Leading learning in the self-evaluating school

John MacBeath*
University of Cambridge Faculty of Education, Cambridge, UK

Governments around the world have adopted school self-evaluation as an essential corollary to the local management of schools. Typically this has been accompanied by prescribed frameworks, indicators and scoring systems, tending to promote a form of ritual self-inspection rather than an ongoing and dynamic process of self-evaluation. It is argued that these have derived more from an accountability imperative than an improvement motive and as such have tended to disempower rather than empower schools. This article examines the inherent tensions between external accountability and internally driven school improvement. The example of Hong Kong’s new relationship with schools is used to illustrate how those issues play out in the implementation of large-scale reform, drawing parallels and contrasts with Ofsted’s New Relationship with Schools in England. The responsiveness of Hong Kong’s Education Development Bureau to evidence from independent research carries lessons for other countries in which such evidence has been ignored or marginalised by policy-makers. It is, in many respects, an exemplary case study of change management.

Keywords: self-evaluation; inspection; review; leadership; learning

Introduction

School self-evaluation has, from the beginning, been an important research and development strand within the Leadership for Learning network in Cambridge. It was one of the inducements for schools in seven countries to take part in the Leadership for Learning (Carpe Vitam) Project described elsewhere in this issue. In each of those countries, encompassing the US, Australia, Austria, Greece, Denmark and Norway as well as England, school leaders were struggling with the tensions between policy directives and a desire to stay true to the educational values which brought them into the profession. Although in each of these disparate cultures ‘leadership’ was understood and practised in very different ways, it was seen as the key to the embedding of self-evaluation in classroom and school life. While for the research team the integral connections between leadership, learning and self-evaluation may have been obvious, in the seven participating countries self-evaluation was widely perceived as occupying its own separate world, a form of...
occasional audit and systematic data gathering, often for the benefit of an external inspectorate or review teams.

In many ways it is in England that the tensions between external and internal forms of evaluation have perhaps been at their sharpest with a direct impact on school leadership. In England the Government’s promise of a New Relationship with Schools (DfES 2004a, 2004b) was an explicit recognition that the old relationship was neither economically viable (Gershon 2004) nor demonstrably effective in improving schools. A substantial body of evidence gathered over the previous decade was highly equivocal as to Ofsted’s impact on improvement. Cullingford and Daniels’s 1998 study reported an adverse effect on examination performance for a sample of schools, although dismissed by then Chief Inspector of Schools (Woodhead 1999, 5) as ‘deeply flawed, ineptly executed and poorly argued’. Rosenthal’s (2001) study in the following year, however, also found ‘a significant negative effects of Ofsted visits on school exam performance in the year of the inspection’.

Ofsted visits seem to have adverse effects on the standards of exam performance achieved by schools in the year of the Ofsted inspection. Perhaps the efforts required by teaching staff in responding to the demands of the school inspection system are great enough to divert resources from teaching so as to affect pupil achievement in the year of the visit. (16)

Employment of inspection consultants and rehearsal for the forthcoming event had become an increasingly common feature of school life. A report by the University of Brunel (Jones et al. 1999) referred to ‘anticipatory dread’, impairing normal school development work and effectiveness of teaching, an impact on a school which, it was claimed, could last for over a year.

Such reactions are to be expected in systems where inspection carries high-stakes consequences for teachers. While Ofsted’s strapline is Improvement through Inspection, ex-Chief Inspector David Bell (now Permanent Secretary at the Department for Family School and Community) was ready to admit that inspection does not of itself-improve schools:

I have always been cautious in saying that inspection causes improvement because, frankly, we don’t. But it has to be an important part of our thinking about inspection. You do try to understand what contribution inspection can make to improvement and that is a statutory base of the organisation. It forces us to be more articulate and explicit about that. (in MacBeath 2006, 30)

Ofsted’s own analysis in 2004 found that in some cases schools made greater progress than those that were not inspected, in other cases they did not. Ofsted commented that ‘there is little significance to be read into this except to say that inspection is neither a catalyst for instant improvement in GCSE results nor a significant inhibitor’ (Matthews and Sammons 2004, 37).

The move towards a new relationship between government and schools and between inspection and self-evaluation in England is exemplary of a larger trend. In Europe, the Standing International Conference on Inspection (SICI) has for a number of years argued for a ‘sequential’ approach (Alvik 1996) in which the
school’s own self-evaluation provides the focus and centrepiece for external review and in which initiative lies with school leaders to place self-evaluation at the heart of school and classroom practice. As in Europe, in Asia, in Australia and New Zealand, and in North America there has been a progressive move away from more traditional forms of quality assurance to more school-owned, school-driven forms of evaluation and accountability. In this respect the development of self-evaluation and external review in Hong Kong is exemplary in three critical respects.

One: the Education Development Bureau commissioned longitudinal research to evaluate impact over a four-year period, responding at every reporting stage to evidence, and feeding formatively into the next cycle of development. Two: support and development for leadership at all levels was seen as integral to systemic reform and to the embedding of self-evaluation as a mindset rather than a discrete event. Three: a focus on learning was made explicit through the sharing of self-evaluation tools and the development of an online interactive resource for school management committees, senior leaders, middle leaders (subject ‘panels’), school improvement teams and classroom teachers (http://hk.sitc.co.uk/English2/E-1-1b.html, last accessed 1 June 2008).

The Hong Kong story

In 1997 the Hong Kong Education Department, as it was then, introduced a Quality Assurance framework and a programme of school inspections conducted by officers of the Quality Assurance Division (QAD). The programme of inspections continued until 2003 when feedback internally within the department and from the education sector identified three key areas for improvement:

1. the implementation of a robust system of school self-evaluation (SSE) together with a repertoire of tools, processes and performance measures;
2. equipping schools with key performance measures to help them assess their strengths and areas for improvement;
3. complementing SSE with a process of external school review (ESR) in a timeframe consistent with a school’s developmental cycle.

A letter dated 9 May 2003 from Mrs Fanny Law, Permanent Secretary for Education and Manpower, to all principals stated that a strong school development and accountability framework should be based on systematic internal school self-evaluation (SSE), complemented by external school review (ESR). She ‘strongly believed’ that rigorous and systematic SSE was the driving force for continuous improvement and an internal quality approach whose key purpose was to enhance student learning.

During 2003, documentation to support the new approach was developed together with a training programme for QAD inspectors plus a further training programme for external reviewers (principals seconded from Hong Kong schools to ESR teams). The model was constructed carefully with a watchful eye on international experience.
It involved intelligence gathering from consultants and visits to numerous other countries, in order not to repeat the mistakes of other regimes, identifying what appeared to work well and tailoring policy and practice to the Hong Kong context.

**A substantial database**

Research into the implementation of SSE and ESR was commissioned as an independent study, and carried out by a team from the University of Cambridge and Cambridge Education between 2002 and 2008, with a brief to report back to the Department on the impact of reform at successive stages and make recommendations for improvement (MacBeath and Clark 2004; MacBeath 2008).

For four years, between 2003 and 2007, the research team followed three successive cohorts of schools, complemented by data gathering from school sponsoring bodies, school management committees, parents and parent bodies (PTAs) and external review teams.

The database over the four years of the study encompassed questionnaire responses (quantitative and qualitative) from 32,000 teachers and 203 School Improvement teams in 635 schools, 1009 External Review Team members and 165 external reviewers. A total of 17 half-day focus groups were conducted with school management committees, school sponsoring bodies, principals, School Improvement Teams, classroom teachers, parents, and students from special, primary and secondary schools. Field observations of the ESR process were conducted in 20 schools plus in-depth case studies of 16 schools which had undergone ESR in 2003–2004 and 2004–2005. The overall response rate for questionnaire returns from school was 81%.

The disaggregation of responses between school phases, socioeconomic status and achievement levels revealed a spectrum of differing views on implementation and impact while within-school analyses by stakeholder group exposed systemic differences in perception of purposes, readiness, pressure and value attributed to the experience. High-achieving schools (particularly those well versed in data analysis) were generally better informed, more positive and less reactive than low-achieving schools (some in highly challenging circumstances), while there was a consistent trend for principals to be more informed and more positive than middle leaders, who in turn were consistently better informed and more positive than classroom teachers. Most informed and positive of all were members of School Improvement Teams, a function of their close involvement in, and ownership of, the self-evaluation process.

**Achievements**

The summary headline at the end of the four-year period would depict a progressive embrace of self-evaluation and welcome by schools for external review. There was clear evidence of a deepening understanding of the purposes of self-evaluation and its
contribution to continuous school improvement, much of this owed to the quality of leadership at system and school level. There was an increasing degree of comfort with the use of the data and evidence base, and a greater degree of openness to the views of students and parents.

The final Impact Study (MacBeath 2008) identified the following achievements:

- a deepening understanding and heightened confidence of school staff in relation to SSE and ESR;
- classroom teaching becoming more engaging, learning-centred, and open and receptive to student voice;
- a welcome for the insights of ESR teams and setting of clear agendas for improvement following external review;
- the enhanced skills of ESR teams in conducting review;
- sharing of thinking and practice by teachers beyond the classroom in a whole-school dialogue;
- a growing willingness to engage with evidence, to move from impressionistic evaluation of quality and performance to a more systematic, rigorous and informed approach to assessing practice.

Nonetheless, for many of the schools within this pilot group there is some distance still to travel in genuinely embedding self-evaluation, while the roll-out territory wide to all schools is only now beginning. The significance of achievements to date has also to be understood within a four-year compass of a pilot, which met with considerable resistance, not least from the Legislative Council. The experience of the first cohort of 92 schools which took part in the 2003/2004 academic year tells a story of impact in which the reformers’ intention to create a bottom-up culture of self-evaluation was experienced widely by school staff as top down, with the attendant pressure and anxiety that rapid high-stakes change incurs.

In the first cohort three questionnaire items in particular testify to the nature of the impact: ‘ESR did not exert much pressure on me’, ‘ESR did not affect much of my daily duties’ and ‘My school was well prepared for ESR’.

These data both reveal and conceal a story. They tell a story of pressure experienced by a substantial proportion of the teaching population and a widespread diversion of energies from teaching to preparation for external review. This is

| Table 1. Three items from the post external review questionnaire (2003–2004). |
|------------------------------------------|--------|--------|--------|--------|----------|----------|
| ESR did not exert much pressure on me   | Strongly agree | Agree | Disagree | Strongly disagree | Standard deviation |
|                                         | 3.3    | 19.7   | 29.3    | 30.3    | 17.4     | 3.2       |
| ESR did not affect much of my daily duties | 4.1    | 26.5   | 27.6    | 25.4    | 16.4     | 3.2       |
| My school was well prepared for ESR     | 33.0   | 50.8   | 12.4    | 3.0     | 0.4      | 16.6      |

Note: \( n = 3581 \) school staff (including senior and middle leaders).
reflected in the highly positive figures in relation to being well prepared for ESR. As questionnaire open-ended comments and interviews disclosed, this was achieved over a six-month extended period of familiarisation with indicators, performance measures and the amassing (or creation of) documentation.

What is concealed by these data is partially revealed by the standard deviations. These point to a significant range of difference in relation to being ‘well prepared for external review’ (SD = 16.6) while the much smaller SD figure for perceptions of pressure and diversion from duties (3.0) suggests a more commonly held set of opinions. Yet, probing further, disaggregation of the strongly disagree response category reveals a significant depth of feeling on these issues. The statement ‘ESR did not exert much pressure on me’ produced a distribution at school level from 3.0% to 63.4%, around a mean of 17.4%.

In follow-up interviews, focus groups and in-depth case studies, in schools where anxieties were high, pressure stemmed, as one teacher put it, from ‘fear of the unknown’, while the ‘known’, based on war stories from other schools, had generated a siege mentality in which interviewees talked about ‘preparing for battle’. Rumours which circulated at second and third hand inflated the degree of stress and demands of documentation. Words used to describe the prospect of ESR were ‘frightened’, ‘worried’, ‘anxious’, ‘stressed’, ‘confused’. ‘We worried for at least three months to six months’, said one teacher. In another school a teacher claimed she ‘couldn’t sleep in September because of the number of things to prepare for ESR’. Another teacher said she felt so anxious during the initial stage she had lost considerable weight. A common apprehension was ‘not being good enough’, as practice became publicly exposed.

In pursuing the explanation for the differential levels of pressure, anxiety and stress, leadership, learning and their interconnections emerge as the most salient explanatory factors.

The leadership equation

Leading change is expressed at different levels within the system and, most saliently, at the interface between and among those different levels. In the Hong Kong study the locus and impetus for change was sometimes clearly located in influential and conviction-led individuals but more often was difficult to separate out from the flow of activity between school principals and school improvement teams, in exchanges between members of the Quality Assurance Division, external school review teams and with members of School Management Committees.

In many instances it was the leadership of the principal that distinguished schools in which self-evaluation and external review were seen more as an opportunity than a threat. Anxiety, although impossible to dismiss in a highly competitive and high-stakes system such as in Hong Kong, was attenuated by the calm, self-confident modelling of senior leaders, by their demonstrable trust and confidence in their staff, and a progressive embedding of a culture of reflection and inquiry. The contrast
between the steadying influence of these principals and their less measured counterparts was most vividly portrayed by one secondary school principal who kept his staff on a state of alert over the Christmas period, asking them not to take the customary holidays ‘out of town’ but to stay close, ready to be called in at any time. Its effect in cranking up the anxiety level spread like a virus through the school.

A concomitant of the self-confident leaders was the emphasis they placed on learning – at student, professional and organisational level. They were themselves exemplary learning leaders, some having initiated 360 degree forms of appraisal by staff and students. One principal of an all-girls school which had received an outstanding review from the ESR team on every indicator except value-added took immediate issue with the team on their misunderstanding of ‘value’. It was enshrined, she pointed out, in the quality of learning they had witnessed and commented so favourably upon – the moral, spiritual and intellectual growth and the self-assured leadership of so many of the girls.

The impact and differential quality of leadership was also manifest in the approach of School Improvement teams (SITs), varying from largely ineffective to exemplary models of shared leadership. In some cases their responsibilities had been ‘distributed’ to them by the principal, in other cases they had grasped the space and latitude afforded to them to take the initiative and lead creatively. The Impact Study concluded that their efficacy and credibility could be explained by a number of key factors:

- Team membership includes a cross-section of staff with high credibility among their colleagues.
- There was scope to exercise initiative and creativity.
- There was a willingness and capability to ask hard questions and instil an ethos of accountability.
- There was teamwork which synergises the capacities of all its members.
- Initiative and ownership were displayed, which create confidence and shared leadership.
- Vision was evident as to what self-evaluation can achieve and how it can feed into learning and school improvement.

The vitality and compass of a SIT’s endeavours owed much to the qualities of senior leaders. They needed the perspicacity and courage to genuinely share leadership, to provide space and support for SITs to take the initiative, enabling them to take risks while maintaining a sense of accountability to their colleagues. They needed to be aware of how to deploy SITs most creatively, ensuring that the composition of the team included people with credibility among their colleagues, inviting participation by newly appointed teachers as well as experienced staff. There had to be tangible opportunities for them to support their peers in school-wide improvement with a focus on learning at the centre of their endeavours.

Within a framework of School Development and Accountability these senior leaders and SIT members implicitly grasped the difference between contractual,
professional and moral accountability (Becher and Eraut 1981), managing the sensitive balance among them, and rendering unto their political masters that which did not compromise the accountability they owed to staff and to students. The principal who argued against the corruption of ‘value-added’ contrasted a contractual, essentially accountancy perspective with a professional and moral accountability.

Passionate leaders with a strong sense of the professional and moral, feisty individuals prepared to argue their corner, were most often those most highly regarded by the Quality Assurance Division (QAD), their schools deliberately chosen to exemplify the leading edge of practice. In the leadership for learning equation the influence of the QAD team was a key factor. Theirs was, and continues to be, a presence in the wings, working to create the stage on which leadership flourishes at every level.

In this respect the secondment of principals to external review teams has been one of the strongest planks in developing capacity at system, school and individual level. Scrutiny of evidence, observation in classrooms, conversations with teachers, students and parents has, for these external reviewers, illuminated the nature of weaknesses as well as leading-edge transformative practice. They have been required to engage in rigorous analysis of what they have seen and heard, bringing a new respect for evidence, a skill in winnowing it out from surface behaviour and surface learning, bringing new insights on what leadership can achieve through the building of resilient self-evaluation cultures.

The leadership of ESR teams is also part of the interlocking jigsaw. Team leaders play a vital role in supporting and inducting seconded external reviewers into teams. They draw on the skills of managers, counsellors and arbiters, inter-personal skills which come to the fore when there is conflict within the team or between team members and staff. At these critical moments a steadying influence is required. Shadowing students, observing classroom practice, examining students’ work, interviewing parents, managing meetings, negotiating grades, writing reports, giving formative feedback, each brings challenges, often a steep curve of professional development and accountability. These require a skilful balancing act. External reviewers have to be authoritative while recognising the authority of school staff. They have to listen sensitively but give their own opinion. They have to support but also challenge. These are all skills which feed back powerfully into leading their own schools and into the tripartite learning agenda – student, professional and organisational.

Leading learning has been to the fore at each stage of the four-year study. It has been a central element in the ongoing dialogue with the Education Development Bureau, with staff of the Quality Assurance Division, with schools invited to seminars and workshops to share practice, with presentations to School Management Committees and with the Education Panel of the Legislative Council and the press. Inter-school visits and networking have expanded horizons while there is virtual access to practice through the online interactive resource (OIR), still being expanded to incorporate a specific focus on classroom learning. Most powerful of all are video
clips of students talking critically about the quality of learning and teaching in their schools – a remarkable cultural shift within a relatively short space of time. The following excerpt from that OIR resource, an 11-year-old talking confidently in English about her learning, is there to be used a stimulus for discussion among school staff, of leadership issues and how, as a collective, to create an environment for learning.

When we get older, teachers don’t tell us the answers. They ask questions, and let us find the answers by ourselves. We may surf the Internet, and go to the library to find some books. When we do it in this way, we can learn how to learn. We will be more interested in the things we learn. We also think that if teachers just stand in the classroom and talk and talk, it will be so boring. This is active education and we like it very much. We keep on learning, maybe just watch the news, it’s so useful for our learning. We may care (about) our world, our home Hong Kong, and China, and to know more about our society. (Lui Chiu Yee, Kylie, Plover Cove Primary School)

The Online Interactive Resource was one of many channels designed to keep the dialogue open, sharing practice, tool building, progressively remodelling and redefining the SSE/ESR relationship. This resulted in significant modifications including no longer requiring schools or ESR teams to use a four-scale numerical rating on the 14 Performance Indicator (PI) areas; withholding the uploading of ESR reports to the website for public access; recognising the selective reporting in the media, often to the detriment of the schools’ reputation; drastically reducing the number of school documents to be prepared for ESR. This series of responsive measures also included streamlining of reporting processes, the development of an e-Platform to enhance systematic data management, an online data-collection tool and revised more user-friendly versions of stakeholder surveys.

Lessons from large-scale reform

The introduction of school self-evaluation in Hong Kong followed a pattern now familiar to countries around the world, and Hong Kong could never be accused of failing to monitor practice elsewhere. In common with their counterparts in Singapore, New Zealand, Australian states, Canadian provinces and, in recent years, Nordic countries, the gift of government to schools has been a framework, a set of indicators, documentation requirements and the protocols and for implementing and measuring change. The Hong Kong nomenclature – the School Development and Accountability Framework – signalled the two key purposes of self-evaluation as improvement and accountability, an arranged marriage which has served to produce as many tensions in Hong Kong schools as it has elsewhere in the world (Pang 2004; Nicholls and Berliner 2007; Bajunid 2008).

As research into implementation with the first cohort of schools revealed, the tensions between these two driving logics created resistance from senior and middle leaders and from teachers who viewed the impending visit as essentially designed to call the school to account. Typically it resulted in immediate diversion of energies
from teaching to the amassing of documentation. After the event, there was a widely voiced complaint that the one-size-fits-all formula failed to take account of the unique contexts of schools and the character of the communities they served.

These are familiar themes. They speak loudly to impatient policy-makers who thrive on standardisation. They resonate with politicians for whom the horizon of electoral life is too close to allow the long-term growth that a genuinely self-evaluating system would entail.

In the English context diversion of energies, as seen from a student perspective, was captured in a Hertfordshire school where a group of secondary students conducted their own study of inspection (Dannawy 2001). The students reported a tenser relationship with their teachers, special lessons being rehearsed beforehand, and students and teachers having to be constantly ‘on show’, ever ready for the inspector’s visit. ‘Trouble students’ were sent away to outdoor pursuits centre to partake in a week-long alternative education programme. Students also wrote ‘Teachers are too busy being stressed’; ‘Some of them have no time to teach, they are so busy getting ready’; ‘Everyone is telling us what to say and how to act. What is this dictatorship? Are we expecting Stalin or Hitler next week?’.

The introduction of shorter and sharper Ofsted inspections with very little advance notice was in part designed to address this issue but has in the process lost much of the vitality of self-evaluation approaches that existed prior to adoption and standardisation by the New Relationship. With a prescribed formula, fed by performance data and attended by the scrutiny of a School Improvement Partner, it is difficult to reconcile with a more organic bottom-up epidemiology of improvement (Rogers 1962; Gladwell 2000; Hargreaves 2003).

While there are clear parallels in England with Hong Kong’s own new relationship, a signal strength of the latter administration is the credibility given to the voices of key stakeholders, mediated through the findings of the Impact Study. Before the rolling out to a second cohort, and then again before embarking on a third cohort, key findings were accepted and key recommendations implemented, including the abandoning of the four grades for a more qualitative profiling. Losing the numbers also helped to lessen the resentment of the review team’s prerogative to raise or lower the school’s own grades.

While the withdrawal of a four-level grading system may seem on the face of it a minor concession, one has only to examine the deep impact that comes from grading schools and the simplistic labels attached to those grades. The more nuanced language of quality has, in a number of countries, ceded to a numerical discourse in which a 1 or a 4 is a defining statement of worth. The Ofsted term ‘satisfactory’ (accompanying a grade 3) has been the subject of endless debates in England, as being satisfactory was clearly not satisfactory enough. The original Scottish terminology of major weaknesses, major strengths, more strengths than weaknesses and more weaknesses than strengths carried a different kind of message about the balance of evidence, and allowed greater scope for seeing and negotiating judgement. However, that more nuanced approach has given way to the simpler one-word descriptors to which policy-makers appear addicted.
It is into the large middle ground of ambiguity, a rating of 2 and 3, that most inspection judgements fall. The nature of those middle ground judgements does, however, depend to a great extent on where you sit, what you see, what prior ways of seeing you bring to your judgement and the context in which that judgement is made. An inspector’s rating of a lesson made in half an hour in a classroom is likely to differ from a student’s long-term perspective or that of the classroom teacher. While such variance is obvious and well documented, top-down evaluation schema often fail to recognise this. Bypassing those differences misses the very heart of the process, that is, the discourse which such mismatches engender. Exploring such dissonances of view lies at the very heart of self-evaluation and of dialogue (‘dia logos – meaning flows through it’).

The parallels with grades and marks given to the quality of children’s work are salutary. As research has shown (Kluger and DeNisi 1996; Black and Wiliam 1998a, 1998b), summative grades can be a disincentive to learning and inhibit the very goals teaching aspires to achieve. While embracing assessment for learning, governments are at the same time failing to address the tensions between formative assessment and a competitive accountability framework with public reporting of school-by-school performance. In similar fashion self-evaluation for learning is not easily reconcilable with high-stakes inspection.

In this respect Hong Kong has invested heavily in retraining of external review teams recognising that a new relationship requires a shift from an inspection mindset to a review approach. It conveys the message that in this brave new relationship no longer will inspectors be the sole arbiters and narrators of the school’s story but rather they come as mediators, encouraging and supporting teachers to speak for themselves.

**Self-evaluation or self-inspection?**

In well-developed inspection regimes there is a compelling logic for governments to devolve their own frameworks to schools, providing the goals, criteria and protocols so that schools engage in a form of self-inspection. Therein lies a paradox, however. It appears that the more governments provide the template the less inventive and spontaneous the process at school and classroom level becomes. Self-evaluation all too easily becomes a ritual event, a form of audit in which senior leaders assume the role of an internal inspectorate applying a set of common criteria arising from quite differently held assumptions about the nature of accountability and improvement. Self-evaluation is centred on capacity-building. It understands the iterative relationship between classroom life and school life, and between school learning and out of school learning. It recognises that students’ learning and teachers’ learning are integrally connected and that teachers’ learning stems from, and feeds into, organisational learning. It is this complexity and dynamic that is the missing ingredient in ritualised and formulaic approaches to self-evaluation, the box-ticking and form-filling that makes it such an onerous and tedious process for teachers and
school leaders. Grasping the complexity and dynamic of school as a living growing
entity is what Arnold Tomkins, a New York administrator, wrote about over a
century ago:

The organisation of the school must be kept mobile to its inner life. To one who is
accustomed to wind up the machine and trust it to run for fixed periods, this
constantly shifting shape of things will seem unsafe and troublesome. And
troublesome it is, for no fixed plan can be followed; no two schools are alike; and
the same school is shifting, requiring constant attention and nimble judgement on
the part of the school leader. (1895, 4)

Keeping a school mobile to its inner life is what self-evaluation is about, a continuing
process of reflection that becomes implicit in the way people think and talk about
their work. It is a process in which teachers construct their own knowledge by
‘surfacing tacitly held data about classroom life’ and exploring conditions conducive
to learning (Guskey and Hubermann 1995, 253). This implies a paradigm shift from
a passive and compliant role to an active role in which teachers are the prime movers
in self-evaluation and take charge of their individual and collective professional
development. A sense of agency is fostered within a supportive climate in which
teacher colleagues and senior leaders continually maintain a focus on learning (see
Frost in this issue), making practice explicit and discussable.

The school’s story develops from the quality of dialogue that occurs in the
classroom, in staffrooms, in formal meetings, in ‘the essential conversations’
(Lawrence-Lightfoot 2003) that take place at various levels within and between
schools, with parents, with other agencies and with authorities at local and national
level. It is a slow-burn developmental process for which the system owes teachers a
significant measure of support and critical friendship.

Internal accountability

Richard Elmore (2005) makes an important distinction between internal and
external accountability. Internal accountability describes the conditions that precede
and shape the responses of schools to pressure that originates in policies outside the
organisation. The level or degree of internal accountability is measured by the degree
of convergence among what individuals say they are responsible for (responsibility),
what people say the organisation is responsible for (expectations), and the internal
norms and processes by which people literally account for their work (accountability
structures). Elmore concludes that with strong internal accountability schools are
likely to be more responsive to external pressure for performance.

Internal accountability, moral and professional, implies an openness to dialogue,
to the nature of evidence, a form of self-evaluation that is genuinely embedded in
teachers’ thinking and day-to-day practice. Opening up of practice to colleagues
whose intentions are to learn, rather to judge, removes – or at least attenuates –
anxiety and pressure. It both rests on and engenders trust. When there is a measure
of professional trust it is possible for there to be mutual support, a relationship in
which people experience a genuine intention on the part of the other to help without a hidden agenda, without a sense that support come with caveats and some form of payback. This is why critical friendship is best exercised when there are no hidden agendas, no differential power bases, and it is not premised on accountability rather than improvement (see Swaffield in this issue). When there is intelligent internal accountability and the critical support of a trusted adviser schools are likely to respond more positively to external pressure, confident in the knowledge that they have a rich and unique story to tell, one which rises above the mean statistics and pushes against prevailing orthodoxies of competitive attainment.

However far removed culturally and geographically, these issues resonate with many other countries. For example, Malaysian educators have witnessed the loss of the vitality that had characterised previous practice, sacrificed on the altar of their own version of the new relationship. As Bajunid (2008) reports, the impetus has been lost in new waves of legislation that have failed to spot the inherent professional capital. In the rush to modernise and bureaucratisation political leaders failed to build on the cultural legacy. He argues for a ‘reclamation of the intellect’, a graphic metaphor for a process in which a terrain has fallen into disuse and needs to be rebuilt from the ground up. This implies, of course, that leaders are comfortable with risk and ambiguity, that they are more interested in learning than in outcomes, that they trust teachers and students to work their own magic in the classrooms. It also implies that leadership is distributive rather than distributed, arising from many different sources.

Writing in a Canadian context Ben Jafaar (2006) describes the tensions between economic bureaucratic accountability and ethical professional accountability. These can, she argues, be addressed by ‘inquiry-based accountability’. In this model evaluation at classroom, school and external levels is used in each case as an entry point for professional discussions about learning experiences, opportunities and outcomes and the priorities for the young people that schools are expected to serve.

The history of self-evaluation in many countries shows that it has often had its roots in inquiry and teacher-led research, driven by a natural desire for evidence, the professional ground-bed for improving practice. In many countries (England, Germany, Australia, Singapore, Thailand, Hong Kong) self-evaluation preceded the interest and blandishments of policy-makers. Individual schools or clusters of schools initiated their own process, typically through partnership with a university, school district or commune. These self-initiated approaches often had a vitality and drive because they were ‘owned’ by schools themselves but also suffered because there was no wider networking or systemic support. The lesson for governments is to provide that systemic support, to nurture those home-grown systems while enhancing their efforts by providing facilitative (not mandated) frameworks and appropriate tools and strategies. DfES guidelines insist that the SEF (the Self Evaluation Form) is ‘NOT [their capitals] self-evaluation’ yet it is clearly a message that has not permeated to headteachers or to inspection teams themselves as many previous grounded initiatives have been shelved to accommodate the SEF formula (MacBeath 2006).
As the Hong Kong experience shows, and as we know from those schools in the UK that have not abandoned their commitment to self-evaluation from the bottom up, this is the place where leadership and learning meet. Improvement takes place when learning is centre stage, when there is a learning culture in a school and when heads and senior leadership team are lead learners. It is in the process of building such a learning culture that self-evaluation takes root and external review, or inspection, plays a valuable supportive and challenging role.

Note on contributor

John MacBeath was elected to the Chair of Educational Leadership at the University of Cambridge in 2001. He is the director of ‘Leadership for Learning: the Cambridge Network’ and has written extensively on school leadership, school improvement and self-evaluation. He has acted in a consultancy role to organisations such as the OECD, UNESCO, the European Commission, DCSF, the Scottish Executive, the Swiss Federal Government, and the Hong Kong Education Development Bureau.

References


Dannawy, Y. 2001. ‘Should we sugar coat the truth then miss?’ Unpublished M.ED paper, Faculty of Education, University of Cambridge.


