criticisms can be levelled at these experimental programmes. Notwithstanding all of this, the experience of school reform programmes provides the most important evidence on what works and what does not work in school improvement, evidence which is required to guide us in the reform of South Africa's schooling system as a whole.

Some indication of the range and scope of the experimental programmes that are underway in different parts of South Africa is provided by the following brief descriptions of seven major interventions. Most of these programmes are in early stages, but they are all pioneers in one way or another and thus demand analysis, even if our analysis is necessarily tentative or preliminary.

READ began work in Soweto over twenty years ago, and has since extended to primary schools across the country. READ 'offers pupils a sound, book-based education foundation' (READ, 1998: 3). Its work entails teacher training and mentoring, together with the provision of a range of books (to replace reliance on a single textbook), with the goal of improving pupils' reading and writing skills. As READ emphasises: 'Teachers cannot teach and pupils cannot learn effectively, nor can the new curriculum be efficiently implemented, without the delivery of cost effective teaching and learning materials, together with training in their use' (READ, 1999: 28) Much of READ's work is organised and funded as separate projects. For example, READ ran a pilot project in what used to be the Transkei and is now part of the Eastern Cape, beginning in 1995 and concluding in 1998. In the four years of this particular project, 187 courses were attended by an average of twenty-three teachers. At an annual cost of only R53 per pupil, READ argue that their model is one that can be widely replicated (and, moreover, would cut repetition rates and thus save money). In 1999 READ launched its Learning for Living project, with an award of R153 million from the Business Trust. This project is intended to reach 12 000 teachers in about 900 schools, and hence a total of one million pupils.

Imbewu (meaning ‘seeds’ in Xhosa) is a school improvement project in the Eastern Cape. It was initiated in October 1997 with a grant of R75 million from the British Department for International Development (DfID). In 2001 it was extended for a further five years, with a new budget of R240 million. In its first phase it focused on five hundred primary schools in clusters each of five schools. The project sought to improve school management (including the functioning of governing bodies formed by parents), to supply teaching and learning materials, to improve teaching through short courses and workplace support, and to improve management at the district, regional and provincial levels. There was also systematic monitoring and evaluation, with a full evaluation undertaken at the end of the first phase (1997-2000). The second phase is expanding the project to cover 3 500 schools.

Mahlahle, like *Imbewu*, seeks to improve school management as well as classroom practice. It is located in the Giyani and Phalaborwa districts of the Northern Province, where it has targeted fifty primary and thirty-six secondary schools. As with *Imbewu*, most of the participating schools are in rural areas that were part of the former bantustans. *Mahlahle*, meaning ‘the bright morning star’, had an initial budget of just R 29.5 million for the four years 2000-2003.

The **Education Quality Improvement Programme (EQUIP)** is a project of the National Business Initiative, which was founded in 1995 by senior business leaders with the goal of helping to ‘make the new South Africa work’. The NBI explicitly identifies its programmes as experimental, intended to produce ‘systemic change that could be replicated or implemented at scale in socio-economic delivery mechanisms’. EQUIP is one of the NBI’s major programmes. The NBI assessed that high failure rates in schools were primarily due to a ‘lack of governance and management skills in schools and districts, as well as a lack of quality teaching by teachers’ (NBI, 2000: 10). EQUIP would focus on local management and planning. Schools would be helped to prepare their own development plans, which would have to be discussed and approved in the school governing body. The necessary resources would be secured through sponsorship by business, to the tune of about R20 000 p.a. for three years. The development plan would be implemented with the assistance of whatever service providers were required. EQUIP first began to work with schools in 1996. The pilot phase included fifty schools, but the programme has since expanded to 320 schools in Gauteng, the Western Cape, KwaZulu-Natal and (most recently, in 2001) the Eastern Cape. EQUIP serves as a partnership between business and government, helping communities to help themselves. As NBI puts it: ‘Communities will have to take the initiative in upgrading their schools, since

the government cannot do it alone' (NBI, undated: 4). EQUIP's role is to guide, support and encourage community efforts.

The **General Education and Training In-Service Project** (GET-INSET) is a collaborative project between a consortium of NGOs and the Western Cape Education Department. Initiated in 1998, it is now involved in 85 historically disadvantaged primary schools in urban and rural areas. Its four-year budget amounts to about R10 million. It provides subject-specific teacher training in workshops, together with textbooks and other 'learner support materials', and in-school support. It sees itself as a pilot project, with an experimental design.

The **Quality Learning Project** (QLP) is a new programme, focused on five hundred secondary schools in twenty districts across the country (including six in the Eastern Cape but just one in the Western Cape). The QLP seeks to improve the maths, reading and writing abilities of high school students in the selected schools. The focus is on mathematics and language because it is believed that these provide the foundation for all other learning. Like *Imbewu* and *Mahlahle*, the QLP also entails improving district and school management, teacher development and systematic monitoring and evaluation. The total budget for the QLP is R139 million over five years, funded by the Business Trust.

A programme that focuses entirely on management is the **District Development and Support Programme** (DDSP). The DDSP began to be implemented in 2000. It is focused on district-level management, and seeks to strengthen the links between district officials and schools. It is funded by the USAID.

The **Education Action Zones** (EAZ) programme is the only one of the programmes discussed here which is entirely a departmental initiative and not an NGO initiative (albeit run in partnership with provincial education departments). The Gauteng Department of Education initiated the EAZs programme in January 2000, targeting seventy-one 'profoundly dysfunctional' secondary schools (all scoring below 20 percent in the 1999 matriculation examination). The programme did not involve much additional funding, but rather intensive monitoring (Fleisch, 2001).

Taylor has distinguished between different 'generations' of programmes, with the second generation learning from the experience of the first generation. READ and GET-INSET are what 'first generation' programmes, focused heavily on what goes on inside the classroom, and aiming to enhance pupil achievement through teacher training and the provision of books and other

learner support materials. *Imbewu* and *Mahlahle* are ‘second generation’ programmes, combining a concern with teaching methods with a concern with educational management. Both programmes invest heavily in training and support for principals, district level managers and personnel in provincial education departments.³

The second generation programmes benefited from the lessons learnt from their first generation predecessors. By 1996 it was clear that teacher training alone was having too little effect on pupils’ performance (JET, 1996). The shift to the second generation also reflected the prevalence in official rhetoric of a concern with ‘governance’ and management issues (see the Education White Paper and the 1996 South African Schools Act). The experience with the second generation programmes has, in turn, informed the design of the most recent programmes. Taylor includes the QLP in his category of ‘second generation’ programmes, but it is perhaps not so easy to classify. In addition to emphasising classroom practice and managerial reform, the QLP also stresses the importance of teachers’ subject knowledge. Perhaps this might be thought of as a ‘third generation’ programme, informed by the experience of *Imbewu* and other second generation programmes which alerted planners to the problem of teachers’ lack of basic subject knowledge.

An alternative typology for categorising programmes might distinguish between the two kinds of educational reform that are conducive to experimentation. Some reforms or programmes entail the direct allocation of *additional resources*. For example, funds are provided for additional teachers or, more often, for additional training for teachers or managers. A second kind of reform is *institutional* in nature, involving changes to systems of governance, incentives or accountability. These institutional reforms would have indirect effects on the *use* to which resources are put. In practice, programmes generally entail elements of both kinds of reform, but in different proportions.

³ As Roberts points out, *Imbewu* was primarily concerned with teacher development, and might be thought of as on the front edge of the second generation rather than being firmly in that generation.

Table 1: Some characteristics of selected programmes

<i>Programme</i>	<i>Additional resources invested</i>	<i>Institutional reform</i>
READ	LSMs (plus teacher training and mentoring)	-
<i>Imbewu</i>	Teacher/manager training; LSMs; Technical support for provincial Department of Education	Increased community involvement; ‘Vision-crafting’
<i>Mahlahle</i>	As above	-
EQUIP	Up to the school to decide	Development planning (including increased community involvement)
GET-INSET	Teacher training (plus LSMs, in-school support)	
QLP	As with <i>Imbewu</i>	-
DDSP	-	-
EAZs	In-school support; LSMs	Accountability (to provincial education dept)

LSMs: ‘Learning support materials’, i.e. textbooks etc.

Most of the programmes described above concentrate on the provision of additional resources for targeted purposes (see Table 1). Unsurprisingly, sums are invested in much the same general categories: training for teachers, books and other learning support materials, and in-school support. Some (the ‘second generation’) programmes also include management training, others (the ‘first generation’ programmes) not.

Most of these programmes have only a minor institutional component. Only the Education Action Zones has a primary focus on institutional reform; in the case of the EAZs, the focus is on accountability. Most of the programmes have only a limited conception of accountability and limited provision for incentives. This is because these programmes are driven by the private sector (albeit in partnership with provincial education departments). NGOs and donors clearly cannot call schools, teachers or education officials to account in the same way as provincial departments, which use threats of disciplinary or other action. Nor can they restructure the governance systems of schools beyond what is permitted by national or provincial policy.

NGOs and donors have only two, weak ways of demanding accountability. First, they can threaten to withdraw. In cases like EQUIP, where there are financial incentives to persevere, the threat of withdrawal might prove significant. Secondly, they can try to boost the capacity of the ‘community’ to hold schools to account. For sure, some of these programmes require a demonstrable commitment from school governing bodies, which might serve to maintain some accountability. But these are lines of accountability that presumably did not prove effective before. It is difficult to avoid the impression that programmes like READ and *Imbewu* and the QLP have in common an assumption that the provision of skills, whether to teachers or managers, will result in improved performance. In other words, they diagnose the problem as primarily one of a lack of skills, not a lack of incentives or pressure. In this they run counter to much of the international literature on educational reform, which attaches considerable importance to issues of accountability and incentives (see Hanushek *et al.*, 1994).

While there are good reasons why most experiments have been driven by the private sector, there are constraints on the both the scale and range of experiments that the private sector can drive in public schools.

- Firstly, the sums spent by the public sector dwarf those spent through NGOs; the scale of experimentation potentially available within the public sector therefore dwarfs the possibilities for the private sector;
- Secondly, the private sector has a limited ability to experiment with institutional reforms;
- Thirdly, it is the state’s responsibility to ensure thorough evaluation of all experiments in order to assess what is ‘best practice’ in different situations, and to ensure that the investment of public monies in education is an *informed* investment.

For these reasons it is especially imperative that the public sector – meaning the national and provincial education departments – embraces an experimental approach to educational reform.

The Design of Experiments: A Case-Study of *Imbewu*

The case of *Imbewu* provides a useful study of how these independently-initiated programmes are designed and implemented. *Imbewu* is probably the most advanced of what Taylor terms the ‘second generation’ programmes. It is concerned not only with practices inside the classroom but also with governance and management within the schooling system. According to Taylor (of JET):

‘*Imbewu* is the Xhosa word for ‘seeds’. Botany offers a very apt metaphor for schooling. Classrooms, which nurture the minds of our future citizens, may be likened to the leaves of a plant where its food is manufactured. But the leaves cannot exist without the branches which support and supply them with their needs. Similarly, classrooms are maintained by a set of institutions, starting with budgetary allocations and policies at the national level, which are implemented through an array of management and supply lines through provincial, regional, district and school offices. And the entire edifice would not stand without deep roots in the community and the ongoing leadership of parents and civic leaders through community-based governing structures. Public schooling is a truly systemic enterprise: no one part of the system can function effectively if all the other parts are not in good working order’ (Taylor, 1998: 7).

The *Imbewu* programme was designed through a process of ‘negotiations’ between the donor (the British Department for International Development, DfID) and the national and provincial departments of education. Details were then worked out ‘by the project team, led by JET, in close discussion with the [provincial] department, and under the overall authority of the project steering committee’.⁴ Planning took into account international models, in part through the involvement of British advisors brought in by the donor, DfID. Planning began in 1995, with implementation starting in October 1997. The programme was located in the Eastern Cape, officially because of the dire state of education there, but perhaps also because the British government wanted to be seen to be doing something in the ANC’s heartland.

The project team diagnosed the problem in Eastern Cape schools as a systemic problem. The need for systemic reform was underscored by the difficulties in

⁴ Nick Taylor, personal communication.

both establishing a baseline study of administrative capacity in the provincial education department, and evaluating progress during the *Imbewu* programme (see JET, 2000a). Baseline studies of classroom practice, student performance and school management and leadership, by Eric Schollar, confirmed the appalling state of education in the province (ESA, 1999).

Imbewu combines a strong emphasis on classroom practice with a similar emphasis on management and governance. In-service training is provided for three 'key' teachers from each participating school. This training comprises short courses interspersed with practical assignments and workplace support. The programme also provides a package of 'learning support materials' (stationery, books and other teaching materials), as well as training on their use and management. With respect to school management, managers at every level (i.e. school, district, regional and provincial) are provided with training (and technical assistance) in financial, information, provisioning and human resource systems. School principals are recognised as especially important players. Community involvement is also encouraged.

Classroom and management/governance training are partially integrated through a comprehensive training programme. This comprises thirty-four training modules, including four introductory modules, five EMD (education management development) modules, twenty INSET modules, and five further modules. The introductory modules cover the basic *Imbewu* approach, namely 'vision crafting', 'whole school development', managing change and the theoretical approach of 'practice-based inquiry'; these modules are attended by principals and selected teachers from participating schools. The EMD modules are targeted on principals. The INSET modules are organised into four sets of five modules for teachers in mathematics, science and technology, English and the foundation phase (JET, 2000b: 9-10). Modules are taught in a five-week cycle.

The programme design also provides for the benefits to be passed onto non-participating personnel or schools. Observers refer to the lessons being 'mainstreamed', i.e. incorporated into the mainstream of schooling practice in the province. The replication of improvements is to be achieved in three ways. Firstly, the programme adopts a 'cluster' model of participating schools.

'Project schools are grouped in clusters of five: these are generally within easy reach of each other. Training and support is delivered directly to each cluster, in one of the member schools. Three key

teachers from each school, together with the principal, members of the school governing body and district officials receive training. This team, in turn, is responsible for sharing the knowledge and skills with other members of staff, with in-school support from the training providers' (Taylor, 1998: 8).

This cluster model can be contrasted with the more traditional 'cascade' model, according to which training is initially provided to a small number of officials, who themselves train the next layer of personnel, who in turn train the next layer, and so on. Unfortunately, it seems that the benefits get diluted as they 'cascade' down through the layers (*ibid*). Further, participants in each cluster attend 'cluster meetings' in addition to training workshops. These meetings are intended as a forum for discussion of what had been learnt and of people's experiences in implementing this (JET, 2000b: 11). District-based development teams provide support.

Secondly, the programme is said to have been 'entrenched' in the provincial Department of Education. JET describe this as one of the 'groundbreaking innovations developed by *Imbewu*'! Whilst the programme has its own specialist management team, the provincial government has authority, and government officials have been seconded to the programme. Initially, much of the training and other work was done by NGOs; later, departmental personnel assumed some of these roles. The provincial department also receives technical assistance (JET, 2000b: 13; Perold, 1999: 8-10). Thirdly, *Imbewu* was designed with a systematic performance of monitoring and external evaluation, to ensure that the lessons can be learnt readily and applied elsewhere. This is discussed further below.

One hundred schools participated in the pilot phase, with a further 423 schools joining later. Most were in deep rural areas in former bantustans. A few were in townships around Cradock and Grahamstown. In total, schools in 28 districts participated.

The programme supposedly entails a particular philosophy of school change. Participants talk of 'being *Imbewu*-ised', referring to the programme's emphasis on attitudinal change, with transformation coming from within as much as without (see JET, 2000b: 6-8; Perold, 1999: 11-20). An integral component of the programme is what is called 'vision crafting', i.e. the initial deliberation among participants about their goals and objectives (see JET, 2000b: 10). The emphasis on an improved ethos is evident in the following description of the

‘*Imbewu* model for school development and transformation’, produced by the Eastern Cape Education Department. The ECED diagnoses that the problem is a lack of ‘spirit’ in schools:

‘Teachers complain of principals who are both incompetent and authoritarian. Principals complain of teachers who abandon their classes, who are alcoholics or who show little interest in the welfare of the school or their learners. Learners raise issues of harassment, sexual abuse, assault and a general lack of caring. Some teachers feel physically threatened by learners who are out of control. The litany is endless.

The key, then, is to determine how to rekindle the spirit in individuals and restore the essence that gives the school its life. The transformation strategy must find the means to ensure that the spirit of individuals and the uniqueness which they bring to the school are both transformed into positive action’ (ECED, 2000: 2-3).

This is an extreme version of an ‘inside-out’ approach (see Muller with Roberts, 2000), with little or no emphasis placed on external accountability or incentives. It is reported that some of the foreign ‘technical’ advisors were hostile to the formal assessment of pupils. Whether this kind of transformation from within is viable or leads to improved pupil achievement is, of course, an empirical question, and one that needs to be tested. It is unfortunate that a programme as large as *Imbewu* simply assumed that it was.

The experience with *Imbewu* has informed the design of subsequent programmes, most notably *Mahlahle* and the QLP. This is evident with respect to the methodology of evaluation, the content of the programme, and the link between these. First, baseline studies are integral to the evaluation. Two baseline studies were done for *Mahlahle*: the first, in late 1999, examined management, administration and governance at regional, district and school levels. The second, in mid-2000, examined teaching itself. Secondly, these baseline studies are being used to inform project design to a much greater extent (JET, 2000c).

Thirdly, the *Imbewu* experience affected the design of *Mahlahle* and the QLP directly. The original Business Plan for the QLP identified three elements in the systemic approach to be used for the QLP:

- ‘Building capacity and systems at the District and other levels of the provincial Departments of Education. This is an essential element in institutionalising any external intervention. Well-functioning district offices are key to sustaining support for schools once the project has come to an end, and to replicating the intervention in non-project schools;
- Improving the ways schools function. Building the capacity of individuals – such as INSET for teachers, or management training for principals – has been found to have no effect on improving learning if schools, the key institutions of delivery, are not functioning effectively.
- Intervening directly at the classroom level. Local and international experience indicates that changing the classroom behaviours of teachers and learners is the most intractable element of any school development programme, and that if this is not addressed directly then learning outcomes are unlikely to change. [Previous research] ... strongly suggests that the biggest problem at this level is the poor conceptual knowledge of teachers concerning the subjects they are responsible for’ (JET *et al.*, 1999: 8-9).

Both *Mahlahle* and the QLP comprised five ‘programmes’. These included, in both cases, programmes on ‘district development’, focused on management and administration at the district level, and ‘school development’, focused on management, administration and governance at the school level. Both also included a programme on monitoring and evaluation. *Mahlahle*’s other two programmes covered ‘resource provisioning’, i.e. the distribution and use of textbooks and other teaching materials, and ‘classroom support’, i.e. interventions in classroom practice. QLP’s other two programmes covered ‘learning outcomes and assessment practices’, i.e. the development of assessment practices to guide and monitor teaching and learning, and ‘teacher development and curriculum intervention’, concerning other aspects of classroom practice (JET, 2000d: 4; 2001).

The experience of *Imbewu* was also supposed to inform the design of what is sometimes known as ‘*Imbewu II*’, i.e. the second phase subsequent to the initial three-year programme. One identified lesson was the problem of a ‘one size fits all’ approach, i.e. an approach that assumes that the same package of services is appropriate in every participating school. This is a growing concern with the larger programmes. Other than this, however, the available documentation from

the ‘*Imbewu* Transitional Phase Operational Team’ seems to include little critical reflection on the experience of *Imbewu*’s first three years.

The Methodology of Evaluation: *Imbewu*

The value of *experimental* projects in school improvement depends on the quality of the *evaluation* of the experiment. The programme of evaluation for *Imbewu* serves to illustrate what needs to be done, and also what problems can be anticipated. The documentation available on the evaluation of *Imbewu* includes: baseline, midterm and final studies of classroom practice, school management and student performance (ESA, 1999, 2000a, 2001b); an assessment of the first set of *Imbewu* training modules (Moyo and Rembe, 2000); a series of studies of management at the provincial and district levels; two additional evaluations at approximately midterm (Crown Agents, 1999; Perold, 1999); and a further assessment done by Roberts for the JET (JET, 2000b), as well as studies of vision crafting and the secondment of provincial department officials. The evaluation of *Imbewu* required a separate and dedicated managerial structure, in the form of a sub-committee of the Project Steering Committee.

Imbewu’s objective was to improve student performance, especially in basic literacy and numeracy. The outcome of improved performance should be measurable through numeracy and literacy tests. *Imbewu* also has ‘intermediate’ objectives, which are less easily measured but are observable. It is as important to evaluate success in these as to measure student performance, because the evaluation of success in achieving intermediate objectives helps us to identify which factors or mechanisms serve to effect changes in student performance. These intermediate objectives include such things as classroom practice and school management.

The core of the evaluation process is the series of studies of the impact on *Imbewu* on participating schools. After a competitive process, the tender was awarded to Eric Schollar and Associates (ESA), who applied a methodology that they had already used in an evaluation of READ in the same area. The schools study comprised four main components:

- Tests of student performance, using separate literacy and numeracy instruments designed by ESA;
- Observation of lessons and other activities at school;

- Interviews with principals, teachers and members of school governing bodies;
- Scrutiny of records, for example school registration records and teachers' lesson plans.

There are two differences between these school studies and the Whole School Evaluation proposed by the national Department of Education. Firstly, the ESA studies included externally-run tests of student performance, which are absent in the Department's proposals. Secondly, the Department's proposals are seen as having a developmental purpose as well as a monitoring process, meaning that the experience of evaluation is explicitly designed to feed immediately into improved practices, whereas the ESA evaluation is purely a monitoring exercise.

The baseline study covered twenty selected schools that were participating in *Imbewu*, together with a control group of four non-participating schools. A total of 666 students sat the tests in the twenty participating schools, and a further 160 in the control group. Fifty-one teachers were interviewed, and their classes were observed. All twenty principals were also interviewed, and management systems and the curriculum in the twenty schools were described (ESA, 1999: 7-11). The midterm study, conducted twelve months later, comprised the same set of components. At the same time the same evaluation was carried out in a further twenty-three schools which had not participated in the first year of *Imbewu* and thus were at the baseline stage. The final evaluation covered twenty-four schools which had been participating in *Imbewu* for two years and twenty-five which had been participating for one year only. A total of 1 748 pupils were tested, and 125 teachers were interviewed and their classes observed (ESA, 2001b: 7-8).

The 'evaluation' process gathered a lot of 'baseline' data on the problems in the schooling system. This data, and data from the mid-term study, proved useful in redesigning aspects of the programme, and perhaps in reminding participants of its objectives. But the crucial data in terms of evaluating whether *Imbewu* 'works' – or, more precisely, how its various components have 'worked' separately and collectively – is the data on changes over time. Have students' performances in tests improved? Have classroom practices improved? Has school management improved? Has district and provincial management improved?

The evaluation reports indicate the many difficulties involved in evaluating a programme like *Imbewu*. Can one assess changes over such a short period of time? JET cautiously warned that 'because *Imbewu*'s energies are directed

across a number of levels in order to achieve systemic change, it is not expected that significant learning gains will be achieved by *Imbewu* during the first three years' (JET, 2000a: 3). In fact, for reasons that are not clear, the time gap between the baseline and midterm studies, and midterm and final studies, was only one year in each case, so that the total time-span was two rather than three years (ESA, 2001b: 5). Given that *Imbewu* could only be expected to have an effect on pupils' performance *after* teachers and managers had attended training and implemented the lessons in their daily practice, two years is a short period of time in which to expect much change in measured pupil performance.

Further problems arise from the limits of selective observation and interviewing. What can one tell from *observing* classes, given that teachers might put on a non-random class to impress the observer? As ESA point out in their midterm report:

'It is a commonplace that the presence of an observer affects, to one degree or another, the routine functioning of a school or classroom that he/she wishes to observe. Virtually all field researchers would assert, for example, that they are frequently aware that a lesson they are observing has been presented previously, and that pupils are often 'coached' beforehand' (ESA, 2000a: 12).

ESA report that when researchers visited schools unexpectedly, for whatever reason, teachers would usually say that they were 'not prepared for *Imbewu*' that day! Many of the teachers who had been forewarned 'happened' to present a class on 'Water is the source of life' which had been a model lesson in an *Imbewu* training module. Researchers therefore observed the same lesson being presented many times, and in different grades! The obvious conclusion is that, when there was forewarning of the presence of observers in class, the researchers observed very atypical lessons in terms of teacher (and possibly student) preparation, as well as teacher-student relations. As ESA note, it is at the very least unclear whether trained teachers are able 'to transfer what they learned about methodology through a sample lesson to the rest of their teaching' (*ibid*: 14).

Interviewing or other self-reported data is also problematic. How does one interpret what one is told by teachers or principals? ESA note the following:

'It has been our experience that principals and teachers routinely provide positive endorsement of all intervention projects. Perhaps the fact is that schools have received so little support over the years that

they fear support they are currently receiving will be lost if they are critical. Whatever the reason, we have experienced the phenomenon in so many different contexts and provinces, that we consider impact data derived from interviews alone to be unreliable. While, in other words, one would not expect interviewees to be critical of interventions which are genuinely effective, blanket approval in itself does not necessarily prove that the intervention has, in fact, been effective. Further, interviews which provide closed sets of possible responses to attitudinal questions provide too many specific cues to interviewees appearing to have a more detailed knowledge and understanding of an intervention than is actually the case' (*ibid*: 11).

Referring to previous work by ESA, Taylor has also warned about the uncritical use of self-reported data. Self-reported data can be influenced not only by obvious self-interest but also by 'excessive good faith: programme staff are idealistic and dedicated; participants have worked hard and are excited about the possibilities of the new methods solving what seem to be insurmountable problems; everyone wants the programme to succeed. Under such circumstances it is easy to confuse best intentions, and changes in the outer forms of teaching and learning, with qualitative improvements' (in JET, 1996: 52).

Two constraints on what can be learnt from the evaluation resulted from choices made by the programme managers. Optimally, the performance of a panel of students would be assessed over time, with the performance of participants in *Imbewu* classes being compared with that of a control group of non-participants. This, indeed, was the original intention. Instead of doing this, however, the evaluation was limited to testing the performance of successive cohorts passing through each grade. In other words, the performance of successive cohorts of students was taken as an indicator of the effectiveness of the teacher. In addition, there was no real control group (although schools that only joined after the first year of the project could be treated as a quasi-control group for the first year). The number of students assessed was also reduced markedly. As ESA (1999: 2) note: 'The eventual design inherently lacks much of the control over variables affecting teacher and pupil performance available to more robust designs and, consequently, questions of causality can only be answered at a much lower degree of confidence'. These changes were reportedly made for financial reasons (ESA, 1999: 6-7). It is unclear what the extra cost would have been, or precisely how and why it was decided to use funds in training activities rather than systematic monitoring.

The final evaluation of *Imbewu* by ESA was conducted in 2001. In their draft report, ESA emphasised once again the deficiencies in the evaluation process (ESA, 2001b: 25). ESA underscored especially the absence of any external control group of schools that did not participate in *Imbewu*. If, as the Department of Education itself claimed, the improved matric exam results in 2000 reflected an improvement in the quality of education, then one would expect similarly improved test scores in the lower grades also. The lack of any external control group rendered it impossible to distinguish between the impact of *Imbewu* specifically and general changes in the schooling system.

ESA also noted a series of problems with the data. For example, baseline test scores from one district were suspiciously high. In ESA's view, they were probably 'contaminated in one way or another'. The same schools produced much, much lower test scores in the midterm and final evaluations, giving the impression that *Imbewu* has led to 'a catastrophic decline' in pupils' performance! ESA sensibly, if somewhat arbitrarily, decided to omit the data from this district in its analysis (*ibid*: 25-6).

Improving the Quality of Evaluation

Work by JET indicates a number of lessons from past evaluations in South African education. JET's conference on 'Quality and Validity in INSET Evaluations' helped to define the need to assess evaluations critically. *Getting Learning Right* (Taylor and Vinjevoid, 1999) helped to sustain this concern. JET's current project (undertaken by Roberts) promises to take it forward.

Roberts provides a pithy summary of the weakness of many of the earlier evaluations:

'Many impact evaluations tend to focus on process variables and aspects of school culture and internal relations, rather than looking at whether the project has had an impact on the classroom (and if changed classroom practices are resulting in improved learner achievement). Many projects (and evaluations) focus on doing things differently, rather than the quality of what is done, especially at classroom level' (personal communication, April 2001).

JET's design of a framework for evaluating its *Mahlahle* project indicates some of the lessons that JET has learnt:

‘The [evaluation] combines quantitative and qualitative approaches, with learner performance testing being conducted in a larger sample of schools and in-depth qualitative studies being conducted in a smaller sample of these schools. The qualitative studies include interviews, a short questionnaire which solicits factual information on the schools (e.g. enrolment, staff size), classroom observation and an audit of learning support materials. Efforts are made to corroborate what is reported in interviews with observed behaviour, administrative and management documents and what is actually seen in the school. Information on teaching and learning practices is supported by lesson observation, discussions with teachers and reviews of learner workbooks, assessment records and work plans. The learner performance test which is administered is a standardised test which is analysed according to performance on each item.

Despite efforts to address many of the limitations, the methodology still has its flaws. One of these is the lack of sustained observations of teachers over a period of time – the project does not have the funds nor the personnel to allow researchers to visit schools for long blocks of time or to visit them several times a year. Each school is visited for a full day and an expert researcher observes at least four lessons in each school’ (Roberts, 2001: 3).

The evaluation process of *Imbewu* suggested a number of specific lessons. Overall, it is apparent that more care must be given to evaluation procedures and the necessary requisite funds must be provided in the project budget. Hitherto, too much has been done in a rush, with last minute planning, and with insufficient funding.

Two of the key weaknesses of the *Imbewu* evaluation were the inadequacy of the control group of schools and the absence of data on the progress made by fixed cohorts of pupils over time. The inclusion of a control group is important in order to distinguish between the effects of the intervention and other changes that might be occurring in schools generally. The testing of a cohort is important because it prevents the results of testing being affected by normal variations between the ability of successive years of students. Each of these issues has been addressed in the evaluation of some other educational interventions. Both controls and cohort-testing were included in the evaluation of the READ Umtata programme (READ, 1999), the READ Kei-Komga programme (ESA, 2000b)

and the GET-INSET Special Project in Atlantis-Malmesbury (although this tracked changes over only seven months) (Fredericks *et al.*, 2001). Cohort-testing without a control group was included in the ongoing evaluation of the Pilot EQUIP programme (ESA, 2001a).

A notable strength of the *Imbewu* evaluation was the close integration of quantitative and qualitative research methods. Careful observation of lessons and other activity at schools is crucial in assessing the extent to which training and materials are actually being put to use. Combining such observation with testing makes it possible to assess systematically different components of a programme as well as the overall impact of the programme as a whole. Optimally, the test results and qualitative research would be combined with information from the education departments' EMIS database.

Perhaps the best design for an evaluation is the one for the READ/Business Trust 'Learning for Living' project (ESA, 2000c). Only the baseline study has been completed yet, but the design of the prospective follow-up studies over the five-year life-span of the project is very impressive. The evaluation seeks to answer four key questions:

- Has the project been delivered as designed?
- Have the practice of teaching and running a school changed in the project schools to the extent and in the ways predicted?
- Has pupil performance in tests improved as predicted?
- Can these hanging practices and improved results be ascribed to the project?

Tests will be administered to a sample of pupils in each of four grades at each of ninety schools (fifty-four project schools and thirty-six control schools). In total, 7 200 pupils will write the tests each year for five years. The baseline study was conducted in 2000 (*ibid*), with the first retest at the end of 2001. With a large sample being tested across a long period of time, this study promises to provide very valuable information of what affects student performance and how it can be improved.

There are signs that lessons have been learnt from the experience of evaluating programmes such as *Imbewu*. The project documentation for *Mahlahle*, the QLP and the READ Learning for Living projects all emphasise the importance of evaluation and monitoring, including thorough baseline studies (in part to help to finalise the design of the projects themselves) and regular follow-up. In the QLP, for example, evaluation includes not only qualitative research and pupil

testing, but also the collection of data on the schools, community and students' family background.

It is not clear, however, who is the intended audience for evaluations such as these (see also Taylor, in JET, 1996: 51). In part, I imagine, it is the donor. In large part it is certainly the project managers, who look to 'evaluations' for indications of how to modify their projects' design during the course of the projects; this is, in other words, a diagnostic function. They may serve a role as a 'consultation' mechanism, and may also serve to jolt actors into a recognition of the nature or scale of the problem. But it is not clear that the evaluations are intended to help identify lessons for the much larger task of reforming schools countrywide.

The process of evaluation is shaped, if not driven, by the particular concerns of the donors. Most donors have their own agenda. Foreign donors (such as DfID) have to account for their use of funds. Sceptics suggest, with good reason, that such donors are more interested in ensuring that funds were spent on the correct budget items than they are in achieving the best possible results in terms of improved pupil performance. Given the role of political considerations in donor generosity and officials' more immediate self-interest in ensuring the continued flow of funds, some donors might want to avoid overly-thorough evaluations lest they generate embarrassing results. It is difficult to avoid some cynicism about DfID given that it committed to the extension (and enlargement) of *Imbewu* before it had received the final evaluation of the first three-years of the programme.

Of the donors active in South Africa, the Business Trust seems to be the most thoroughly concerned with evaluations and efficiency. One of the points raised by the Business Trust Review Panel on the draft QLP business plan was the need for an evaluation process 'that will ensure the spread of lessons to the rest of the country and support the replication of the project' (JET *et al.*, 1999: 6). The response of the QLP team was that the correlation of planned assessments of student performance with planned situation analyses at district, school and classroom levels would make clear what works and what does not work, and thus 'contribute to the derivation of a model for improving the nation's schools' (*ibid*: 24). About R1 million per year was allocated to evaluation and monitoring; this amounted to just 3.8 percent of the overall budget. This is a great improvement on (for example) *Imbewu*. It is important, however, to ensure that the concerns of the donors (in the case of QLP) are translated into a thorough evaluation and monitoring process. It is somewhat worrying that the

QLP business plan seemed to substitute soothing sentences for critical analysis (see eg *ibid*: 37), and that the appointment of consultants to conduct the evaluation may reflect considerations other than maximising quality.

What Evaluations Tell Us: *Imbewu*

The baseline study of *Imbewu* found that school-level management and leadership were poor: basic records were not kept, there was little planning, and there was inadequate provisioning of textbooks. The study found that many students had language difficulties in the classroom, teaching practices were poor, and that teachers seemed to have inadequate knowledge of their subjects. Very few teachers provided feedback on students' written work or administered regular tests. The study found that physical factors (including the condition of school buildings) were much less important than the extent of commitment among principals, teachers, students and parents.

What changes were observed, measured or at least claimed in the midterm and final studies? The ultimate objective was, of course, to improve student literacy and numeracy. The numeracy and literacy tests used to evaluate student performance showed 'encouraging' but nonetheless weak indications of a slight improvement in student performance between the baseline and midterm studies. The sample of students in grades 5 and 7 who were tested was sufficient to have some confidence in the results. Among grade 5 students there was only a negligible improvement in literacy and numeracy test scores, with the mean test scores rising from 21.1 percent to 22.4 percent and from 32.8 percent to 34.6 percent respectively. Among the grade 7 students, however, there was a significant improvement in both literacy and numeracy tests, with mean test scores rising from 33.8 percent to 39 percent and 46.6 percent to 51.9 percent. Students in five districts were tested. Not all the changes were in the same direction: in one of the five districts where pupils were evaluated in both baseline and midterm studies, test scores declined in both literacy and numeracy tests for both grades 4 and 6; in a second district, test scores declined in two of these four categories; but in the other three districts, all mean test scores rose. These were small changes, but they were over a short period of time. On the other hand, considering that they were rising off a very low base, surely larger gains might have been expected?

If the effects on the ultimate goal of student performance were ambiguous, there was more evidence of progress with respect to some of the intermediate

objectives. In view of the ESA midterm study and the earlier study by Perold, the JET put a cautiously positive spin on the findings:

‘There is no question that *Imbewu* is succeeding in providing the prerequisite conditions for improving pupil performance in project schools. ... There is also evidence that capacity is being built at the district level for replicating the project beyond the 523 target schools, and for sustaining these gains after completion of [the first three year phase of] *Imbewu*. ... These are remarkable achievements indeed after only two years...’ (JET, 2000a: 9-10).

The main change, it is claimed, is in the atmosphere or ethos prevailing in schools (JET, 2000b: 14-15). Perold (1999) highlights quotations such as the following: ‘*Imbewu* has brought a renaissance, a rebirth in our schools’ (District Manager); ‘*Imbewu* transforms our teachers, especially their attitudes’ (Principal); ‘We were inspired by *Imbewu*’ (Teacher).

The first activities and modules in the *Imbewu* programme certainly aimed at changing attitudes: the objective of ‘vision crafting’ was to change the attitudes of teachers and principals and create the desire to improve schools (JET, 2000b: 13). Roberts reports that some of the personnel involved in designing and implementing *Imbewu* stressed the change in ‘atmosphere or ethos in participating schools (e.g. changed spirit, having a vision, being able to take initiative and having pride and ownership of schools)’ (*ibid*: 14). Perold provides anecdotal evidence of improvements in some participating schools (Perold, 1999: 31-6). But, as she asks, does enthusiasm deliver results? The government’s policy of redeploying teachers serves to undermine gains (*ibid*: 37).

Enthusiasm might be widespread but, as Roberts found, the precise objectives of the programme were understood less and less well as one moved away from the core project personnel (JET, 2000b: 5). Thus project personnel described the programme in terms of its stated goal of systemic transformation, whilst participating principals and teachers tended ‘to reduce the project’s objectives to bite-sized components, often corresponding with the interviewee’s personal experience of a particular component of the project or with the development of specific skills’ (*ibid*: 6; see also 14).

There are said to have been problems with the delivery of training. The use of district officials from the Department of Education caused some problems, as they were over-stretched and lacked capacity (*ibid*: 10).

Roberts, in an assessment based on a small number of interviews with principals and teachers, found that the main reported benefits concerned improved management and administration in schools (*ibid*: 14-15). Parents were also said to have become more involved in school governance. But ESA found that there was only a tiny improvement in record-keeping and administration at the school level (ESA, 2000a: 39-40).

At the same time, JET acknowledges shortcomings, especially with respect to management within the provincial education department. *Imbewu* has invested considerable effort in assisting the Department of Education to improve its operations (Perold, 1999: 26-8). Although the evaluation of this is still largely impressionistic (see above), ‘it would appear that *Imbewu* have been very active in the areas of strategy and policy, but has had minimal effect in building more effective management practices...’ Management systems remain ‘hopelessly inadequate’.

‘At the district level it would appear that the influence of *Imbewu* is confined to involving individual officials in the training and support programme. While this is undoubtedly a key element in linking the work of districts and schools in support of improved teaching and learning, more thoroughgoing and lasting change must be dependent on the strengthening of district offices, including the building of sound management practices. Similarly, at the school level, while *Imbewu* is having a profound influence on the morale and motivation of the entire community, impact on institutional management appears to be piecemeal rather than systematic’ (JET, 2000a: 10).

JET concluded that

‘The first and most important problem that the second phase of the [*Imbewu*] project needs to get to grips with is the very poor state of management which exists at all levels of the EC DoE [Eastern Cape Department of Education]: province, district and school. Until an entry point can be found for establishing systems for planning, regulating and monitoring workflows, and for building the capacity to do this,

schooling will continue to limp along, with projects like *Imbewu* leaving, at best, a transitory legacy' (*ibid*:11).

The problems within the Department of Education combined with disappointingly weak links between *Imbewu* and core departmental work meant that the project has not achieved the sustainability hoped for at the outset (JET, 2000b: 19).

JET also acknowledged that changing practices in the classroom 'are largely in the area of form rather than substance' (JET, 2000a: 11). An external evaluation found that *Imbewu* had brought about more changes in school management than in classroom practice. But improved management was largely confined to administration, rather than quality assurance. There continued to be weak monitoring of quality in the classroom. The ESA found, in their midterm study, that:

'The majority of principals continued to report that teachers will accept a check on their presence in the classroom, and their adherence to the formal work plan, but they will not accept quality assessment and feedback. This may also explain why no schools reported classroom visits, or quality monitoring, by the DoE either. The principals and teachers continue to assert that the new developmental teacher appraisal system will obviate the problem.

It should be noted that, during the course of research, there were frequent instances of teachers stopping to take a cell phone call, pupils wandering around during class time, remaining outside after break for long periods and leaving classes without reference to the teacher. None of these indicate a high level of managerial supervision of the whole school – if it is so obvious to a researcher, surely it must also be obvious to the principal, at least?' (ESA, 2000a: 40-1).

Many classes were left unsupervised by teachers, who congregated in a separate room. There even appears to be a problem with absenteeism. Within the classroom, however, there does appear to have been a slight improvement in the practice of teaching, with (for example) more planning and more group work. But the improvements are marginal, and uneven. The lessons observed and reported by ESA include numerous terrifyingly bad examples.

The evaluations indicated clearly the need to improve teachers' knowledge of the content of their subjects. According to the ESA: 'The understanding of many teachers of the methodologies encouraged by *Imbewu* remains simplistic and, in some cases, an increase of disordered and undirected activity was observed in classrooms as a result of their greater use'. Moreover: 'Teachers tend to overemphasise the methodological components of their lessons to the detriment of the content they are presenting. When they are, in addition, not fully in control of these methodologies, the level of content presented in lessons often appears trivial indeed' (*ibid*: 5). Perold makes a similar finding (1999: 49).

The final evaluation of *Imbewu* provides further dispiriting evidence of the lack – or at least slow pace – of change. The study found no or weak evidence of improved pupil performance in the tests. The test results might suggest very minor improvements in literacy test results in the younger grades, but there was no improvement whatsoever in numeracy test scores. Moreover, pupils continued to use woefully inadequate methods for solving arithmetic problems. Over the two-year period, as we saw above, schools in Phase 1 showed an improvement of just 4.2 percentage points in the grade 5 literacy test, 0.9 percentage points in the grade 5 numeracy test, and 2.0 percentage points in the grade 7 literacy test. The scores in the grade 7 numeracy test declined by 0.4 percentage points. With the exception of the grade 5 literacy test results, the changes are insignificant in comparison with the error factor, and even the grade 5 literacy test improvement is very small. Moreover, a comparison of the baseline, midterm and final evaluation test scores shows that there were no clear trends. Indeed, overall test scores had declined between the mid-term and final study, suggesting that participation in *Imbewu* might have had a temporarily positive but unsustainable effect (perhaps related to the conduct of the evaluation, i.e. what sociologists call a 'Hawthorne effect'). The Phase 2 schools also showed insignificant changes. Trends in numeracy tests were, however, also worse than trends in literacy tests.

ESA suggest that the lack of impact on measured numeracy performance might reflect the severity of prior inadequacies in numeracy teaching (2001b: 27). In other words, numeracy teaching is so bad in many schools that even a well-designed intervention will have little or no effect in the short-term (or perhaps even the medium- or longer term). The strongest evidence for this came from a close study of the notes that pupils made in solving arithmetic problems. Even in classes where teachers had benefited from *Imbewu* training, pupils continued to tackle multiplication and division problems by altering them into addition and subtraction problems. ESA commented:

‘Since these methods (the baseline report referred to them as a lack-of-a-method) realistically amount to nothing more than counting on one’s fingers, and have not altered at all since the initiation of the project, there is really no surprise in the fact that pupils demonstrate no impact in numeracy of any kind. Indeed, I would say that the average child was fundamentally innumerate before the project started and remained fundamentally innumerate after it ended.’ (*ibid*: 29)

The mean test scores by grade are set out in Table 2. Note that the Phase 1 participants had received two years of *Imbewu* activity, but the Phase 2 participants just one year. (The suspect results for Grade 5 in one district are omitted, as noted above). This data is not for fixed cohorts of students, but rather for successive cohorts of students in each of the two grades. The results are not encouraging.

Table 2: Mean test scores by grade, Imbewu

Grade	Phase	Literacy		Numeracy	
		Pre	Post	Pre	Post
5	Phase 1	19.7	23.8	31.7	32.7
	Phase 2	14.8	14.1	27.5	26.4
7	Phase 1	33.8	35.8	46.6	46.2
	Phase 2	22.6	23.3	41.8	39.5

Source: ESA, 2001b. The G5 Phase 1 scores are calculated excluding the deviant district.

Their observation of lessons in practice served broadly to confirm some of the misgivings already noted in the mid-term report. ESA found that there were clear changes in teaching methods and practices within the classroom but it is far from clear that the quality of teaching has really improved. It seems that many teachers have learned to adopt the new and ‘approved’ teaching methods but ‘without becoming observably more proficient in the essential skill of choosing and controlling methodologies appropriate to topic, class, and content’ (*ibid*: 40). Too few classes passed beyond trivial and superficial levels, too many classes were clearly replicating a ‘model’ lesson learnt in a *Imbewu* training session, worryingly many teachers seemed to have an inadequate grasp of the subject matter, and some appeared to struggle to teach in English. There was some evidence that pupils spoke more often in class than two years previously, but there was no change in the time spent reading or writing. Pupils spent more time in group work during class, but it was unclear whether this time was spent

effectively. ESA report that in only 20 percent of the lessons observed did pupils ask a single question. One of the few clear improvements was in the availability of textbooks, with a sharp drop in the proportion of classes where only the teacher had a textbook.

Researchers also found many schools where too little teaching was taking place. ‘All of the researchers commented on the number of pupils who regularly wander around schools at all times, teachers absent from classes while pupils sat unattended, and those who presented cursory lessons when a visitor was not expected’ (*ibid*: 38). This reflects poorly on school management, as well as the teachers concerned. ‘Those schools in which ordered management and teaching/learning took place on a routine basis were almost all the schools in which such practices were already evident at the time of the baseline study’ (*ibid*: 38).

ESA also collated for the final evaluation a mass of information on what principals, teachers and so on said about the effects of *Imbewu*. Participants were almost universally positive but were typically vague as to what precisely had changed for the better – just as ESA had found in the mid-term evaluation. When pressed, principals claimed most often that teachers were more motivated, more positive, and combined better as a team. Teachers mentioned most often improved teaching methods (*ibid*: 32-4). Principals reported that *Imbewu* had improved their financial and other managerial skills. There was said to be more community and parental participation in the school, but this had not prevented vandalism of the schools concerned.

The final ESA report has only just been completed in draft form, and is yet to be finalised. It is unclear whether there will be some kind of composite assessment of *Imbewu*. Given that DfID have already committed massive resources to the extension of the programme, regardless (it seems) of the assessment to date, it seems unlikely that there will be a strong impetus for such a composite assessment from the donor’s side! The provincial education department also has mixed incentives to conduct a thorough-going, composite assessment.

The documentation prepared for and by JET is remarkable in its candour, and deserves considerable praise. But it neglects entirely the question of cost. There is no mention, and apparently no evaluation, of the costs of different reforms.

Comparing the Evaluations of Different Programmes

Imbewu has been scrutinised more closely than most programmes, but it is not unique. A number of other programmes have been evaluated in comparable (and sometimes, as we have seen, better) ways. These include a number evaluated by ESA, whose work generally is of the highest quality. Given that the goal of treating educational interventions as experiments is to learn what works and what does not, what have we learnt from interventions to date?

Two READ programmes have been evaluated thoroughly: the READ programme in Umtata district over the period 1995-98 and the programme in Kei-Komga district over 1997-99. In both cases tests of pupil performance indicated sharp improvements in reading and writing skills across both short and longer periods of time. Table 3 sets out the improvement in test scores of fixed cohorts of students across just one year, comparing pupils in schools participating in the READ programme with pupils in a control group of non-participating but otherwise similar schools in the area. The first cohort of pupils tested was in Grade 5 in 1997 and Grade 6 in 1998. The second cohort was in Grade 6 in 1997 and Grade 7 in 1998.

These test results show that READ achieved significant and consistent improvements in reading and writing. Table 4 summarises the difference between the improvement for pupils in READ schools and pupils in the control group. Given the low cost, at just R53 per pupil per year, this seems to be a remarkably effective intervention.

Table 3: Improvement in test scores for fixed cohorts: READ Umtata project

<i>Cohort</i>		<i>Improvement in reading test score</i>	<i>Improvement in writing test score</i>
I (G5 to G6)	Project	+ 27.4	+ 36.3
	Control	+ 2.9	+ 23.2
II (G6 to G7)	Project	+ 31.6	+ 41.1
	Control	+ 1.0	+ 23.0

Source: READ, 1999: 20

Table 4: Difference in improvement in test scores, project compared to control cohorts: READ Umtata project

<i>Cohort</i>	<i>Difference in improvement in reading test score</i>	<i>Difference in improvement in writing test score</i>
I (G5 to G6)	+ 24.5	+ 13.3
II (G6 to G7)	+ 30.0	+ 18.0

Source: READ, 1999: 21

The evaluation of the READ programme in the Kei-Komga area entailed testing of four cohorts of students over a period of between two and three years. The baseline study was conducted early in 1997, and the final study at the end of 1999. The first cohort was in grade 3 at the beginning and had completed grade 5 at the end; the other cohorts were in grades 5/7, 7/9 and 9/11 respectively. The differences between changes in test scores for each cohort of pupils in READ schools and pupils in the control group are set out in Table 5 below. The positive scores indicate that pupils in READ schools were improving considerable faster than those in the control group (with the exception of the third cohort's reading test scores, but these may reflect an influx of pupils into grade 9 who had not been in READ primary schools; this interpretation is supported by the fact that successive cohorts of grade 9 pupils in the READ schools improved their reading test scores relative to successive cohorts in the control group – ESA, 2000a: 21). By the end of the programme, pupils in READ schools were performing at about the same level in the reading and writing tests as pupils *two grades higher* in non-READ schools.

Table 5: Difference in improvement in test scores, project compared to control cohorts: READ Kei-Komga project

<i>Cohort</i>	<i>Difference in improvement in reading test score</i>	<i>Difference in improvement in writing test score</i>
I (G3 to G5)	+ 8.9	+ 17.0
II (G5 to G7)	+ 23.1	+ 17.2
III (G7 to G9)	- 0.1	+ 15.9
IV (G9 to G11)	+ 4.8	+ 20.7

Source: ESA, 2000c: 21, under para 3.3.1

It must be noted that the ‘cohort’ being tested is not a pure cohort in every case. Optimally, the cohort would comprise a fixed panel of pupils, who are retested at regular intervals regardless of whether they are promoted into successive grades or required to repeat. Assessing a fixed panel of pupils requires that firstly, pupils’ names are recorded, secondly there is low rates of movement between schools or out of school, and thirdly, that the listed pupils attend school on the days that retests are conducted. (In a perfect evaluation, pupils should be tracked and retested even if they move schools or drop out of school altogether, but that would require considerable extra cost.) Attrition out of school means that an initial panel of (say) twenty-five pupils per class might be reduced to fifteen pupils per class two or three years later. A comparison of the test results then requires careful attention to the character of attrition, i.e. who is falling out of the panel and why. These problems are small, however, compared to the problems of analysing data if a pure panel is not used. In some evaluations the names of pupils were not recorded in the baseline study. So a ‘retest’ of a class one year later would probably only catch those pupils in the original panel who were promoted to the higher grade, excluding students who were held back and including for the first time students that were in the higher grade the previous year but were held back in it.

An evaluation of the impact of EQUIP at mid-term, i.e. after one year, indicated that cohorts of pupils in project schools had made absolute progress, but that only one of the two cohorts had made progress relative to the control group. Table 6 sets out the improvement in mean test scores for pupils in the project schools and the control group, whilst Table 7 summarises the difference.

Table 6: Improvement in test scores for fixed cohorts: EQUIP

<i>Cohort</i>		<i>Improvement in literacy test score</i>	<i>Improvement in numeracy test score</i>
I (G4 to G5)	Project	+ 10.5	+ 13.2
	Control	+ 12.4	+ 14.4
II (G5 to G6)	Project	+ 13.2	+ 13.1
	Control	+ 6.6	+ 7.8

Source: EQUIP, 2000 workshop overhead

Table 7: Difference in improvement in test scores, project compared to control cohorts: EQUIP

<i>Cohort</i>	<i>Difference in improvement in literacy test score</i>	<i>Difference in improvement in numeracy test score</i>
I (G4 to G5)	- 1.9	- 1.2
II (G5 to G6)	+ 6.6	+ 5.3

Source: EQUIP, 2000 workshop overhead

It would be wise to avoid reading too much into these results, and to wait for the final project evaluation to assess trends in performance over a longer, two-year period. But at first glance, the literacy test score results are very disappointing, even if the numeracy test results are more promising.

The evaluation of the first year of GET-INSET's special project in the Atlantis and Malmesbury area of the Western Cape was on a smaller scale. Pupils in grades 4 to 6 in five schools that participated in the project were tested in April and November of the same year, and the results compared to a control group of pupils in two neighbouring schools that were not part of the project. The small samples may explain why the results were very mixed and it is hard to discern any overall pattern. The absence of any pattern, however, is disappointing, since it suggests that there is no discernable improvement in the performance of pupils in schools that participated in the project relative to pupils in the control group.

Of the programmes discussed above, *Imbewu* and GET-INSET appear to have disappointing outcomes in terms of measured pupil performance, whilst READ's programmes result in a dramatic improvement; EQUIP's mid-term results are somewhat mixed. These findings need to be treated with some caution, however. Firstly, the evaluations are not strictly comparable, given that

different test instruments were used. Secondly, careful attention needs to be paid to details of the studies: How large was each sample of pupils? Was the control group really a control group? Were cohorts pure or pseudo-cohorts? And so on. Thirdly, and perhaps most importantly of all, some programmes might be expected to have an impact over a longer time period than others. READ interventions are heavily geared to what happens inside the classroom. Improved teacher training and teaching materials might translate into sharp improvements in pupils' performance in writing and reading tests. *Imbewu*, however, is focused in significant part on school management also. Improved school management might translate into improved pupil performance very slowly. This is not an argument for disregarding the results of an evaluation, but rather for designing and interrogating them more carefully, looking for discernable relationships between (say) observed improvements in management and measured pupil performance.

If there is a time-lag between the intervention and measurable results in terms of improved pupil performance, are there perhaps indicators which could be used in the meantime that would enable us to predict the subsequent improvement in pupils' performance? Several programmes have sought to develop sets of indicators of performance with respect to the key mechanisms by which the intervention is expected to affect (eventually) pupils' outcomes. EQUIP's set comprises twenty indicators in seven 'performance areas'. Unfortunately, the only way of telling what indicators are really 'useful' in that they do predict future changes in pupils' performance is to conduct a series of experiments and demonstrate that there is a clear relationship. Well-designed evaluations should provide information not only on what works, but also on what indicators help to reveal what is working.

It is possible that pupils will sometimes not benefit from an intervention however long the time period over which the intervention—and the evaluation—run. In reflecting on the slow improvement in numeracy test scores, Schollar wonders whether some pupils are so disadvantaged by the inadequacy of their initial education that they can never make significant progress thereafter. This would be a worrying finding. If true, it would require efforts to be directed to the very first stages of primary or even pre-primary education. Experimental interventions could help to indicate whether or not this is true.

Designing Evaluations and Interventions to Assist in Policy-Making

Crouch (1999: 1) has suggested that running an education system without indicators is like flying a jumbo jet without an altimeter or a compass or any of the other gauges that provide essential information! The analogy should not be overdone, however. The field of education is unlike flying a jet plane in that we lack a basic understanding of the mechanics of education. Pilots know how an aeroplane works, and know *how to respond* when they receive certain data from one of their gauges. If (I imagine) the gauges reveal that the engine speed is dropping, the pilot knows that the solution is to step on the gas. In education, acquiring precise data on a whole lot of variables is only one step in the right direction, as we need to know how the different variables are (or can be) linked. We don't know which control to push even if our gauges are working well. Will pupils' test scores improve more if we pump extra gas into post provisioning, if we turn up instead the supply of textbooks, or we adjust the flaps of management training? In order to learn how the education machine works, and how it might be improved, we have to study it in operation, and especially study the effects of experimental interventions in management, textbook supply or whatever. The effects of these different interventions need to be compared, critically, and with an eye on their cost.

Several years ago Jansen advocated clearly that comparative evaluations be done:

'[I propose] that comparative evaluation studies be designed and conducted with the deliberate goal of examining the value of rival programmes not with the goal of discarding the less successful (although this might be necessary in an extreme case) but with the aim of transferring lessons learned across projects. This means that comparison is built into the design (and the selection of comparable cases) from the start rather than in retrospective studies on set designs' (Jansen, in JET, 1996: 18).

This is little evidence that this call has been heeded.

The comparison of different programmes or interventions would, of course, be much easier if they had employed broadly standardised test instruments. The changes in READ schools could be compared directly with the changes in

Imbewu schools and the changes in GET-INSET schools, and so on. If information was also available on the costs of the interventions, then policy-makers would have the information needed to make choices as to what kinds of interventions achieve the best results at the least cost.

This is where the provincial and national departments of education can and should be playing a far more active role. Only the departments have the interest and capacity to assess all of the various interventions in their jurisdiction that are working with similar groups of students. The way to do this assessment is surely to organise a standardised assessment of cohorts of students in schools that are participating in each of the many programmes (as well as control groups of students who are participating in none). None of the existing forms of evaluation and assessment will serve precisely this purpose. What is needed is focused external testing of cohorts of students, linked to careful observation and interviewing, on an annual basis. Few – if any – of the provincial education departments are organised in ways that facilitate this kind of research.

A set of experimental interventions in the Philippines in the early 1990s shows what can be done. The Dropout Intervention Programme entailed four experimental interventions, randomly assigned to primary schools in selected low-income areas. Data was collected on dropout rates and pupils' performance in from the schools prior to and following the intervention, as well as from a control group of schools over the same time period. The interventions focused on a school feeding scheme on its own, the same scheme combined with parent-teacher partnerships, the provision of multi-level learning materials on its own, and the same scheme combined with parent-teacher partnerships. The experiment was flawed (with inadequate testing of pupils' performance, especially) and the actual findings are irrelevant. But the experiment indicates that it is possible to run routine systems for evaluating alternative policy options (Tan *et al.*, 1999).

Huge sums are invested in education in South Africa, but tiny sums are invested in learning the lessons of this investment. Even when evaluations are done, they are done in an uncoordinated way, and little effort is put into learning general lessons from specific evaluations. South Africa is currently making an astonishingly *uninformed* investment in education. This is not only a possibly colossal and absurd waste, but is also a huge missed opportunity. If we get the investment wrong, it costs the taxpayer a lot of money and children better opportunities in life.

References

- Case, A. & A. Deaton. 1999. 'School Inputs and Educational Outcomes in South Africa'. *Quarterly Journal of Economics* 458 (August), 1047-84.
- Crouch, L. 1999. 'Education indicators in South Africa: Who needs them? What for? Some Theoretical and Practical Considerations'. mimeo (April).
- and T. Mabogoane. 2001. 'No magic bullets, just tracer bullets: The role of learning resources, social advantage, and education management in improving the performance of South African schools'. *Social Dynamics* 27 (1), 60-78.
- Crown Agents 1999. *Eastern Cape Primary School Improvement Project: Imbewu Project*. Mid-Term Evaluation Report for UK Department for International Development. Sutton, UK: Crown Agents, Institutional Development Group.
- Department of Education. 2000a. *Education for All: The South African Assessment Report*. Pretoria: Department of Education.
- , 2000b. *Projects Audit in Provinces*. National Centre for Curriculum Research and Development (July). Pretoria: Department of Education.
- ECED (Eastern Cape Education Department). 2001. 'Celebrating the Success: The Seeds are Planted... The Growth Continues!' Report on Imbewu Project.
- ESA (Eric Schollar and Associates). 1998. *Evaluation of the pilot EQUIP project of the National Business Initiative*. Unpublished report, Eric Schollar and Associates, Johannesburg (October).
- , 1999. *First report of the Imbewu project evaluation: Baseline study*. Unpublished report, Eric Schollar and Associates, Johannesburg (July).
- , 2000a. *Imbewu midterm report*. Unpublished report, Eric Schollar and Associates, Johannesburg.

- . 2000b. *Baseline study of the evaluation of the READ/Business Trust project*. Unpublished report, Eric Schollar and Associates, Johannesburg (18 December).
- . 2000c. *Final report of the evaluation of the READ/Royal Netherlands Embassy project in the Border region of the Eastern Cape Province*. Unpublished report, Eric Schollar and Associates, Johannesburg (February).
- . 2001a. *The evaluation of the pilot EQUIP project of the National Business Initiative: Second report*. Unpublished report, Eric Schollar and Associates, Johannesburg (31 March).
- . 2001b. *Final Report of the Evaluation of the Imbewu Project: First Draft*. Unpublished report, Eric Schollar and Associates, Johannesburg (4 July).
- Fleisch, B. 2001. *Project Profile: Education Action Zones*. Report for JET.
- Fredericks, T., Powell, G. & A. Schaffer. 2001. *An Empirical Study of the First Year of GET-INSET's Atlantis-Malmesbury Special Project*. Report for ... (14 February).
- Hanushek, E. *et al.* 1994. *Making Schools Work: Improving Performance and Controlling Costs*. Washington DC: the Brookings Institute.
- JET (Joint Education Trust). 1996. *Quality and Validity in INSET Evaluations*. Braamfontein: JET.
- JET. 1999. *Mahlahle: Report on the Phase 1 Baseline Study*. Braamfontein: JET.
- JET *et al.* 1999. *Quality Learning Project: Business Plan*. Plan submitted to the Business Trust (5 November 1999).
- . 2000a. 'Imbewu Evaluation Overview'. Summary document drafted by Jennifer Roberts for JET.
- . 2000b. 'Project Profile: Imbewu'. Report drafted by Jennifer Roberts for JET survey of school improvement projects.

- , 2000c. *Mahlahle: Report on Phase 2 Baseline Study*. Braamfontein: JET.
- , 2000d. 'Mahlahle Project'. Brochure. Braamfontein: JET.
- , 2001. 'QLP'. Information package. Braamfontein: JET.
- Moyo, G. and S. Rembe. 2000. *Analysis of the Delivery of Module 1 [of the Imbewu School Improvement Project]: Draft Report*. Eastern Cape Education Department.
- Muller, J. 1999. 'Reason, Reality and Public Trust: The Case for Educational Research for Policy'. In Taylor, N. & P. Vinjevold (eds) *Getting Learning Right*. Johannesburg: Joint Education Trust.
- Muller, J. with J. Roberts. 2000. 'The Sound and Fury of International School Reform: A Critical Review'. Unpublished report for JET.
- NBI (National Business Institute). 2000. *Annual Report 1999-2000*. Johannesburg: NBI.
- NBI. Undated. 'Education Quality Improvement Partnerships: EQUIP'. Information brochure.
- Perold, H. 1999. 'Reflections on the Imbewu Project'. Report commissioned by JET.
- Porteus, K. 2000. 'The Personal and the Global: Grinding Gears? Reconsidering Efficiency Debates in General Education'. *Quarterly Review of Education and Training in South Africa* 7 (3).
- QLP. 2001. *Quality Learning Project Annual Report Year 2000*.
- READ. 1996. *A Survey Report on the Reading and Writing Skills of Pupils Participating in READ Programmes*. Johannesburg: READ Educational Trust.
- , 1998. *No South African Pupil Should Fail Matric*. Johannesburg: READ Educational Trust.

- 1999. *READ Delivers: A Successful Schools-Based Reading and Writing Programme*. Johannesburg: READ Educational Trust.
- 2001. Website pages downloaded from www.read.org.za (May 2001).
- Roberts, J. 2001. 'Project evaluations – some limitations and challenges'. Draft notes.
- Seekings, J. 2002. *Indicators of Performance in South Africa's Public School System*. Centre for Social Science Working Paper no. 10, University of Cape Town, CSSR.
- Tan, J., Lane, J. & G. Lassibille (1999) 'Student Outcomes in Philippine Elementary Schools: An Evaluation of Four Experiments'. *World Bank Economic Review* 13 (3), 493-508.
- Taylor, N. 1998. 'Imbewu: The Seeds of Learning'. In *JET Bulletin* 9, 7-9 (Johannesburg: Joint Education Trust).
- 2001. 'School Reform: How Should Government, Donors and NGOs Relate?' Unpublished text of a presentation.
- & P. Vinjevold (eds). 1999. *Getting Learning Right: Report of the President's Education Initiative Research Project*. Johannesburg: Joint Education Trust.

The Centre for Social Science Research

Working Paper Series

RECENT TITLES

- 16/02 *Popular Attitudes towards the South African Electoral System* By R. Southall & R. Mattes
- 17/02 *Are Urban Black Families Nuclear? A Comparative Study of Black and White South African Family Norms* By M. Russell
- 18/02 *AIDS and Human Security in Southern Africa* By N. Nattrass
- 19/02 *Public Works as a Response to Labour Market Failure in South Africa* By A. McCord
- 20/02 *Race, Inequality and Urbanisation in the Johannesburg Region, 1946-1996* By O. Crankshaw & S. Parnell
- 21/02 *The "Status" of Giving in South Africa: An Empirical Investigation into the Behaviour and Attitudes of South Africans towards Redistribution* By C. Pengelly
- 22/02 *How Important is Education for Getting Ahead in South Africa?* By M. Keswell & L. Poswell
- 23/02 *Missing Links? An Examination of the Contribution made by Social Surveys to our Understanding of Child Well-Being in South Africa* By R. Bray
- 24/02 *Unemployment and Distributive Justice in South Africa: Some Inconclusive Evidence from Cape Town* By J. Seekings
- 25/02 *Comparing Alternative Measures Of Household Income: Evidence From The Khayelitsha/Mitchell's Plain Survey* By J. Skordis & M. Welch
- 26/02 *The Employment of Domestic Workers by Black Urban Households* By M. Russell
- 27/02 *Poverty, Survival and Democracy in Southern Africa* By R. Mattes, M. Bratton & Y.D Davids
- 28/03 *The Cost Of HIV Prevention And Treatment Interventions In South Africa* By N. Geffen N. Nattrass & C. Raubenheimer
- 29/03 *Predicting the Social Consequences of Orphanhood in South Africa* By R. Bray

The Centre for Social Science Research

The CSSR is an umbrella organisation comprising five units:

The Aids and Society Research Unit (ASRU) supports quantitative and qualitative research into the social and economic impact of the HIV pandemic in Southern Africa. Focus areas include: the economics of reducing mother to child transmission of HIV, the impact of HIV on firms and households; and psychological aspects of HIV infection and prevention. ASRU operates an outreach programme in Khayelitsha (the Memory Box Project) which provides training and counselling for HIV positive people

The Data First Resource Unit ('Data First') provides training and resources for research. Its main functions are: 1) to provide access to digital data resources and specialised published material; 2) to facilitate the collection, exchange and use of data sets on a collaborative basis; 3) to provide basic and advanced training in data analysis; 4) the ongoing development of a web site to disseminate data and research output.

The Democracy In Africa Research Unit (DARU) supports students and scholars who conduct systematic research in the following three areas: 1) public opinion and political culture in Africa and its role in democratisation and consolidation; 2) elections and voting in Africa; and 3) the impact of the HIV/AIDS pandemic on democratisation in Southern Africa. DARU has developed close working relationships with projects such as the Afrobarometer (a cross national survey of public opinion in fifteen African countries), the Comparative National Elections Project, and the Health Economics and AIDS Research Unit at the University of Natal.

The Social Surveys Unit (SSU) promotes critical analysis of the methodology, ethics and results of South African social science research. One core activity is the Cape Area Panel Study of young adults in Cape Town. This study follows 4800 young people as they move from school into the labour market and adulthood. The SSU is also planning a survey for 2004 on aspects of social capital, crime, and attitudes toward inequality.

The Southern Africa Labour and Development Research Unit (SALDRU) was established in 1975 as part of the School of Economics and joined the CSSR in 2002. SALDRU conducted the first national household survey in 1993 (the Project for Statistics on Living Standards and Development). More recently, SALDRU ran the Langeberg Integrated Family survey (1999) and the Khayelitsha/Mitchell's Plain Survey (2000). Current projects include research on public works programmes, poverty and inequality.
