

review

Liz Slater

**Building high-performing
and improving education
systems**

Quality assurance and accountability

Acknowledgements

We would like to express our thanks to NFER (National Foundation for Educational Research) for carrying out the scoping and bibliographic work on which this review was based.

Thanks are also due to Methuen and Co. Ltd for permission to reproduce an extract from: *Instruction to deliver: Tony Blair, the public services and the challenge of achieving targets* by Michael Barber (2007).

The views and opinions expressed in this publication are those of the author and do not necessarily represent the views of CfBT Education Trust.

© Copyright CfBT Education Trust 2013

Contents

1	Introduction: why monitor and evaluate?	2
2	What's new about monitoring and evaluation?	3
3	Structures and responsibilities	5
4	What are the processes, tools and models?	9
5	What data and evaluation tools are there?	11
6	Considering some models: balancing accountability and improvement	15
7	Evaluation: what works?	23
8	Conclusions	28
	References	31

1 Introduction: why monitor and evaluate?

Why are systems^a investing in monitoring and evaluation?

Monitoring, evaluation and quality assurance in their various forms are seen as being one of the foundation stones of high-quality education systems. According to Akkari,¹ for instance, the Tunisian Ministry of Education believed one of reasons for the ‘mediocrity of the internal efficiency of the education system’ was the lack of a ‘culture of evaluation’. Similarly, De Grauwe,² writing about ‘school supervision’ in four African countries in 2001, linked the decline in the quality of basic education to the cut in resources for supervision and support.

Countries and states monitor and evaluate not only to drive educational improvement but also for accountability.^b There are difficulties with putting monitoring and evaluation into practice, and some opposition to it.³ However, there seems to be a general recognition that evaluation is part of the process of education and that accountability is a normal and necessary part of the supply of public services.⁴ For instance, in its document on inspection of public services,⁵ the UK Office of Public Services Reform described the purposes of inspection as including accountability^c and service improvement.

Monitoring and evaluation are used to achieve educational improvement for two reasons: to build or maintain the nation’s economic competitiveness and to realise individuals’ potential. Writing about the United States, for instance, Mayer et al.⁶ believed that the ‘nation’s economic and social health’ depended on the quality of its schools and that there was therefore a need to define, assess and monitor quality to ensure the existence of high-quality schools. While the targets for educational improvement may differ, these overarching purposes are common. Realisation of individuals’ potential is not only in the interests of the state, but is also linked to equity.⁷ So, in Chile, the Ministry of Education is responsible for drawing up policies and programmes for ‘improving the quality and equity of the system’;⁸ Ghana sees high-quality education as ‘essential to achieving the goals of the country’⁹ (i.e. economic targets, together with those relating to access to education) and Indonesia¹⁰ has three strands to its education policies: the expansion of access; the improvement in quality and relevance; plus good governance and accountability. (The overarching purpose is to improve the country’s competitiveness.) Similarly, aware that differences in student achievement had been linked to differences in socio-economic status, culture and whether students lived in cities or the country, Poland took account of these issues in planning and piloting its monitoring systems as part of wider education reforms launched in the 1990s.¹¹ The draft National Strategic Reference Framework Plan (2007-2013) continued the emphasis on quality, giving priority to the creation of a quality assessment culture.

^a Levels of jurisdiction differ. In some instances, it will be a country, such as Ghana or Chile; in another it may be a subsidiary jurisdiction, such as Victoria in Australia or Ontario in Canada.

^b UNESCO reports Canada as undertaking assessment of performance because “With their vision of responsibility and accountability, [there is] great emphasis... on the importance of reporting to students, parents and taxpayers on outcomes achieved by educational institutions and administrations...” World Data on Education 2006/07: Canada

^c Defined as providing assurance to ministers and the public about safe and proper delivery of those services.

There are other reasons linked to the introduction of change. While introducing its educational reforms following independence in 1991, Slovenia¹² used a range of monitoring and evaluation tactics to find out what was happening in its school system, to learn from other countries and to benchmark its performance against international standards. With increased pressure on public finances – either because of increased demands for access (as in Tunisia and Ghana, for instance) or because of budget cuts (as in England) – there is a related interest in ensuring value for money.

2 What's new about monitoring and evaluation?

What has changed?

Monitoring and evaluation are not new, though both the scale and range of activity have expanded considerably. Ofsted, in England, can trace its foundation to the appointment of two inspectors in 1837 with a remit to monitor the use of grants to schools operated by the National Society and the British and Foreign Society.

However, several writers have identified a change in approaches in monitoring and evaluation towards the end of the twentieth century. This is partly due to a change in purposes (see below). It also shows the influence of other disciplines.

Hopkins¹³ for example, traced the interest in evaluation to organisational development work and evidence from the 1970s that survey or data feedback was 'the only treatment associated with substantial improvement' and linking this with school self-evaluation. Consequently, he argued that:

- a. the quality of students' learning outcomes was linked to the school's organisational health,^d and that
- b. school review linked to an internal improvement process was a proven strategy for improving organisational health.

At the same time, international interest in large-scale educational change has focused on decentralisation (i.e. including to school level) and raised standards. Decentralisation has, in turn, led to increased demands for accountability. Therefore, many systems aiming to improve the quality of their schools have put in place external review. (Hopkins warned that while evidence showed this was a necessary first step, there also needed to be processes in place to ensure there was follow-up so that evaluation outcomes had an effect.) Putting a greater emphasis on involving the school in the review process is not only seen as helping the individual institution, but also as a building block for implementing large-scale, system-wide changes.¹⁴

^d e.g. the morale and commitment of the school's staff, the school's expectations and standards.

Why the change?

Work on developing evaluation techniques has been undertaken partly because policy-makers believe traditional inspection has not led to improvements in school performance. De Grauwe and Naidoo¹⁵ and other analysts suggest some explanations:

- a. Schools are complicated organisations. All the staff will not necessarily have the same views and priorities. Consequently, the conclusions from evaluation may be taken to mean that existing ways of working are satisfactory. In addition, while external evaluators are generally trained thoroughly,¹⁶ there is often much less training for schools on how to follow up on evaluation reports.
- b. Schools need guidance on how to improve, in addition to information on their performance. While there are quite comprehensive evaluation frameworks (including criteria and scales), and even monitoring regimes (cf. the Netherlands),¹⁷ systematic approaches to the implementation of change and improvement strategies are less evident. Making schools responsible for improvement requires that heads are able to take on leadership of evaluation and innovation. This may demand a considerable change of role, as in Chile.
- c. There may be a lack of coordination between the various evaluation mechanisms and the personnel involved. In England, for instance, while there is a link between school self-evaluation and Ofsted inspection, the link between inspection report findings and school rankings in performance tables is less clear. Moreover, Ofsted-contracted inspection providers are not encouraged to build relationships with local authorities, which have a responsibility for school improvement.

Consequently, countries have either changed their inspection systems or added in other arrangements including school self-evaluation and evaluation using the outcomes of examinations and tests. The Netherlands, for instance, changed its approach to inspection in 2002, moving from using external inspection for compliance to combining school self-evaluation with external inspection and publication of reports on the internet.¹⁸ The Education Review Office in New Zealand has made similar changes. Both now adopt a proportionate approach to external inspection, inspecting weaker schools more frequently than stronger ones. However, 'Decentralised management and proportional monitoring are only possible if there is adequate quality assurance within the sector and the institutions themselves.'¹⁹ In 2003, Hong Kong moved from top-down inspection by officers of the Quality Assurance Division to school self-evaluation alongside external school review and access to key performance measures so that schools could assess their strengths and find areas for improvement.²⁰ In the period between 2003 and 2008 Hong Kong used an impact study, including consultation with stakeholders, to monitor and refine its approach.²¹

Interest in school improvement is international. Not surprisingly, as a result, governments draw on international expertise. David Hopkins led a seminar for Chile on its plans for an inspectorate ('Superintendency'),²² and Cambridge University and Cambridge Education were contracted to assess the implementation of changes to evaluation in Hong Kong²³ which had been designed following a survey of international practice. Some governments choose specifically to include an international dimension in their quality assurance policies. This can include participation in international benchmarking studies such as PISA (e.g. Korea, Finland, New Zealand, Australia, the Netherlands, Poland, Slovenia, Chile, Hong Kong, China, Singapore, Tunisia and Indonesia).²⁴

3 Structures and responsibilities

Decentralisation, control and accountability

One of the most widespread changes has been decentralisation – freeing up schools to take more control of areas such as staff appointments, curriculum delivery and finance. However, this has been matched by a requirement for greater accountability.

School autonomy is also balanced by rules (policy frameworks) for inputs, processes and outputs. In Poland²⁵ for instance, greater freedom in terms of the curriculum, choice of textbooks and teacher education has been matched by policies to ensure greater consistency in school-leaving examinations.⁶ In the Netherlands there is central prescription of curriculum frameworks and standards.²⁶ In Canada there is greater control over teacher training and evaluation, student testing and reporting.²⁷

Integrating accountability into structures and systems

In general, effective systems, as well as those trying to improve, focus their attention on schools because several studies have shown that, notwithstanding policy initiatives at national, regional or local level, it is schools that make the difference: it is in schools that these policy initiatives do – or do not – have an impact on teaching quality.²⁸

Consequently, this chapter focuses on schools, though it notes where high-performing and improving systems, such as British Columbia (Canada), Slovenia and Poland, place their school monitoring and evaluation within a framework which ties together arrangements for local and national accountability and support.

Who does what? The central/regional/school split

The desire to achieve decentralisation and greater local accountability is reflected in changes to structures and responsibilities. Generally, national and/or federal governments take responsibility for more strategic matters, municipalities have a largely operational/executive role and schools either seek or are required to involve their local communities in their work.²⁹

⁶ The examinations are used to decide on student destinations, as well as a measure of school output. So they are used for accountability purposes.

In Slovenia,³⁰ the Ministry of Education has responsibility for feedback on the performance of the education system drawing on statistics, targeted evaluation and research, state-wide examinations and international research (e.g. PISA). Institutes ('developmental bodies') provide expert input including documents to inform decision-making and curriculum development proposals for ministerial approval; assist in the evaluation of schools; contribute to the preparation of state-wide examinations; promote development in schools (including offering professional advice to teachers) and initiate education-related projects. Municipalities have largely executive/operational responsibilities, providing capital and maintenance funds. At school level, councils (similar to school governing bodies) made up of teachers, parents and representatives of the founding body, have responsibilities which include the adoption of the school development plan and the appointment of the headteacher. The headteacher is responsible for the professional and administrative leadership of the school, including leading internal school evaluation. Broadly similar arrangements are in place in Poland, except that external oversight of quality ('pedagogical supervision' – i.e. activities including inspection, evaluation, control, support and advice for principals and teachers) is carried out through regionally-based superintendents (*kurator*) who are responsible to the regional head who, in turn, reports to the Prime Minister.³¹ In both cases, the examinations system is used as an additional level of quality assurance and to identify problems.

Australia devolves most responsibilities to federal level and below. However, an overview of national quality is maintained by testing all pupils aged 8–9, 10–11 and 12–13 against national literacy and numeracy benchmarks. Additionally,³² there are national plans to improve accountability and transparency, including data to support benchmarking of schools and clearer reporting on performance. The results are reported to parents/carers and schools.³³ There is further decentralisation at state level. For instance, in Victoria, the Office of School Education sets policies, regulates and manages the overall system (including a curriculum and standards framework with seven age-related levels for reporting student achievement), but schools plan (and are therefore accountable for) courses and curriculum delivery.

British Columbia introduced a more formal accountability structure in 2002.³⁴ School planning councils draw on evidence from classroom, school, district and provincial levels to develop plans that are shared with the school community and must be approved by school boards. School boards (i.e. municipalities) are responsible for submitting their district accountability contract to the Minister each year. This focuses on student achievement, taking account of any areas for improvement identified by the Minister. The third component is district reviews conducted by expert teams who report to the school board and the Ministry of Education and make recommendations designed to improve pupil achievement. In 2006/07, British Columbia reported initiating an 'accreditation programme for primary and secondary schools' to 'strengthen public accountability'.³⁵ New Zealand adopted a similar approach in the late 20th century,³⁶ aiming to involve parents and communities in the management of schools 'in partnership with principals

and teachers' through school boards of trustees who, in turn, are responsible for meeting objectives set out in their charters.

Creation of the decentralised structure and linked accountability measures often run alongside setting up an autonomous or quasi-autonomous body responsible for aspects of monitoring and evaluation, such as the Inspectorate in the Netherlands.³⁷ New Zealand's Education Review Office, for instance, nominally has its own Minister.^f Ofsted was established in 1992 in England. Some writers³⁸ have argued for an entirely independent body. In principle, this sounds sensible. However, possibly of more importance is the professionalism of its members, the quality of the output, honest handling of issues and a willingness to accept responsibility for failures as well as successes up to and including ministerial level.

Some states choose not to have an inspectorate, but to establish arms-length bodies with quality assurance functions. In Finland, the Ministry of Education established a separate Education Evaluation Council in 2003 with responsibility for 'planning, co-ordinating, managing and developing the evaluation of basic and upper secondary education'.³⁹ (However, the authors of this report noted that the system relied 'on the proficiency of... teachers... to carry out the objectives laid down in the criteria' and the importance of both self-evaluation and external evaluation.)⁴⁰ The Ontario Education Quality and Accountability Office⁴¹ not only conducts province-wide assessments of student achievement, but has developed the *Education Quality Indicators Framework* which provides background information (e.g. on demographics) to support the interpretation of its findings at school, board and provincial level.

Where does the inspectorate come in?

It is important to be sure of consistency in the use of criteria and in making judgements. While longer-established inspectorates will have focused on compliance, the emphasis tends now to be on improvement and accountability. However, precise roles differ, shaping their relationships with schools and other parts of the education system. Initially, Chile chose not to have an independent inspectorate, but, in planning to introduce the inspection function, its purpose was seen as strengthening the state's ability to guarantee the quality of education.⁴² The Presidential Advisory Commission's report recommended a superintendency to evaluate compliance with quality standards through inspection, use of international benchmarking, more information on outcomes and to intervene in underperforming schools. Legislation established the superintendency in 2009.⁴³

In principle, the specialised evaluation/inspection units are staffed by professionals with specified levels of knowledge and skill, often including experience in education and management. In the Netherlands there is an expectation that inspectors will have around nine years' experience as a teacher or in a school management post and will undertake specialised training. In Slovenia, inspectors are civil servants recruited by open

^f In fact, the Hon. Hekia Parata is both Minister of Education and Minister Responsible for the Education Review Office (March 2013).

competition. In addition to a minimum of seven years' experience in education or related disciplines, they need to have passed a state examination in administrative procedures.⁴⁴ A year after appointment they must pass a school inspector examination, followed by re-certification via tests of professional efficiency at four-year intervals. In some countries value for money and efficiency are audited using specialist teams with technical expertise. In others, the central ministry may set up arrangements, including guidance and benchmarks. The issue then is monitoring and follow-up (as with any inspection).⁹

De Grauwe⁴⁵ noted the difficulty inspectorates may experience in changing what they do, and particularly the difficulties inspectors had when they were expected to be both a representative of government and a critical friend. Changing from examining compliance to reviewing accountability and improvement requires development of a different skill set, which inspectorates may find difficult. There will be a need for re-training, new appointments and the careful development of a cohort of suitably skilled professionals. While this development is taking place, systems need to make sure the inspectors focus on the correct criteria and to check the robustness of judgements. In switching from inspection to review, Hong Kong reportedly 'invested heavily in retraining of external review teams, recognising that a new relationship requires a shift from an inspection mindset to a review approach'.⁴⁶

What drives the evolution of monitoring and evaluation systems?

Compliance, accountability or improvement?

Approaches to monitoring and evaluation evolve in line with the development of education systems. The initial focus is likely to be on compliance. Later, the emphasis is increasingly on a holistic drive towards school improvement and accountability.

Evaluation for improvement and accountability is concerned with looking at performance, drawing on all the information available.

The purposes of school evaluation may be to ensure compliance with administrative requirements, to fulfil accountability requirements and to set the scene for improvements in teaching and management. In practice, monitoring for compliance may take place alongside evaluation for accountability and improvement. There may be control and quality assurance of inputs (e.g. resources, teacher quality, access to education, textbooks), of processes (e.g. management, teaching and learning, relations with the community) and of outputs (e.g. examination results, retention, progression).⁴⁷

Accountability aims to provide information to policy-makers and stakeholders about meeting standards,⁴⁸ and value for money. It is summative. If schools fail to meet standards, they may be placed on probation and given a timeframe within which to make good the weakness.^h Some states, such as Chile, New Zealand, the Netherlands

⁹ "The Department is accountable to Parliament and the public for the proper management of the money given to schools. ...The Department has set standards for financial management in maintained schools, and has a framework in place for gaining assurance which relies on local authority oversight of schools. The Department needs to know that this framework is meeting the intended objectives, and is capable of alerting it to systemic issues with schools' financial management that require action or intervention as they emerge." Report by the comptroller and auditor general, HC 1517 Session 2010-2012. 19 October 2011, Department for Education, *Oversight of financial management in local authority maintained schools*.

^h Where Ofsted inspections of non-association, independent schools find that schools fail to satisfy the regulations for registration and there is a failure to make the required improvements, the school may be deleted from the register of independent schools and forced to close.

and Indonesia, publish reports on inspections and/or examination results to be used by parents in choosing schools. By doing so, they are trying to motivate schools to improve or face the loss of funds and, possibly, closure.

Evaluation focusing on improvement includes access (and associated equity issues) and educational performance (i.e. quality and efficiency). It is formative, aiming to bring schools up to a uniformly high standard and raise the performance of all students.⁴⁹ Systems may choose to achieve improvement through the pressure of public accountability (see above), the introduction of support or a combination of the two. While agreeing about the importance of accountability, writers⁵⁰ advised that, particularly in countries where resources are limited, the emphasis should be on improvement, since there is little to be gained from blaming schools for failures for which they cannot be held accountable.

Whereas accountability is often seen as contractual (e.g. schools meeting the requirements of the system and delivering specified levels of output) associated with the use of the market, Faubert, drawing on Learmonth and Gurr⁵¹ identifies two other types which may have a more subtle, yet powerful impact. These are moral accountability (i.e. meeting the needs of students, their parents and the wider public) and professional accountability (i.e. meeting one's own expectations and those of colleagues).⁵²

The challenge is to design evaluation systems that bring together accountability and improvement purposes, gain the commitment of all those involved and build capacity.

4 What are the processes, tools and models?

Processes and tools: a summary

The core monitoring and evaluation processes are the regular collection of information, the evaluation of that information, and using the evaluation to decide on action.⁵³ The main monitoring and evaluation tools, which may have a variety of names, are external supervision/inspection/audit, internal evaluation/self-evaluation and examinations/tests.

External supervision/inspection

External supervision (or inspection) is often organised by the national or regional education ministry (or its agent) and may include checking that administration and professional activities conform to the legislation and are effective. Areas covered tend to include teaching and learning, the curriculum, responding to student needs, leadership and management as well as relations with parents and the wider community. It generally takes place on a cycle – often of 3–6 years – which may be adjusted to take account of the strength or weakness of particular institutions. The process will include some or all of the following:⁵⁴ notice of the inspection; preparation of pre-inspection and background information (by the inspectorate and/or the school, including questionnaires for school

staff and parents); a site visit; interviews with school leaders, teachers and sometimes parents and pupils; inspection of documents and observation; reporting (including feedback to school staff and, possibly, publication of the outcomes) and follow-up (which may include compliance or support measures). External evaluation, using common criteria, helps to ensure consistent judgements across schools, demonstrates the standards expected and produces a national picture of education quality. It may also encourage improvement. However, it requires a skilled team of inspectors, is expensive (as well as involving an opportunity cost) and may not be sufficiently flexible to take account of the particular circumstances of some schools. It may encourage weaker schools in particular to concentrate on meeting inspection criteria rather than genuinely improving performance.

Internal evaluation/self-evaluation

Internal evaluation/self-evaluation (which can include individual self-evaluation) is usually led by the headteacher, possibly with stakeholder involvement. It may be self-initiated or a statutory requirement. Sometimes it is a precursor to external evaluation, though it generally occurs more frequently – on an annual cycle – as part of school review and planning for improvement. Even where schools use criteria and benchmarks devised at national or regional level, there is a risk they will not be able to carry out a sufficiently rigorous self-review, particularly when the approach is new.⁵⁵ The process relies on the strength of the headteacher's educational and leadership capabilities. With the transfer of responsibility for management, resources and the delivery of the curriculum, many systems seek a 'stronger form of accountability' i.e. external evaluation – alongside.⁵⁶

Examinations and tests

Examinations and tests may be used for formative assessment to inform teaching and learning; as a mechanism for selecting students; as a means of checking whether students are learning the mandatory knowledge and skills; and (in the case of international tests, such as PISA, TIMSS) to benchmark pupils' achievements against those of other countries. They may be used to carry out a range of functions simultaneously; alternatively, different examinations and tests may be used for different purposes. Examinations and tests are increasingly used to judge the performance of schools (and education systems) and may be used to rank schools. Where tests and examinations are standardised, they can be used to give an indication of the performance of schools and even students. However, unless the results take account of the value added by schools (which requires the ability to track individual students), comparatively better performance may simply reflect the fact that examination results and socio-economic status are closely linked. Without student tracking, tests and examinations will also fail to account for the impact of student mobility.⁵⁷ They can also limit the breadth of the taught curriculum and there is some evidence to suggest that high-stakes testing (i.e. where it determines student promotion or graduation), may lead to higher drop-out rates among low-achieving students.⁵⁸

A review of practices in OECD countries⁵⁹ suggested that where examinations were based on a common curriculum with established performance standards (and possibly targets), they helped schools to make comparisons between students; gave broad comparisons of schools' school performance; showed which schools needed help and made it possible to evaluate the effectiveness of that assistance; showed up gaps or weaknesses in curriculum delivery and were useful in planning resource allocation. Publication of student results can raise public awareness and provoke pressure for improvement. Harris⁶⁰ quoted Michael Barber on the benefits of performance tables:

'They make evidence about performance public, they focus minds on the priorities they encompass, and they make sure, in whatever system they are applied, that something is done about the individual units at the bottom of the league table...'

As the preparations for national monitoring in Poland⁶¹ showed, setting up a national assessment system is a demanding project. It requires the development of test items, reliable mark schemes, data on schools (e.g. for sampling), training for teachers to enable them to administer the tests and effective systems for disseminating the information, not just up to central government, but also to local partners so that they see benefits and become committed. However, where there are few professionals with the knowledge and skills to carry out external evaluation, national assessment may offer a more realistic solution to those seeking to monitor and evaluate system performance while developing a shared understanding of standards.

Where they are the only mechanism for public accountability, the way in which examination results are published and explained will determine whether they are of any use to parents.

5 What data and evaluation tools are there?

Defining information requirements

Interest in improvement has driven the growth of indicators and benchmarking. There are many definitions of education indicators. Visscher,⁶² for instance, defined them as 'a number by means of which the quality of the functioning of an institution or a system (e.g. a policy area) is expressed' identifying, in particular, the use of indicators for public accountability and improvement purposes. Criteria describe what is measured by the indicator, while standards define thresholds or levels against which measures can be undertaken.⁶³

Soundly-based decisions about the range of data to be collected, the benchmarks against which the data is to be evaluated and the way it is to be collected are fundamental to effective evaluation. There is a risk of selecting indicators simply because they are easy to measure. States and institutions need to be clear about the purposes of monitoring

and evaluation (e.g. accountability, improvement, reward), the information needed and how it will be used.⁶⁴ This will help them to develop a set of indicators which allow them to make judgements about particular aspects of the system, as well as the system overall. It is important not to choose short-term targets which may skew schools' focus away from real improvement.⁶⁵

The choice of indicators needs to reflect the policy background (e.g. whether the state's aim is universal access to basic education) and take account of the resources available to collect, interpret and act on the data (e.g. data processing capabilities, the capacity to track individual student outcomes, the existence of a workforce skilled in evaluating teaching). China, for instance, reported⁶⁶ investments to improve data collection and analysis. Issues identified included the gap between urban/rural and regional educational levels, where relevant indicators included participation in compulsory education, conditions in schools and teacher quality.

It is easy to focus on the collection of meaningful amounts of data and evidence and to lose sight of the burden placed on the institution being evaluated.

Quantitative indicators are likely to relate largely to inputs and outputs. To be meaningful, they will need to include appropriate targets or benchmarks. There is a range of sources. The UNESCO Institute for Statistics, for instance, produces a list of education indicators, together with guidance on their calculation⁶⁷ and data on individual countries. It also provides information on outputs such as literacy levels. There are publications such as the OECD's *Education at a glance* and Eurydice's *Key data*. International pupil sampling also provides benchmarking data on outcomes as well as trends over time, including the opportunity to make comparisons with similar countries. The Trends in International Mathematics and Science Study (TIMSS) has, for instance, been administered four times since 1995 and enables participating countries to compare outcomes with those of others in the programme.

High-performing and improving countries usually have (or are developing) curricula with age-related performance standards in order to judge educational standards. These may include additional features to take account of student intake (i.e. socio-economic status) and/or value-added. In England, for instance, based on the National Curriculum standards, schools, local authorities and inspectors can access and edit data relating to student progress and attainment in Key Stages 1, 2, 3 and 4; set and monitor progress towards targets; take account of contextual information about the school and compare its performance with other schools nationally.ⁱ Hopkins reported in 2006⁶⁸ that Chile was introducing standards in order to clarify expectations about age-related attainment and Slovenia's reforms included the introduction of subject-related attainment targets.⁶⁹

ⁱ www.raiseonline.org

Other indicators, which will tend to relate to processes (such as teaching or management) and outcomes, may be defined by reference to criteria and require the exercise of professional judgement. Ofsted⁷⁰ provides inspectors with criteria for judging aspects of teaching, such as formative assessment, together with syntheses which exemplify what outstanding, good, satisfactory and inadequate practice look like. Ontario, Canada, has developed the *School Effectiveness Framework*, designed to enable schools to undertake self-evaluation against 'research-based indicators' and plan 'for focused improvement'.⁷¹

How are information and data collected?

The collection and use of data can seem deceptively simple, but establishing systems to collect even the most basic data can prove problematic, as Ghana has discovered with its education management system (EMIS).⁷² One of the reasons given for Finland's decision to abandon, temporarily, the collection of data in the 1970s was the lack of 'technological readiness to create indicators'.⁷³

The extent of the rules about how information is gathered varies. However, high-performing and improving countries seem to focus on compliance, outcomes, teaching, management and quality processes. For instance⁷⁴ the inspection framework used by the Dutch inspectorate includes indicators and norms for aspects of quality. Quality is organised into five categories: results, learning process, student care, quality assurance and compliance with legislation. The framework contains criteria to enable inspectors to determine whether the school is 'at risk' or 'failing'. A similar approach is used by Ofsted in England.

Some of the high-performing systems are showing interest in outcomes which go wider than academic performance. The Singapore School Excellence Model asks schools to evaluate the extent to which they contribute to the achievement of the Desired Outcomes,⁷⁵ which relate not only to knowledge and skills, but also aspects such as citizenship and the capacity to learn.

In addition to reviewing documents, lesson observation is used (e.g. in the Netherlands and England). In some cases, such as in England, this is structured through the use of checklists.

The views of the community, parents, members of the workforce and sometimes students may be collected via questionnaires. However, attempts to gather data which can be processed relatively cheaply may lead to simplistic questions and produce poor response levels. In common with other states, British Columbia⁷⁶ was disappointed by the low levels of parental responses to questionnaires designed to test their views on the quality of education. Some states and institutions supplement or replace questionnaires with meetings and/or interviews. For instance, in Hong Kong, lay members of inspection teams' roles include participating in the parents' meeting and talking to students and staff.⁷⁷

Where external examinations and test data are available, they may be used by internal and external evaluators to compare performance between classes or institutions.

In some countries, schools are either required or encouraged to use the framework and criteria used by the external evaluators for their own self-evaluations. The outcomes of these self-evaluations may then form part of the evidence base used by external evaluators. In Slovenia,⁷⁸ the headteacher's report analysing performance in the previous year forms part of the evidence for external evaluation. In England, until recently, schools were required to submit self-evaluations prior to inspection by Ofsted inspectors.

International benchmarking

Many countries that are either high-performing or have made marked improvements in student outcomes have chosen to join in international benchmarking studies, such as TIMSS, PIRLS or PISA. PISA, for instance, has carried out assessments covering reading, mathematics and scientific literacy at three-yearly intervals since 2000. Samples from participating countries comprise from 4,500 to 10,000 15-year-old students, who also complete a questionnaire about their background. Principals of participating schools also complete a questionnaire about their schools. The resulting reports and analyses allow countries to benchmark their students' performance against that of other participating countries.

Benchmarking, particularly when cultural neighbours or economic competitors are used as comparators,⁷⁹ helps countries to identify areas (such as the achievement levels of students and curriculum coverage) that may be weak compared with other countries. It also shows different ways of doing things and challenges assumptions about what schools can achieve (cf. equity in Finland). Regular participation can enable countries to identify issues – such as the US' poor performance in the 1997 TIMSS.⁸⁰ Some countries, such as Canada, have revised their national assessment (which is based on sampling) as a result. The Pan-Canadian Assessment Programme, introduced in 2007, which tests a random sample of 30,000 13- and 16-year-olds every three years was deliberately designed so that it matched international assessments such as PISA.⁸¹

What happens after evaluation?

The most difficult stage of evaluation is follow-up. Without follow-up, the investment in evaluation is unlikely to pay off (by leading to improvement). For instance, Straus⁸² reviewed what happened after Slovenia had taken part in international assessments. He wrote that, as well as identifying strengths and weaknesses, it would be necessary to find out the reasons for weaknesses and follow up with action to solve problems, to 'complete the monitoring cycle'.

In some systems, there is feedback to the school as well as publication of a report of the findings after external evaluation. There may also be meetings with parents. For instance, in the Netherlands,⁸³ inspectors discuss their findings (including teacher quality, the

curriculum, testing and pastoral care, as well as how quality of provision is managed and assured).

Steps following evaluation depend on the purposes of that evaluation. Where the objective was to check for compliance with standards and processes, the follow-up will depend on how seriously rules have been broken. The school may be given time to improve, have its registration cancelled or be closed. Where the non-compliance with the rules is significant, or there is a failure to improve, then more severe punishments may result: in the Netherlands, where schools have performed poorly for two years, the Inspectorate advises the Minister to intervene.

In some countries, identification of weaknesses will lead to additional support or funding (e.g. in England, the Netherlands, Chile, Australia, Finland). In Ontario,⁸⁴ schools are encouraged to network with other similar schools that have made 'significant and sustained progress in student achievement'.

Where the purpose of external evaluation is improvement, the next evaluation will be used to check that recommendations have been followed up. Schools may be required to prepare a plan for improvement. The time between inspections may depend on the quality of the school and its capacity to improve. This is described as 'proportional supervision' in the Netherlands Supervision Act of 2002.

Some systems specifically set out to reward high standards of performance. In addition to publication of reports or even league tables, they may publicise outstanding schools such as in England and Ontario, Canada.⁸⁵ Singapore makes a range of awards to schools⁸⁶ linked to the School Excellence Model.

6 Considering some models: balancing accountability and improvement

What are the variables?

In considering what approaches to adopt, there are three issues:

- a. The purposes of external evaluation/inspection/supervision: whether it is meant to check on schools' compliance or provide support
- b. How information from the various evaluation arrangements (i.e. results of external evaluation, self-evaluation and tests/examinations) is brought together and how decisions are made about the relative importance of each type of information
- c. The context, i.e. the strengths and weaknesses of the existing education system and, in particular, of the evaluation arrangements already in place so that the methods selected support improvement⁸⁷

States develop quality assurance and evaluation models which, either through careful advance planning, or through a process of adjustment, are used individually or in combination to achieve their policy purposes. Poland, for instance, launched its education reform programme⁸⁸ in 1999 after several years of planning and preparation, drawing on financial support from the EU through its Strategic Measures for Achieving Reform Targets programme and involving experts from England and Scotland. Bahrain took a similar approach. Conversely, changes in Ofsted's role in England and that of the inspectorate in the Netherlands⁸⁹ (in 2002) were adjustments to existing models.

Compliance monitoring to control inputs and processes

The narrowest approach is monitoring schools' compliance with requirements relating to inputs.⁹⁰ It deals with the establishment and maintenance of state standards. It is not appropriate for monitoring school effectiveness, though it can be used to monitor input measures linked to laying the groundwork for that effectiveness, such as access to education, textbooks, teacher qualifications, teacher attendance, levels of investment and administrative requirements. The MOESS EMIS in Ghana, for instance, gathers a wide range of information through the Annual Education Census on the school's profile, management, infrastructure, materials and equipment, enrolment, teaching and other staff.⁹¹ In China, the State Education Inspection Team has monitored compliance with the Law on Compulsory Education since 2003.⁹² Monitoring of inputs may be preferable to trying to carry out diagnostic or performance monitoring where the numbers of expert professionals are insufficient. It can be used for planning purposes.

Monitoring may be extended to include processes such as 'management of financial, human and material resources needed to contribute to the quality of a school's performance in compliance with the regulations'⁹³ including the use of mandated textbooks and the delivery of the specified curriculum to a particular timescale.⁹⁴ It is generally summative and may include measures to improve compliance where necessary. While checking for compliance may relate to putting in place the most basic building blocks of an education system, it can also be used to monitor equity (e.g. in the provision of support for pupils with special educational needs) and a level of quality (e.g. the certification of teachers, the match between teacher qualifications and their deployment).

Integrating evaluation and examinations

Slovenia⁹⁵ uses external and self-evaluation, as well as examinations and research to control and evaluate inputs, processes and the quality of outputs:

- a. The external Inspectorate for Education and Sport ensures schools have implemented educational legislation and are using funds appropriately, and evaluates the quality of educational provision. The inspections, which are scheduled (every five years) or unscheduled (i.e. unannounced) therefore cover a wide range of administrative, financial and professional checks.

- b. Since legislation in 2008, schools have to report every year on the results of self-evaluation. While the headteacher has overall responsibility for evaluation as educational leader, teachers are closely involved. Since the 1990s, schools have had a range of information including data from national and international studies, results from students' national examinations and national averages (benchmarks) to help them to assess the quality of their work. Teachers work together to analyse academic results at least once at the end of each assessment period. The outcomes of the analyses are presented to parents and discussed with individual students. At the end of each academic year, Teachers' Assemblies, Parents' Councils and Schools' Councils discuss the school report. The outcome feeds into the subsequent institutional development plan.
- c. The main mechanism for assessing outputs is national testing conducted at the end of the second three-year cycle in basic school (i.e. for students aged circa 15 years), upon completion of basic school and then at the end of upper secondary school. These evaluations are conducted by expert bodies: the National Testing Committee at basic school level, the National Examination Committee for the practical part of the final examination at vocational education level; the National Committee for the Vocational *Matura* and the National Committee for the General *Matura* (both of the latter with subject committees). The National Examinations Centre, the leading institute for external assessment in Slovenia, provides technical and administrative support to the committees.

Among the issues identified by the Government, is the need⁹⁶ to achieve a more consistent approach to evaluation and quality assurance in pre-school institutions and schools. The 2008 Act announced the introduction of national quality indicators and the Ministry of Education invited bids for a project to develop a consistent and coherent system of evaluation at national level.

Shifting the balance between external and internal evaluation

Early models of external monitoring and evaluation aiming to bring together accountability and improvement were described by De Grauwe⁹⁷ as the 'classical supervision' model. While the model has the advantages of universal coverage and a comprehensive role (i.e. control, evaluation, support and elements of advice via feedback and/or a report), it is costly, relies on the existence of a team of competent inspectors, risks role conflict and an emphasis on administration rather than on the quality of teaching, is bureaucratic and gives limited follow-up.

De Grauwe notes this model has evolved into an approach which, while still based on central control, takes account of school self-evaluation and data on outcomes. It is designed to be less costly and bureaucratic, leaving schools to make improvements. Schools use their own budgets to buy support under the supervision of a board of trustees/governors. External inspectors audit each school as part of a programme and

publish a report which makes the school accountable to their stakeholders in government and the community. The emphasis of the external evaluators has tended to move from looking at processes, such as classroom teaching, to organisation and management. Theoretical advantages include lack of role conflict, potential reduction in the size of the central inspection service and giving responsibility to schools to respond to the report. Weaknesses include the lack of support available to struggling schools, the pressure on schools and the potentially damaging impact on the future of the school with a poor report. There are variants to the model which give more or less weight to the school's own self-evaluation.

Self-evaluation can serve as a preparation for external review (as in England and New Zealand) and feed into the preparation of schools' annual plan, or, alternatively, the external evaluation can take the form of a validation of the self-evaluation, with more weight given to the latter. In Victoria, Australia, for instance, each school prepares a self-review annually, using questionnaires provided by the state Department of Education to survey opinion among teachers, parents and students. Every three years there is a more complex self-review. External evaluators validate the process and outcomes. The validation takes a day, with the validator acting as critical friend and an emphasis on accountability.

In the Netherlands inspectors are required to base their inspections on those carried out by the individual institutions and not to 'place a greater burden on the school than is strictly necessary for careful supervision'.⁹⁸ Inspectors must report on compliance with statutory requirements relating to basic quality standards and conditions for funding, as well as student outcomes and the 'structure of the learning process' (e.g. syllabus, learning/teaching time, school climate, pedagogy, tests and examinations).

In Hong Kong⁹⁹ starting in 2003, systematic internal school self-evaluation was introduced, supported by external school review. As well as a re-training programme for Quality Assurance Division inspectors, there was a further training programme for external reviewers (who were principals seconded from Hong Kong schools to join the external school review teams). Schools were given performance indicators to use in self-evaluation. The model was refined based on extensive evaluation from 2002 to 2008.

There is some evidence that external inspectorates believe there is a need for increased emphasis on the importance of making sound judgements on teaching. The approach, including involvement of the headteacher, was discussed at length in Ofsted's guidance to inspectors.¹⁰⁰ There is also a risk that a move to give more weight to self-evaluation may be premature where schools have not developed a sufficient understanding of educational standards¹⁰¹ nor the ability to carry out rigorous self-review.

¹ Subsequent changes introduced in September 2012 have put even more emphasis on the inspection of teaching. (Ofsted, 2012, *School inspection handbook*)

Some writers have looked at how to tailor monitoring and evaluation to make them more effective. One option is to subject all schools to a uniform evaluation process and criteria in order to satisfy the need for accountability, while tailoring feedback and support to identified need in order to increase impact. (A variant on this is, additionally, to match the frequency and depth of monitoring and evaluation to risk factors.) Some writers believe¹⁰² that evaluation and support should be separated. However, if they are, there is a risk that unless the evaluation report is of exemplary detail and clarity, there will be a loss of information between evaluation and follow-up. (This is particularly so where inspection reports are targeting multiple audiences – as in England.)

Separating accountability from improvement

In some cases, states establish a clear divide between responsibility for accountability and for improvement. Accountability for performance may be covered through examinations, while more financial, audit-oriented aspects may be dealt with by a team of specialists. In reforming its education system, Poland moved from a supervision system where ‘ideological control... was one of the most important tasks’¹⁰³ to a split between external evaluation of performance exercised via examinations and tests, external evaluation of administrative matters (‘administrative supervision’) carried out by the bodies running schools (*powiats* i.e. districts, and *gmina* i.e. communes) and of professional matters (‘pedagogical supervision’) by inspectors on behalf of the *kurators*, who operate at provincial level. While the inspectors monitor compliance with legislation and evaluate the processes and impact of teaching and care in schools, they are also responsible for advising on improvement and innovation, reflecting the overall shift in their role ‘towards advisory activities’ thus combining a control and support role.

Some states choose to separate entirely the control and support roles, appointing a team of specialists with specific responsibility for supporting the schools in their efforts to improve. De Grauwe¹⁰⁴ describes this as the ‘close-to-school’ support model, arguing that all schools are different and that the weaker schools need support to improve. Therefore, the supervision system is tailored to the needs of individual schools, the audit role is removed, and supervisors have a limited caseload and are based close to their assigned schools. Supervision becomes a part of the process of school self-evaluation and improvement. Advantages include the reduction in bureaucracy, flexibility and focus on weaker schools. However, there is a risk that schools in the satisfactory category may be ignored. The model relies on good quality, up-to-date data (e.g. from examinations) to monitor all schools’ performance and identify accurately and quickly those requiring attention. In Chile, supervisors work with teachers on the preparation and implementation of projects that serve as the basis for continuous self-assessment and school improvement. Administrative supervision, which is largely financial, is left to a team of accountants, called inspectors.

Victoria (Australia) also used evidence about school performance¹⁰⁵ to identify schools in difficulty and give them support using Senior Education Officers (SEOs) operating at

regional level with a caseload of around 40 schools. Their role was to act as leaders and change agents for school improvement. An audit in 2007 found evidence of 'considerable progress', but unresolved problems included finding workable systems for giving appropriate workloads to SEOs; training SEOs for their new role as change agents; accurate and timely matching of demand and support including the development of risk factors (triggers for intervention); and support for schools in using school performance data to plan school improvement and measure progress.

School-based accountability and improvement

Finally, in a limited number of states, the responsibility is largely at school level.¹⁰⁶ This model relies on professional and public accountability to drive improvement. Advantages include modest cost (since there is no external evaluator, nor support) and a strong emphasis on the role of the school (which is where change has to occur if improvement is to be sustained). Schools may use national indicators and benchmarks, as well operating within a national curriculum framework. Disadvantages include lack of capacity to generate improvement on the part of weak schools as well as a risk to the achievement of national policy objectives.

Finland adopted this approach in 1991 when it abolished its external inspection service and introduced a lighter national curriculum framework in 1994. However, in consequence, it took the view¹⁰⁷ that 'decentralisation has placed increased importance on evaluation'. While there is considerable local freedom, there are structures to minimise risk, including:

- a. national core curricula decided by the Finnish National Board of Education which are the norm for local or school curricula
- b. a requirement, since January 1999, that education institutions evaluate their operations and effects (i.e. outcomes). Evaluation looks at achievement of objectives, completion of teaching and curricular reforms and the use of resources. Moreover, municipalities are responsible for evaluating the effectiveness of their provision and have to contribute to the national evaluation of the education system.

There is also some evidence of moves to strengthen consistency and control: the Finnish National Board of Education published a Quality Management Recommendation for Vocational Education and Training (VET) in 2008 to help VET providers develop quality management. The Board was also reported to be developing quality criteria for comprehensive education designed to support local decision-makers in their work and focusing on 10 areas of activity.

In Chile, decentralisation made heads of schools accountable. They were required to report each year on the management of the school, including inputs, processes and outputs relating to professional and financial matters. The development of the Good School Leadership Framework (2005)¹⁰⁸ was a response to the difficulties encountered

in trying to shift the headship model from an 'administrative role, in the most classic and restrictive sense of the word'. It includes criteria relating to leadership and curriculum management as well as to management of the school and of resources.

De Grauwe¹⁰⁹ argued that successful implementation of this approach required a society that was broadly similar in terms of socio-economic status and values, teachers who were respected by the public for their professionalism and strong parental interest in education. It also required other evaluation mechanisms such as examinations, a comprehensive system of indicators and a framework of standards.

Using examinations and tests for accountability and improvement

Examinations and tests play a role in all of the high-performing and improving education systems. At institutional and system level, examinations are used for two purposes: accountability (i.e. by reviewing outcomes) and for improvement (i.e. by using a process of diagnosis and follow up action). In some cases, a further level of complication is added by using the same examinations as a mechanism for grading or selecting individual students.

In some countries they remain either the only, or the most popular tool. In Indonesia, for instance, following the 1990s reforms of basic education, expansion of access and investment in professional development, the management of schools was decentralised to district level in 2001. The examination system was used to assess school quality as well as, in some cases, individual students.¹¹⁰

Separating the evaluation of systems and students

In some states, different examinations are used to evaluate system and student performance. In Ghana, for instance,¹¹¹ the National Education Assessment (NEA) is used to monitor the national performance of the education system at basic level (i.e. up to entry to senior high school). The NEA tests competence in English and mathematics at P3 and P6, sampling performance across Ghana. Reporting an analysis of test scores in 2005 and 2007, the Government noted a decline in scores at P3 and believed it was linked to large increases in enrolment; whereas performance in English and mathematics at P6 had improved slightly. Overall, the evaluation concluded that too few students were achieving a minimum level of competence (a score of 35%) and a 'very low' percentage of students was achieving proficiency (a score of 55%). The NEA also permits comparison of performance across regions and districts. Alongside this, Ghana developed the School Education Assessment (SEA) (also via the Ghana BECAS Project/EQUIP2). This test measures how well students can demonstrate they have met core objectives in English and mathematics at grades 2, 4 and 6. Rather than producing comparisons of school performance, the tests are designed to allow schools to see where improvements are needed in specific areas of the curriculum. The results are not meant to measure student achievement, but parents can be provided with information about the outcomes through School Performance Appraisal Meetings led by circuit supervisors.

The AED Grail Project¹¹² explored how information from the SEA, NEA and EMIS, together with information such as teacher attendance and curriculum coverage, might be used to produce school report cards. These would give a snapshot of the performance of a school and compare it with the performance of other schools. The aim was to trigger action by education service officials, headteachers, teachers and parents. (Separate examinations control entry to senior high school and university on a competitive basis: the Basic Education Certificate Examination and West African Senior Secondary Certificate Examination respectively.)

Testing system and pupil performance together

Some countries use testing to evaluate and improve both system and student performance: in Australia,^k students in Years 3, 5, 7 and 9¹¹³ (i.e. aged 8–9, 10–11 and 12–13) are tested against national literacy and numeracy benchmarks.^l The data is used for reporting to parents and schools as well as for systemic review and evaluation. There is further sampling of student performance in science (Year 6), civics and citizenship (Years 6 and 10) and ICT (Years 6 and 10). (Additionally, the national Government specifically reports on performance in international benchmarking studies such as PISA and TIMSS. Separately, almost all Australian states have external assessment at the end of Year 12 to certify completion of school education and rank students for entry to tertiary education.) In Victoria, the benchmark testing is preceded by reading and writing tests for all five-year-olds in their first month at school. (Presumably, this establishes a threshold for assessment of subsequent value added.)

In Chile, the SIMCE (Education Quality Measurement System) is taken by students in the 4th grade at the end of each year and students in, alternately, the 8th and 10th grades. It tests maths, Spanish, social and natural sciences. Its aim is to enable teachers to review and improve their teaching practice; parents to gain additional information about their children's performance and know how their school is performing relative to others; education authorities to design incentive programmes (e.g. financial bonuses for all teachers at the best-performing schools) and plan teacher training programmes; research centres to consider issues affecting education quality in Chile and the Ministry to offer advice and support.¹¹⁴

Other systems blend assessment for different purposes at different points in students' careers; allocating lead responsibility accordingly. In Slovenia, in primary and lower secondary (6–15 years of age), in-year formative assessment is carried out by the teachers. At the end of the second cycle there are optional national examinations and at the end of the third cycle, compulsory national examinations testing mother tongue, maths and a third subject. In upper secondary school (15–18 years of age), formative and summative assessment is carried out by the relevant subject teacher. However, in order to complete upper secondary school, students must pass either the *Matura* or the Vocational *Matura*. National examinations are organised by the National Examinations Centre in conjunction with other bodies to ensure consistency. Results are evaluated by

^k All children in government schools and almost all in non-government schools participate.

^l This forms part of the National Literacy and Numeracy Plan.

the National Education Institute, which then proposes changes including amendments to national subject syllabuses.¹¹⁵ The outcomes of the analyses of external assessments are used for external evaluation purposes and are fed back to schools to enable them to make improvements to their ways of working.

Following its education reforms, Poland introduced similar assessment arrangements designed on the one hand to provide parents and students with information on attainment and progress and on the other, to use external examinations to diagnose the achievements and shortcomings of students' education, to assess the effectiveness of teaching and ensure consistency of qualifications.¹¹⁶

In introducing radical changes to its education system, Bahrain established a National Examinations Unit which developed examinations designed to draw on best practice and test students' capacity to think rather than simply to recall. It was recognised¹¹⁷ that the questions would be unfamiliar and challenging. However, taken in conjunction with the school reviews, the examinations enabled the Authority to identify, for instance, poorer performance at Grade 6 mathematics than at Grade 3.

7 Evaluation: what works?

Setting up and operating monitoring and evaluation systems is not only expensive, but it also carries a cost in terms of resources and skilled personnel that might have been used elsewhere. In its guidance on inspection, the UK's Office of Public Services Reform¹¹⁸ advised that inspection should not only provide assurance about the safe and proper delivery of services and contribute to their improvement, but also deliver value for money. It also recognised that inspection could be discontinued or replaced by other approaches as the context changed.

So, although countries with high-performing and improving education systems, as well as those with aspirations towards improvement, are introducing or adapting monitoring and evaluation systems, is the investment worthwhile?

Difficulties of assessing impact and limited evidence

Assessing impact is difficult because evaluation is often part of a package of measures designed to improve school performance. Hopkins,¹¹⁹ for instance, cited the UNESCO EFA report's praise for Chile's investment in enhancing the quality of its education system as measured by improvements in outcomes, attributing the country's success to a number of interlocking initiatives. These included tightening up accountability (on the part of schools) and the monitoring and publication of examination results as a tool to support parental choice. But it also included changes to attendance patterns, the curriculum and raising standards for headteachers.

A number of writers have pointed out the limited research in this area. Ehren and Visscher¹²⁰ talk of a 'dearth' of research into the effects of school inspection and the 'conditions which may facilitate improvement' as do Blok et al.,¹²¹ noting that many of the studies chose instead to focus on the experiences and opinions of the schools and teachers involved.

It is easiest to see the effect of relatively simple evaluation of quantitative measures. For instance, following World Bank investments in education¹²² including evaluation at classroom, school and country level in Tunisia, Akkari¹²³ reported that institutional accountability had led to measurable improvements in student retention, though there was still room for progress in terms of using the examination system as a mechanism for monitoring progress or the inspectorate as an improvement rather than compliance mechanism.

What works?

External evaluation and inspection

Surveys of the outcome of school inspections are not definitive, with some pointing to school improvement, others to teacher stress and declining examination results.¹²⁴ Viewed in a rather general way and based on the impact on test scores in a sample of Dutch primary schools, there is some limited evidence¹²⁵ to suggest that 'inspections do no harm, but seem to have little or no effect on student performance'. However, other writers,¹²⁶ have found evidence that parts of the inspection process contribute to school improvement, including the requirement to develop an action plan; and oral and written feedback from the inspectors (particularly if given in private and fitting in with the school's culture).

It is necessary to distinguish between a school's willingness to respond to the outcome of external evaluation and its ability to do so. While transparency (i.e. publication of results) or sanctions (e.g. loss of funding, loss of pupils) appears to make schools more likely to respond, actual acceptance and implementation of recommendations appear to be most successful where they are in keeping with the school's beliefs, approaches and priorities. Conversely, if the school and the inspectors disagree about the school's performance, attention is likely to focus on the credibility of the inspection process itself rather than on school improvement.

Ehren and Visscher¹²⁷ cited studies by Janssens as well as by Matthews and Sammons to suggest that schools producing self-evaluation documents for use in external inspections may not wish (not surprisingly) to report on risks and weaknesses.

Some systems try to reduce this risk and strengthen schools' understanding of standards by delegating a member of the school's staff to join the inspection team: for example, in Hong Kong,¹²⁸ principals are seconded to the external review teams. Moreover, the seconded heads have benefited from the experience which has 'illuminated the nature of weaknesses as well as leading-edge transformative practice... [bringing] new insights on what leadership can achieve through the building of resilient self-evaluation cultures'. A similar approach has now been adopted in England, even though the headteacher is not a member of the inspection team.¹²⁹

Except for the strongest schools, external support is required to help the staff think about how to adopt changes; to implement new teaching and learning processes; to adjust the new ways of working to the particular school; and to involve school staff so that changes include grass-roots developments rather than simply imposed solutions. In a report on one, admittedly limited, study Ehren and Visscher¹³⁰ considered that the 'assumption that the inspection of schools alone will automatically improve schools is naive'. The average school found it difficult to make more than simple changes without help and that 'inspecting schools without follow-up and monitoring activities is probably not very effective'. Two factors appeared to be linked to school improvement: a) where school inspectors giving feedback about poor performance made appointments to follow up with the school; and b) where unsatisfactory assessments in the school report were coupled with feedback that gave clear advice about what to do, as well as agreements about improvements to be carried out within a specified period. Ehren and Visscher warned against limiting the role of external inspectors in these circumstances, particularly since internal boards (governing bodies) may lack the ability to assess and control quality.

Self-evaluation

While there is evidence that self-evaluation can set the scene for school improvement by improving professional practice and increasing the focus on pupil achievement, there does not appear to be conclusive evidence that, in itself, it will raise student outcomes. It is also dependent on a number of school characteristics, such as a positive attitude to change, as well as the commitment of the principal. Small-scale and larger-scale, long-term (five-year) studies found that schools lacked the expertise to design and carry out valid self-evaluations; to analyse the causes of underperformance; or to develop and implement remedial measures.¹³¹

Evidence from research in Hong Kong and the Netherlands suggested that a combination of external and internal review can lead to improvements in classroom teaching, better planning for improvement, improved professional dialogue between teachers and increased willingness to engage with evidence. However, other important factors include an overall focus on assessing student progress, as well as willingness to make a long-term commitment to implementation and substantial investment in support and training.¹³²

External examinations

Similar issues appear in respect of external examinations, which are very challenging to design, implement and follow up when they are supposed to achieve a range of purposes. Drawing particularly on studies of the publication of Chile's SIMCE system, the first barrier was access to the results (with levels of access deteriorating from director to teacher); then accessibility (i.e. the capacity to interpret the results), which showed a similar deterioration; then acceptance, with, not surprisingly, poorer validity attributed to SIMCE outcomes by schools with declining results. There were also issues about professional development capacity. As a result, the authors concluded that in the case of teachers and directors, the presentation of the reports, combined with weaknesses in professional development, limited the usefulness of SIMCE for improving teaching and other school activities.¹³³

While research focusing on the impact of large-scale testing carried out by the Office of Educational Quality and Accountability in Ontario found a number of issues evident elsewhere, it also found evidence of positive impact on teachers' planning and practices in elementary schools linked to the grade 3 assessment process. Teachers said that taking part in marking had been 'one of the best professional development experiences of their careers'.¹³⁴

The impact of publication

Stakeholders, whether as politicians, taxpayers or parents, have a right – and a need – to know how their investment in education is doing.

At a simple level, there is also some evidence that publishing data is linked to improvement. For instance, the OECD reported¹³⁵ that, even when factors such as the socio-economic background of students is taken into account, there is still a significant positive association between schools making achievement data public and having stronger results. OECD's report¹³⁶ on the 2003 PISA test in terms of the link between school accountability, autonomy and choice and equity of educational achievement concluded the results suggested these factors are 'tides that lift all boats'.

Inevitably contextual and cultural differences are important. However, evidence from reviews of the results of publication^m suggests that the practical difficulties of accessing and understanding the outcomes of evaluation mean the impact can be limited. In the case of Chile's SIMCE, for instance:¹³⁷

- a. Parents saw fewest problems with SIMCE. They were worried by two issues: the tendency to label schools good or bad without justification and the illegal exclusion of low-achieving students from SIMCE testing. However, there are other issues: first of all, parents did not necessarily have access to the information;ⁿ secondly, they found interpretation of the (two-page) reports difficult; thirdly, few (14%) took account of SIMCE scores in choosing their child's school, with social make-up, instead¹³⁸ being a

^m (of the outcomes of external reviews and examinations)

ⁿ Taut et al. (2009) reported less than a third of parents surveyed had received the report on SIMCE outcomes for their child's school.

driver of choice (and there is doubt about the extent of choice, particularly for poorer parents); finally, parents were unlikely to risk affecting their relationship with their child's teacher by using the SIMCE report to demand improvements.

- b.** Whereas the pedagogical (TPU) and administrative (director) school leaders had access to the 70-page school report, it was not necessarily passed on to teachers. Similarly, TPUs and directors remembered more about their school's SIMCE results than teachers did. Greatest recall (and dissemination of the generally more positive results) occurred in the private, non-subsidised schools. Teachers tended to remember and interpret the reports incorrectly. Although the reports were meant to drive improvement, in fact, most schools undertook work in preparation for the examination. However, they were not able to interpret the outcomes sufficiently well to use them to achieve school improvement. Schools' concerns included the general failure to take account of the make-up of the student intake as well as students' lack of motivation, since the results had no implications for them personally.

Similarly:

- a.** A review which included an evaluation of the impact of reporting on the outcomes of large-scale assessment programmes in Ontario recognised the need to ensure that it included an explanation of the factors affecting student achievement to avoid crude and inaccurate conclusions.¹³⁹
- b.** Other literature reviews¹⁴⁰ have questioned whether publication of inspection findings would lead to school improvement. Evidence from the Netherlands suggested that parents there were more interested in the school's atmosphere, teaching methods and its reputation, rather than in information about educational quality.

8 Conclusions

Quality assurance is part of an overall system

Quality assurance, whether for accountability, compliance or improvement, cannot exist in a vacuum. Several writers¹⁴¹ have identified mechanisms needed to ensure consistency of delivery and of expectations. These include curriculum frameworks with specified learning outcomes against which students can be measured; appropriate performance indicators for institutions which match goals for individual students; and careful benchmarking, including against international standards.

Looking more widely at the context, lack of impact is attributed by several writers¹⁴² to failures to ensure that the system overall is geared up to carry out and benefit from effective evaluation. Relevant factors include not only the professional capacities of headteachers and teachers, but also arrangements at system level for coordinating the outputs of evaluation and providing guidance and support to schools so they know how to act on the information flowing from it.

Conclusions

This chapter has looked at models of quality assurance in high-performing and improving education systems. It has also looked at evidence about approaches to quality assurance, drawing out the judgements, where applicable. Rather than repeating these conclusions in detail, the balance of this section summarises the main messages.

There will be continuing government interest in external evaluation and examinations, both because of the drive for international competitiveness (seen as dependent on improved student outcomes) and for accountability purposes.

Delegation of some responsibilities to school level is, legitimately, coupled with a requirement to ensure that national policies are delivered; to evaluate the impact of policies; to compare the performance of schools and school systems; to account for public funds; and to ensure that value for money is achieved.

Delegation is also coupled with specification about inputs, processes and outputs, including levels of resourcing and teacher qualifications; pedagogy and management; the curriculum; standards and targets. What is specified varies, depending on the resources available and the homogeneity of the system.

Depending on the context, the options available are as follows:

- a. External evaluation/inspection is justifiable on the grounds of public accountability and shows what the state regards as important. It is probably most effective in terms of school improvement when headteachers are actively involved in the development and

subsequent implementation of the model, linking it into self-evaluation and ensuring it is in keeping with the values and priorities of teaching staff.

- b.** Self-evaluation is generally valued by headteachers and teachers. However, if it is to contribute to improvement, its implementation requires extensive training and support. It is not a substitute for external evaluation, though it can be given a progressively greater role over time. However, hasty moves to reduce external supervision can lead to a fall in standards. School staff may lack the skills. There may also be a conflict of interest on the part of members of governing bodies which may damage their ability to be objective.
- c.** Examinations/tests: the most realistic and effective approach is probably to assign formative assessment to teachers. Where examinations for accountability and improvement purposes are separated from those for individual student assessment, it may be possible to reduce costs through sampling. However, students may be less committed to doing their best, since the results will not affect them.

Inadequate support, both during evaluation and in the follow-up to it, is one of the most significant risks, irrespective of the option(s) selected.

Clarity about aims, together with information about resources and context, is a prerequisite for decisions about the roles of external evaluation, institutional self-evaluation and examinations.

This clarity about overall educational aims will also help to ensure that the information collected for monitoring purposes is appropriate for policy-making, supports improvement at school level and is neither excessively burdensome nor costly.

Explicit attention is needed to ensure that external evaluation, self-evaluation and examinations are sensibly linked together at national, regional and institutional levels. Overall, the system design needs to enable 'intelligent accountability'¹⁴³ that achieves a 'dynamic balance' between meeting the need for accountability whilst increasing educational quality.

Evaluation systems cannot be transported wholesale from one state to another, though lessons can be learned. The challenge is to develop approaches that are not only fit for purpose, but take account of the human, technical and financial resources available, as well as the cultural context. Akkari,¹⁴⁴ for instance, while identifying the need to develop further evaluation in Tunisia, warned of the risk of trying to 'import or transfer models of educational reform' noting that 'Tunisia, like other nations of the South, tends to import models that are not always compatible with its economic, social and cultural context'. There are never sufficient resources available to implement the ideal model: the temptation to spread resources thinly has to be resisted: it leads to mechanistic approaches.

Not only will it take time to develop appropriate knowledge and skills; there also needs to be scope for the system to evolve. For instance, the aims may move on from a drive for universal access to primary education and an emphasis on compliance to successful benchmarking against international standards and continuous improvement. As understanding of standards and quality becomes more embedded, the intensity of external evaluation and prescription may be reduced. High-performing and improving systems, such as in Finland and Singapore, plan and implement with a consistent, steady and long-term perspective.

Similarly, international benchmarking, including tests, can be used to identify issues for consideration and monitor trends over time, if used with discrimination. For instance, one aspect of accountability is about the efficient and effective use of public funds. In part, this is an audit role – including benchmarking expenditure levels either nationally, across comparable groups of schools or internationally. The other part is examining the impact achieved in comparable circumstances.

Parents, teachers and the community, as well as the wider public and politicians, seek a range of outcomes from evaluation. They need to be involved, to the extent that is realistic, in the development of quality assurance systems, so that their knowledge and experience is brought to bear, and the outcomes are relevant and comprehensible, promoting a shared understanding of the purposes of education and engagement with it.

Finally, despite well-founded reservations, publication of the results of evaluation is justified, not only on accountability grounds, but also because it appears to be linked with improved student outcomes. However, there needs to be realistic expectations of the motivations, capacity and expertise of different audiences – whether parents, teachers, politicians or the media – and appropriate investment in meeting those needs.

References

- ¹ Akkari, A. (2005) 'The Tunisian educational reform: from quantity to quality and the need for monitoring and assessment', *Prospects*, 35, 1, 59–74.
- ² De Grauwe, A. (2001) *School supervision in four African countries: Volume 1. Challenges and reforms*. Paris: International Institute for Educational Planning/UNESCO [online]. Available: <http://unesdoc.unesco.org/images/0012/001248/124823e.pdf> [7 November 2011].
- ³ Faubert, V. (2009) *School evaluation: current practices in OECD countries and a literature review* (OECD Education Working Papers No. 42) [online]. Available: <http://dx.doi.org/10.1787/218816547156> [7 November 2011].
- ⁴ Harris, R. (2007) 'Why ranking schools would do more harm than good.' Paper presented to the Australian Education Union Federal Executive Meeting, Melbourne, 12-13 September.
- ⁵ Office of Public Services Reform (2003) *The Government's policy on inspection of public services*. London: Office of Public Services Reform [online]. Available: <http://www.roadmap.depkeu.go.id/Ina/Files/pustaka002.pdf> [7 November 2011].
- ⁶ Mayer, D.P., Mullens, J.E., Moore, M.T. and Ralph, J. (2000) *Monitoring school quality: an indicators report*. Jessup, MD: US Department of Education, National Center for Education Statistics, Office of Educational Research and Improvement [online]. Available: <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001030> [7 November 2011].
- ⁷ Faubert, *op.cit.*
- ⁸ Organisation for Economic Cooperation and Development (OECD) (undated) *Improving school leadership: country background report for Chile*. Paris: OECD [online]. Available: <http://www.oecd.org/dataoecd/13/11/39052630.pdf> [7 November 2011].
- ⁹ Ghana Education Service (2008) *Report on the development of education in Ghana*. Geneva: UNESCO International Bureau of Education [online]. Available: http://www.ibe.unesco.org/National_Reports/ICE_2008/ghana_NR08.pdf [7 November 2011].
- ¹⁰ UNESCO International Bureau of Education (2008) *Country report: Indonesia. Indonesian public policies on inclusive education*. Geneva: UNESCO International Bureau of Education [online]. Available: http://www.ibe.unesco.org/National_Reports/ICE_2008/indonesia_NR08.pdf [7 November 2011].
- ¹¹ Bialecki, I., Johnson, S. and Thorpe, G. (2002) 'Preparing for national monitoring in Poland', *Assessment in education: principles, policy & practice*, 9, 2, 221–236.
- ¹² Straus, M. (2005) 'International comparisons of student achievement as indicators for educational policy in Slovenia', *Prospects*, 35, 2, 187–198.
- ¹³ Hopkins, D. (2006) *Quality assurance and large scale reform: lessons for Chile*. Synthesis report from the international seminar on 'Regulatory models and quality assurance systems', Santiago, Chile, December 2006 [online].
- ¹⁴ Hopkins, *op.cit.*; Leithwood, K., Jantzi, D. and Mascall, B. (2002) 'A framework for research on large-scale reform', *Journal of educational change*, 3, 1, 7–33.

- ¹⁵ De Grauwe, A. and Naidoo, J.P. (eds) (2004) *School evaluation for quality improvement: an ANTRIEP Report*. Meeting of the Asian Network of Training and Research Institutions in Educational Planning (ANTRIEP), Kuala Lumpur, 2–4 July. Paris: UNESCO [online]. Available: <http://unesdoc.unesco.org/images/0013/001398/139804e.pdf> [7 November 2011].
- ¹⁶ Eurydice (2009a) *Key data on education in Europe 2009*. Brussels: Education, Audiovisual and Culture Executive Agency [online]. Available: http://eacea.ec.europa.eu/education/eurydice/documents/key_data_series/105EN.pdf [7 November 2011].
- ¹⁷ Eurydice (2009b) *Organisation of the education system in the Netherlands 2008/09*. Brussels: Education, Audiovisual and Culture Executive Agency [online]. Available: http://eacea.ec.europa.eu/education/eurydice/thematic_reports_en.php [14 August 2013].
- ¹⁸ Hopkins, *op. cit.*, Luginbuhl, R., Webbink, D. and de Wolf, I. (2009) 'Do inspections improve primary school performance?' *Educational evaluation and policy analysis*, 31, 3, 221–137.
- ¹⁹ Eurydice (2009b), *op. cit.*
- ²⁰ MacBeath, J. (2008) 'Leading learning in the self-evaluating school', *School leadership and management*, 28, 4, 385–399.
- ²¹ MacBeath, *ibid.*
- ²² Hopkins, *op. cit.*
- ²³ MacBeath, *op. cit.*
- ²⁴ OECD (2010) *PISA 2009 results*. Paris: OECD [online]. Available: <http://www.oecd.org/pisa/pisa2009keyfindings.htm> [30 August 2013].
- ²⁵ Bialecki et al., *op. cit.*
- ²⁶ Eurydice (2010) *National summary sheets on education systems in Europe and ongoing reforms: the Netherlands [online]*. Available: http://eacea.ec.europa.eu/education/eurydice/thematic_reports_en.php [14 August 2013].
- ²⁷ Pitman, A. (2007) 'Ontario, Canada: the state asserts its voice or accountability supersedes responsibility.' In: M.T. Tatto (ed.) *Reforming teaching globally (Oxford studies in comparative education)*. Didcot: Symposium Books.
- ²⁸ De Grauwe and Naidoo, *op. cit.*
- ²⁹ De Grauwe and Naidoo, *op. cit.*; UNESCO International Bureau of Education (2006a) *World data on education, 6th edition, 2006/07, New Zealand*. Geneva: UNESCO International Bureau of Education [online]. Available: http://www.ibe.unesco.org/fileadmin/user_upload/archive/Countries/WDE/2006/ASIA_and_the_PACIFIC/New_Zealand/New_Zealand.pdf [7 November 2011].
- ³⁰ Eurydice (2009c) *Structures of education and training systems in Europe: Slovenia 2009/10 edition [online]*. Available: http://eacea.ec.europa.eu/education/eurydice/thematic_reports_en.php [14 August 2013].
- ³¹ Eurydice (2009d) *Structures of education and training systems in Europe: Poland 2009/10 edition: [online]*. Available: http://eacea.ec.europa.eu/education/eurydice/thematic_reports_en.php [14 August 2013].

- ³² Ministerial Council for Education, Early Childhood Development and Youth (2009) *MCEETYA Four-year plan 2009-12*. Carlton, South Victoria: MCEETYA [online]. Available: [http://www.mceecdya.edu.au/verve/_resources/MCEETYA_Four_Year_Plan_\(2009-2012\).pdf](http://www.mceecdya.edu.au/verve/_resources/MCEETYA_Four_Year_Plan_(2009-2012).pdf) [7 November 2011].
- ³³ International Review of Curriculum and Assessment Frameworks (2007) *INCA summary profile: education in Australia* [online]. Available: http://www.inca.org.uk/Curriculum_documents_and_guidance.pdf [14 August 2013].
- ³⁴ Hopkins, *op. cit.*
- ³⁵ UNESCO International Bureau of Education (2006b) *World data on education, 6th edition 2006/07: Canada*. Geneva: UNESCO International Bureau of Education [online]. Available: <http://ddp-ext.worldbank.org/EdStats/CANwde07.pdf> [7 November 2011].
- ³⁶ UNESCO International Bureau of Education (2006a), *op. cit.*
- ³⁷ Eurydice (2009b) *op. cit.*
- ³⁸ Hopkins, *op. cit.*; Barber, M., Mourshed, M. and Whelan, F. (2007) 'Improving education in the Gulf', *McKinsey Quarterly*, March [online]. Available: http://www.mckinseyquarterly.com/Improving_education_in_the_Gulf_1946 [7 November 2011].
- ³⁹ Halinen, I., Koivula, P., Kyro, M., Sarlin, H-M. and Volmari, K. (2008) *The development of education: national report of Finland by the Finnish National Board of Education*. Geneva: UNESCO International Bureau of Education [online]. Available: http://www.ibe.unesco.org/National_Reports/ICE_2008/finland_NR08.pdf [7 November 2011].
- ⁴⁰ Halinen et al., *ibid.*
- ⁴¹ Council of Ministers of Education, Canada with Canadian Commission for UNESCO (2008) *The development of education: reports for Canada*. Geneva: UNESCO International Bureau of Education [online]. Available: http://www.ibe.unesco.org/National_Reports/ICE_2008/canada_NR08_en.pdf [7 November 2011].
- ⁴² Hopkins, *op. cit.*
- ⁴³ Malinowski, M. (2008) 'Education reform conflicts continue in Chile', *Santiago Times*, 1 July [online]. Available: <http://upsidedownworld.org/main/chile-archives-34/1359-educational-reform-conflicts-continue-in-chile> [7 November 2011].
- ⁴⁴ Eurydice (2008) *Organisation of the education system in Slovenia 2008/09* [online]. Available: http://eacea.ec.europa.eu/education/eurydice/documents/eurybase/eurybase_full_reports/SI_EN.pdf [7 November 2011].
- ⁴⁵ De Grauwe, *op. cit.*
- ⁴⁶ MacBeath, *op. cit.*
- ⁴⁷ De Grauwe, *op. cit.*; Bialecki, I., Johnson, S. and Thorpe, G. (2002) 'Preparing for national monitoring in Poland', *Assessment in education: principles, policy & practice*, 9, 2, 221–236.
- ⁴⁸ Richards, C.E. (1988) 'A typology of educational monitoring systems', *Educational evaluation and policy analysis*, Vol. 10, No. 2 (Summer, 1988), 106–116.
- ⁴⁹ Faubert, *op. cit.*

- ⁵⁰ De Grauwe, *op. cit.*
- ⁵¹ Faubert, *op. cit.*
- ⁵² De Grauwe, *op. cit.*; Richards, *op. cit.*
- ⁵³ Richards, *op. cit.*
- ⁵⁴ Faubert, *op. cit.*; Ehren, M.C.M. and Visscher, A.J. (2008) 'The relationship between school inspections, school characteristics and school improvement', *British journal of educational studies*, 56, 2, 205–227.
- ⁵⁵ The Quality Assurance Authority for Education & Training, Bahrain (2009) *Annual report 2009* [online]. Available: <http://en.qaa.bh/AnnualReportEn2009.pdf> [7 November 2011].
- ⁵⁶ De Grauwe, *op. cit.*
- ⁵⁷ Faubert, *op. cit.*
- ⁵⁸ Volante, L. (2007) 'Educational quality and accountability in Ontario: past, present and future', *Canadian journal of educational administration and policy*, 58, 21 January [online]. Available: http://umanitoba.ca/publications/cjeap/articles/volante_educational%20_quality.html [7 November 2011].
- ⁵⁹ Faubert, *op. cit.*
- ⁶⁰ Harris, *op. cit.*
- ⁶¹ Bialecki et al., *op. cit.*
- ⁶² Visscher, A.J. (2001) 'Public school performance indicators: problems and recommendations', *Studies in educational evaluation*, 27, 199–214.
- ⁶³ Scheerens, J. and Hendriks, M. (eds) (2004) 'Benchmarking the quality of education', *European educational research journal*, 3, 1, 101–399.
- ⁶⁴ Richards, *op. cit.*
- ⁶⁵ Hopkins, *op. cit.*
- ⁶⁶ Ministry of Education of the P.R. China, Chinese National Commission for UNESCO (2008) *National educational development report to the 48th session of the international conference on education: inclusive education in China* [online]. Available: <http://www.un.org.cn/public/resource/a1de113bf9d7fcf2ae98300485002aa6.pdf> [7 November 2011].
- ⁶⁷ UNESCO (2009) *Education indicators: technical guidelines* [online]. Available: <http://www.uis.unesco.org/Library/Documents/eiguide09-en.pdf> [7 November 2011].
- ⁶⁸ Hopkins, *op. cit.*
- ⁶⁹ Straus, *op. cit.*
- ⁷⁰ Office for Standards in Education, Children's Services and Skills (2011) *Conducting school inspections*. London: Ofsted [online].
- ⁷¹ Council of Ministers of Education, Canada with Canadian Commission for UNESCO, *op. cit.*
- ⁷² Trucano, M. (ed.) (2006) *Education management information system: a short case study of Ghana* (InfoDev Working Paper No 4) [online]. Available: www.infodiv.org/en/Document.502.pdf [7 November 2011].

- ⁷³ Kauko, J. and Varjo, J. (2008) 'Age of indicators: changes in the Finnish education policy agenda', *European educational research journal*, 7, 2, 219–231.
- ⁷⁴ Eurydice (2009b) *op. cit.*
- ⁷⁵ Ministry of Education, Singapore (2009) *The desired outcomes of education* [online]. Available: <http://www.moe.gov.sg/education/files/desired-outcomes-of-education.pdf> [30 August 2013].
- ⁷⁶ British Columbia Ministry of Education (2010) *2009/10 Annual service plan report*. Victoria, BC: Ministry of Education [online]. Available: http://www.bcbudget.gov.bc.ca/Annual_Reports/2009_2010/educ/educ.pdf [7 November 2011].
- ⁷⁷ Government of Hong Kong Education Bureau (2010) *The role of lay members in quality assurance* [online]. Available: <http://www.edb.gov.hk/index.aspx?nodeID=743&langno=1> [7 November 2011].
- ⁷⁸ Faubert, *op. cit.*
- ⁷⁹ Straus, *op. cit.*
- ⁸⁰ Schmidt, W.H., Houang, R. and Shakrani, S. (2009) *International lessons about national standards*. Washington DC: Thomas B Fordham Institute [online]. Available: <http://www.edexcellence.net/publications-issues/publications/international-lessons-about.html> [7 November 2011].
- ⁸¹ Schmidt et al., *ibid.*
- ⁸² Straus, *op. cit.*
- ⁸³ Luginbuhl et al., *op. cit.*
- ⁸⁴ Council of Ministers of Education, Canada with Canadian Commission for UNESCO, *op. cit.*
- ⁸⁵ Council of Ministers of Education, Canada with Canadian Commission for UNESCO, *op. cit.*
- ⁸⁶ Ministry of Education, Singapore (2011) *Recognising school achievements in 2010* [online]. Available: <http://www.moe.gov.sg/media/press/2010/09/masterplan-of-awards-2010.php> [7 November 2011].
- ⁸⁷ De Grauwe, *op. cit.*
- ⁸⁸ Bialecki, *op. cit.*
- ⁸⁹ Eurydice (2009b), *op. cit.*
- ⁹⁰ Richards, *op. cit.*
- ⁹¹ Chung, C-L. (2009) *GRAIL Project: Ghana school report card findings including EMIS support assessment* [online]. Available: http://pdf.usaid.gov/pdf_docs/PNADQ833.pdf [7 November 2011].
- ⁹² Ministry of Education of the P.R. China, Chinese National Commission for UNESCO, *op. cit.*
- ⁹³ Faubert, *op. cit.*
- ⁹⁴ Akkari, *op. cit.*
- ⁹⁵ Eurydice (2009a), *op. cit.*
- ⁹⁶ Eurydice (2008), *op. cit.*

- ⁹⁷ De Grauwe, *op. cit.*
- ⁹⁸ Eurydice (2009b), *op. cit.*
- ⁹⁹ MacBeath, *op. cit.*
- ¹⁰⁰ Ofsted, *op. cit.*
- ¹⁰¹ Hopkins, *op. cit.*
- ¹⁰² Hopkins, *op. cit.*
- ¹⁰³ Eurydice (2009e) *Organisation of the education system in Poland: 2009/10* [online]. Available: http://eacea.ec.europa.eu/education/eurydice/thematic_reports_en.php [14 August 2013].
- ¹⁰⁴ De Grauwe and Naidoo, *op. cit.*
- ¹⁰⁵ Victoria Auditor General (2007) *Improving our schools: monitoring and support*. Melbourne, Victoria: Victoria Auditor General.
- ¹⁰⁶ De Grauwe, *op. cit.*
- ¹⁰⁷ Eurydice (2009f) *Structures of education and training systems in Europe: Finland: 2009/10* [online]. Available: http://eacea.ec.europa.eu/education/eurydice/thematic_reports_en.php [14 August 2013].
- ¹⁰⁸ OECD, *op. cit.*
- ¹⁰⁹ De Grauwe, *op. cit.*
- ¹¹⁰ De Grauwe, *op. cit.*
- ¹¹¹ Ghana Education Service *op. cit.*
- ¹¹² Chung, *op. cit.*
- ¹¹³ International Review of Curriculum and Assessment Frameworks, *op. cit.*
- ¹¹⁴ Taut, S., Cortés, F., Sebastian, C. & Preiss, D. (2009) 'Evaluating school and parent reports of the national student achievement testing system (SIMCE) in Chile: access, comprehension, and use', *Evaluation and program planning*, 32, 2, 129-137.
- ¹¹⁵ Eurydice (2008), *op. cit.*
- ¹¹⁶ Brzdak, J., Donska-Olszko, M. and Jaworski, P. (2006) *Assessment project country report: Poland* [online]. Available: <http://www.european-agency.org/agency-projects/assessment-in-inclusive-settings/assessment-in-inclusive-education-files/poland-indexed-report.doc/view> [7 November 2011].
- ¹¹⁷ The Quality Assurance Authority for Education & Training, Bahrain, *op. cit.*
- ¹¹⁸ Office of Public Services Reform, *op. cit.*
- ¹¹⁹ Hopkins, *op. cit.*
- ¹²⁰ Ehren, M.C.M. and Visscher, A.J. (2008) 'The relationship between school inspections, school characteristics and school improvement', *British journal of educational studies*, 56, 2, 205-227.
- ¹²¹ Blok, H., Slegers, P. and Karsten, S. (2008) 'Looking for a balance between internal and external evaluation of school quality: evaluation of the SVI model', *Journal of education policy*, 23, 4, 379-395.

- ¹²² World Bank (2000) *Project appraisal document on a proposed loan in the amount of US\$99 million to the Republic of Tunisia for the first phase of the Education Quality Improvement Program (EQUIP) May 30, 2000*: (World Bank Report No 20467-TUN). Washington, DC: The World Bank [online]. Available: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2000/07/07/000094946_00061705502666/Rendered/PDF/multi_page.pdf [7 November 2011].
- ¹²³ Akkari, *op. cit.*
- ¹²⁴ Ehren and Visscher, *op. cit.*; MacBeath, *op. cit.*
- ¹²⁵ Luginbuhl et al., *op. cit.*
- ¹²⁶ Ehren and Visscher, *op. cit.*
- ¹²⁷ Ehren and Visscher, *op. cit.*
- ¹²⁸ MacBeath, *op. cit.*
- ¹²⁹ Ofsted, *op. cit.*
- ¹³⁰ Ehren and Visscher, *op. cit.*
- ¹³¹ Blok et al., *ibid*; Schildkamp, K., Visscher, A. and Luyten, H. (2009) 'The effects of the use of a school self-evaluation instrument', *School effectiveness and school improvement*, 20, 1, 69–88.
- ¹³² Blok et al., *op. cit.*
- ¹³³ Taut et al., *op. cit.*
- ¹³⁴ Volante, *op. cit.*
- ¹³⁵ OECD (2007) *PISA 2006: Science competencies for tomorrow's world: executive summary* [online]. Available: http://www.oecd.org/document/2/0,3343en_32252351_32236191_39718850_1_1_1_1,00.html#Vol_1_and_2 [7 November 2011].
- ¹³⁶ Schutz, G., West, M.R. and Wobmann, L. (2007) *School accountability, autonomy, choice, and the equity of student achievement: international evidence from PISA 2003* (OECD Education Working Papers, No 14). Paris: OECD [online]. Available: <http://www.oecd.org/dataoecd/53/57/39839422.pdf> [7 November 2011].
- ¹³⁷ Taut et al., *op. cit.*
- ¹³⁸ OECD (undated) *Improving school leadership: country background report for Chile*. Paris: OECD [online]. Available: <http://www.oecd.org/dataoecd/13/11/39052630.pdf> [7 November 2011].
- ¹³⁹ Volante, *op. cit.*
- ¹⁴⁰ Ehren and Visscher, *op. cit.*
- ¹⁴¹ Leithwood, *op. cit.*; Hopkins, *op. cit.*
- ¹⁴² Hopkins, *op. cit.*; De Grauwe, *op. cit.*; De Grauwe and Naidoo, *op. cit.*
- ¹⁴³ Hopkins, *op. cit.*
- ¹⁴⁴ Akkari, *op. cit.*

About CfBT Education Trust

CfBT Education Trust is one of the world's leading not-for-profit education companies, providing a range of education services in the UK and internationally. Established over 40 years ago, CfBT has an annual turnover of over £100 million and over 2,000 staff worldwide, all working on projects relating to educational quality.

Our clients include major international organisations such as the World Bank and the European Union, together with government ministries worldwide. Currently we are managing important projects in the UK, Dubai, Abu Dhabi, Oman, Brunei, Cambodia, Laos, Australia, Jamaica and Kenya. We have a particular expertise in the evaluation of school quality. In England we manage school inspections for the whole of the northern region of the country on behalf of the national inspection agency, Ofsted. Our work for clients also involves teacher and leadership training, curriculum design, careers guidance for young people and support for school improvement.

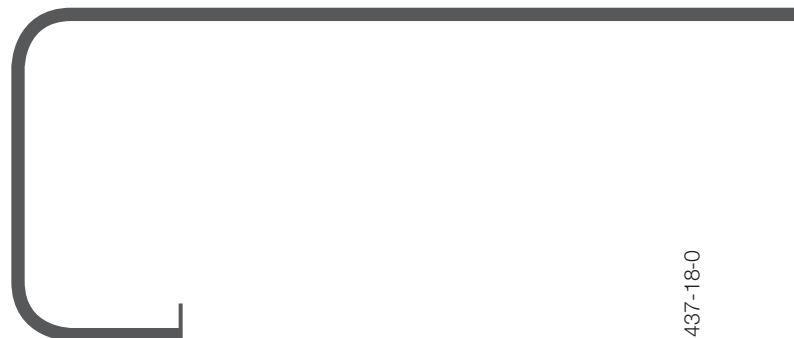
As a not-for-profit company CfBT uses its trading surplus to fund a programme of public research.

Visit www.cfbt.com for more information.

High-performing and improving education systems

In line with its commitment to providing evidence-based and practical support for international education system reforms, in February 2009 CfBT Education Trust commissioned research identifying the critical components of high-performing and improving education systems. The research focuses on: teachers; the curriculum and assessment; quality assurance and accountability; school leadership; and school systems, structures and funding. It examines policy documents, material relating to implementation and academic literature. The challenge was to bring together robust, independently verified evidence and practical knowledge of what is done in effective education systems in an authoritative yet accessible review format.

Systems were selected early in 2010. The process took account of PISA 2000 and 2006 reading scores; PISA 2003 and 2006 mathematics scores; and TIMSS 1995, 1999, 2003 and 2007 (where available) results for mathematics and science for students aged 9-10 and 13-14.



CfBT Education Trust
60 Queens Road
Reading
Berkshire
RG1 4BS

+44 (0)118 902 1000

www.cfbt.com