Chapter 2
Assessment for Learning: Research and Policy in the (Dis)United Kingdom

Mary James

2.1 Assessment Policy in the Context of Education Reform

It is now more than two decades since the UK Government brought in the 1988 Education Reform Act. Its far-reaching powers were designed to create a social market in schooling in England and Wales, which, its Conservative architects believed, would raise standards across the state system. In essence the Act had three linked components. First, it made a commitment to open access to enable parents to choose schools for their children. Secondly, it proposed arrangements for local financial management whereby funds would follow pupils and thus allow successful schools to attract more pupils whilst unsuccessful schools would contract and even close. Thirdly, a new common national curriculum and assessment system would provide parents with a basis for choice because what pupils were expected to learn at various key stages, and how successful schools were in teaching them, would (supposedly) be more transparent. In 1992 the creation of a new framework for inspection through the establishment of the Office for Standards in Education (Ofsted) enhanced this accountability function further.

Although the Act applied only to England and Wales, some aspects were incorporated into policy in Northern Ireland. Scotland, which has always had a separate and distinct education system, watched developments with interest but distanced itself from them. In May 1997, when the Labour Party won the first of three consecutive General Elections, a policy of "devolution" was inaugurated. This entailed the progressive transfer of powers for areas such as health and education from the government's traditional base in Westminster, London, to the new Parliament in Scotland, and to the Assemblies in Northern Ireland and Wales. The argument was that this promotes local control and democracy, although some critics believed it to be wasteful and divisive. This policy of devolution continues under the Conservative

M. James (✉)
Faculty of Education, University of Cambridge, Cambridge, UK
e-mail: mej1002@cam.ac.uk

and Liberal Democrat "Coalition" Government that emerged after the General Election in May 2010.

The development of assessment policy and, particularly, the aspect that is now known as Assessment for Learning (AfL) needs to be understood against this background because a key feature of the following discussion is the ways in which policy has been variously interpreted, developed and implemented, over time, in the four countries of the UK.

2.2 Initial Influences on Policy with Regard to Assessment for Learning

In England and Wales policy discussion about the purposes of assessment was stimulated when the task of designing a framework for national assessment and testing was given to an expert group set up in 1988 and chaired by Paul Black, an academic researcher. The brief of the Task Group on Assessment and Testing (TGAT) was to devise a system to be both "informative" and diagnostic. By the time the group came to report, these two aspects had had become four distinct purposes: formative, diagnostic, summative and evaluative. The system that TGAT proposed was built on four principles; it would be criterion-referenced, formative, moderated and designed to promote progression. The involvement of teachers would be central and group moderation would help them develop common judgments that would be used for reporting purposes but also, crucially, to enable them to plan next steps in teaching. Unfortunately, these proposals received a hostile reception from some prominent politicians and academics. Margaret Thatcher, the then Prime Minister, perceived it to be a subversion by a left-wing "educational establishment", and some academics, including members of the British Educational Research Association (BERA), saw it as a Trojan horse of the political right. So, starting with the idea of group moderation by teachers, the proposals were rapidly dismantled and few recognizable features remain. The system that was eventually put in place focused on increased testing for summative and, especially, evaluative purposes – i.e., to provide performance tables of aggregated results to judge the effectiveness of teachers, schools, local authorities and the system as a whole, in relation to increasingly challenging numerical performance targets. In a rueful reflection, published almost 10 years after TGAT, Paul Black wrote: "With hindsight, it was naïve to imagine that the government, with its commitment to a market where choice would be guided by tests, would support a complex approach." (Black, 1997, p. 41).

With the political imperative to put in place an accountability system based on tests, where even school inspections would take school performance measured in this way as their starting point, the ideas embedded in TGAT about the possibilities of creating a system to meet formative and summative purposes in combination, with the formative purpose uppermost, were almost lost. The community of academic researchers was consumed by the need to react to, and critique, rapidly developing policy based on frequent whole cohort testing, and it largely failed, at that time, to

engage effectively with policy-makers or to clarify the ideas and practices that might offer alternative educational solutions to genuine public concerns about standards of teaching and learning in schools and how to improve them.

A turning point came when Paul Black and Dylan Wiliam, of King's College London, published a pamphlet, *Inside the black box,* derived from an extensive review of research on assessment and classroom learning (Black & Wiliam, 1998). This was commissioned by the Assessment Reform Group (ARG), a group of assessment researchers from all four countries of the UK. The ARG was initially one of several policy task groups set up by BERA in 1989, on the initiative of the then President, John Elliott, who, in response to changes being brought in by the 1988 Act, argued for, "a radically different conception of the primary aim of educational research; namely, to promote worthwhile change by influencing the practical judgements of teachers and policy-makers" (Elliott, 1990, p. 11).

The work of the ARG is described by Richard Daugherty (2007, p. 145) as having two phases. The first, from 1989 to 1995, was characterized as active engagement with policy issues drawing on research evidence and the experience of education professionals to critique policies that were already being implemented. But the second phase, from 1996 to 2010, when the ARG formally "retired", has been more strategic by attempting to influence the policy agenda itself. The commissioning of the review of research by Black and Wiliam, supported by funds from the Nuffield Foundation, was the first step in this more strategic phase.

The central thesis of the *Inside the black box* review was that there is a body of firm evidence that formative assessment is an essential feature of effective pedagogy and its development can raise standards. Moreover, Black and Wiliam put a figure on the size of measured gains and pointed to effect sizes in the range of 0.4–0.7, amongst the largest for any educational intervention. It was probably these figures, and extrapolations that indicated what they might mean in terms of scores on national tests and examinations, and in international surveys of achievement, that encouraged policy-makers to take notice, especially at a time when early gains on the national tests were beginning to level off. However, there was still confusion about what the term "formative assessment" actually meant and, in line with ARG's second phase strategic goal to express issues in a clearer and simpler language, the Group decided to adopt a distinction between assessment *for* learning and assessment *of* learning as a more accessible (less technical) version of the formative/summative distinction. Although the Group continued to debate the wisdom of this – because of similar scope for misinterpretation and confusion – the evidence indicates that the simpler language has, at least, encouraged widespread use. When in 1999 the ARG followed *Inside the black box* with another pamphlet, *Assessment for Learning: Beyond the black box,* and then, in 2002, with a graphically designed poster, *Assessment for Learning: 10 principles* (ARG, 2002a), the new term, and its definition on the poster, was rapidly taken up by policy-makers and practitioners in England and Wales, and elsewhere, including countries beyond the UK. The key event was probably the decision by the Group to find the means to distribute the "principles" to all schools in England, Wales and Northern Ireland.

The next section of this chapter explores the different ways in which the concept of assessment for learning has been interpreted and incorporated into policy in the four countries of the UK. At a time when educational research was criticized for having very little impact on UK policy formation and development (Hargreaves, 1996; Hillage, Pearson, Anderson, & Tamkin, 1998), it is remarkable that developments in assessment for learning provide clear evidence of interactions between research and policy, even if these relationships have not always been straightforward or unproblematic.

2.3 Current Manifestation of AfL Policy in the Four Countries of the UK

At the time of writing, what is most marked is divergence between England and the three "Celtic Fringe" countries: in AfL policy, in assessment policy more generally, and in other aspects of broader education policy. The explanations for this are complex – historical, geographical, economic, political, ideological, cultural and pragmatic – so the ambition of this section is simply to provide a brief description and modest interpretation of some recent AfL initiatives in the four countries of the UK. However, special mention needs to be made of the relative size of these countries. With a population of over 62 million people, England is at least ten times larger, in these terms, than any of the other three countries. This has a profound impact on the ways in which different sections of the education community communicate and interact. There is no doubt that this is more difficult in England, which in the following account, will be dealt with last.

2.3.1 Scotland

Scotland has always taken pride in an education system that is different from that in England. The structures of schooling, the curriculum, the examination system, inspection and the recruitment of teachers have all been developed and managed independently of its neighbour over the border. Scotland was quick to respond to the ideas and evidence on formative assessment/assessment for learning although, characteristically, it started to develop its own distinctive policy. Thus, in 2002, Scotland began setting up a national programme – Assessment is for Learning (AifL) – in which formative assessment is considered as part of a whole assessment system, including pupils' records, personal planning, system monitoring and school evaluation, as well as formative and summative assessment at the classroom level. The intention has been to develop a coherent system for all schools in Scotland that brings together assessment *for* learning, assessment *as* learning and assessment *of* learning. The stated philosophy is to give "considerable freedom to schools and teachers to develop practice within their own context at a pace and in a manner that is suited their needs" (Scottish Government, 2005, p. 3).

Three features of the development of the formative assessment strand are notable: (1) the provision of funding to support teachers' involvement in developing the projects within the programme, (2) support for the participation of university academics (from Scotland and England, especially Dylan Wiliam and Paul Black) which was a source of assurance that the ideas were supported by research evidence, and (3) the fact that the AifL programme was designed using research on transformational change, in particular the work of Senge and Scharmer (2001). The idea was to encourage teachers to engage with the underpinning research and the experience of teachers from England who had worked on the Nuffield Foundation funded King's, Medway, Oxfordshire Formative Assessment Project (KMOFAP) (Black, Harrison, Lee, Marshall, & Wiliam, 2003) and to try out, adapt and develop strategies suited to their own context. Recognizing the failure of previous initiatives in Scotland to embed ideas permanently in the system, those working in AifL in Scotland saw building a sustainable learning community for AifL as a major priority. Thus AifL sought to build learning communities across the education system. In each of the 32 education authorities in Scotland, communities of teachers, headteachers and local authority coordinators worked with national policy-makers, HMIE and researchers from across Scotland. The Research Report (Hayward, Spencer, & Simpson, 2006) on the Scottish AifL initiative notes how important teachers found the opportunities to share their practice through observation and discussion. However, while there was a general impression of successful implementation by those involved in the project, there was a reluctance by some teachers to engage with theories of learning to understand why the strategies "worked" to enhance learning. There was little movement from a pragmatic to a principled rationale. Moreover, there was little evidence that teachers had passed over responsibility for formative processes, especially deciding learning goals, to the pupils themselves. Ownership was still very much with the teachers.

However, by 2008 there was evidence, at least in some local authorities that this was changing. For example, a research report on the formative evaluation of a project designed to explore how teachers were bringing together ideas from AifL with the new curriculum innovation in Scotland, Curriculum for Excellence (see: http://www.LTScotland.org.uk/curriculumforexcellence/ Accessed 22 January 2011). In addition, a second report commissioned by the Scottish Qualifications Authority (SQA) found evidence of teachers developing assessment for learning approaches in the context of high stakes assessment (Standard grade, Intermediate and Higher: all post 16 examination classes). The Highland Council, which was the focus of this study, had taken a particular approach characterized by a professional development programme that encouraged strong and consistent engagement with principles of assessment, that built networks of support, had clear links between assessment, curriculum and learning and teaching, and had contextualized the initiative in a wider policy framework.

Notable, in the Scottish context, is a lack of reference to concerns to raise standards in terms of measured achievements during compulsory schooling. The reason is simple. There are no statutory, standard national tests in Scotland and no league tables of school performance. National monitoring is carried out by means of a

sample survey, rather than whole cohort national testing, so that accountability does not directly drive classroom activity. The first Scottish Survey of Achievement (SSA), a sample monitoring survey of English language and core skills, was carried out by the Scottish Executive in May 2005. Assessment materials from the SSA are used to extend an on-line national bank of assessments that teachers can use, when they judge appropriate, to check their own assessments of pupils' progress. In 2007, AiFL was extended to all schools.

2.3.2 Wales

Following devolution of powers to Wales, and the report of the Daugherty Assessment Review Group (DARG, 2004) which was commissioned to advise on a system for Wales, the Welsh Assembly Government decided to discontinue statutory National Curriculum testing, as carried out in England, and to strengthen teachers' summative assessment as the basis for recording, reporting and accountability. From the school year 2008–2009, schools' responsibilities have been revised in relation to end of key stage teacher assessment arrangements. These now emphasize the importance of internal systems and procedures for standardizing and moderating teacher assessment. For assessments of pupils aged 11 at the end of Key Stage 2, primary schools, and the secondary schools to which they are linked, are required to have effective arrangements in place for cluster group moderation and for transfer of information between primary and secondary schools, in order to increase trust in the reliability of information on attainment. For the assessments of pupils aged 14 at the end of Key Stage 3, secondary schools are expected to supply details of their internal standardization procedures as the first stage in a process of national accreditation of secondary schools (by the Welsh Department for Children, Education, Lifelong Learning and Skills – DCELLS). The DARG Report had recommended the introduction of a limited number of "skills tests" in Year 5 (the year before the end of Key Stage 2 and transfer to secondary school). This proposal has evolved into optional skills assessment materials (not test-based), in thinking, communication and number, piloted by DCELLS in 2008, to support teachers in drawing up a skills profile for Year 5 pupils. Whilst the overall policy thrust stemmed from concerns with over-testing, and the lack of trust in professional judgments that this implied, the new direction has detractors. For example, some teachers, especially those of mathematics and science, have been reluctant to give up the statutory tests, and others fear an increase in workload.

In this broader assessment policy context, the DCELLS carried out a specific development programme for Thinking Skills (TS) and Assessment for Learning, which ran from 2006. This built on the insight that there is much overlap between efforts to develop thinking skills across the curriculum and assessment for learning. Both are interested in the development of metacognition, self-regulation, engagement and autonomy in learners, and ways in which teachers can integrate, or "infuse", TS and AfL strategies into subject teaching. To support the development

of the programme, an advisory group was formed. The author of this chapter was a member, invited because of her membership of the Assessment Reform Group, and her role as director of a development and research project on learning how to learn through AfL (James et al., 2007). Carol McGuinness, from Queen's University Belfast, who was directing a project on the development of thinking skills (ACTS II), was another member (Information on both of these projects can be found at http://www.tlrp.org Accessed 22 January 2011). Carol McGuinness also contributed substantially to the work with schools in Wales by giving talks at the teachers' conferences that were part of the programme. Other support was provided by DCELLS staff and local authority advisers, although the key element was work by teachers, in 42 schools in ten Local Authorities, to develop their practice from the ideas (principles and practices) to which they were introduced.

External evaluation indicated that, in only five school terms, the development programme improved classroom practice and increased the frequency of creative lessons. This was associated with increased learner engagement and improved attainment for all learners, irrespective of perceived abilities. Although evidence of enhanced performance was difficult to discern, because the move from tests to teacher assessment prevented direct comparisons on stable measures, particular improvements in speaking, listening and behaviour were noted. A 3 year extension programme, beginning in late 2008, was therefore embarked upon to ensure that changes in pedagogy are more broadly embedded. The extension included the successful elements of the pilot:

- close partnership working with local authorities
- coaching/mentoring partnerships between DCELLS staff, local authority officers, school senior managers and teachers
- cluster group partnerships, especially between primary and secondary schools
- local and national networks to disseminate good practice
- ownership by local authority and school staff, and some flexibility on implementation so that their pathways reflect local needs
- funded reflection and planning time for practitioners
- monitoring and evaluation by local authorities and DCELLS.

A specific intention was to use the professional networks already established with international researchers and collaborators to enhance the programme and promote its findings.

2.3.3 Northern Ireland

Although, historically, curriculum and assessment in Northern Ireland (NI) has been tied closely to England, this is now changing quite markedly. In 1999, the NI Minister for Education called for a review of curriculum and assessment arrangements, and the resulting revised curriculum is currently being introduced along with

new assessment arrangements (from August 2007). Another important change was a decision in 2001, influenced in part by a devastating critique of the reliability of the 11+ examination (Gardner & Cowan, 2000; but see Gardner & Cowan, 2005 for a more accessible summary), to abolish this selection test for entry to grammar schools, and allow parents and pupils to choose their post-primary schools. One aspect of the new assessment arrangements – the Pupil Profile – is intended to assist them with this choice.

Taken as a whole, the new assessment arrangements are intended to embrace diagnostic, formative, summative and evaluative purposes (i.e., those stated in the 1988 TGAT Report in England). By phased introduction, starting with the first year in each Key Stage in 2007–2008, statutory summative assessment requires every pupil in every year to be assessed by their teachers in (1) areas of learning; (2) cross-curricular skills (communication, using mathematics and ICT); and (3) thinking skills and capabilities (managing information, thinking/problem-solving/decisions, being creative, working with others, self-management). The reliability of these assessments will be assured through teacher moderation. Diagnostic assessment delivered via the Interactive Computerized Assessment System (InCAS) will be used at least once each Key Stage to measure aspects of reading, mathematics and "developed ability". The results of these assessments must be reported to parents and annual parent-teacher meetings arranged to discuss them. An annual pupil profile report will also be produced by the end of May to inform transfer decisions. Alongside all these changes sits AfL to fulfil the formative purpose.

AfL in Northern Ireland represents the "roll out" of a development project begun with 38 primary teachers in 2004 and extended to another 50 primary and post-primary teachers in 2005. The project was described as "action research" in which teachers were encouraged to experiment with aspects of the methodology of assessment for learning and to "adapt the theory and principles of formative assessment to suit their own teaching context and their individual pupils" (CCEA, 2006). The experience of this project has now been distilled into online materials for the Foundation Stage and Key Stages 1, 2 & 3 (See http://www.nicurriculum. org.uk/foundation_stage/assessment/assessment_for_learning.asp for an example. Accessed 22 January 2011) which highlight "five key actions": sharing learning intentions; sharing and negotiating success criteria; feedback; effective questioning; peer- and self-assessment and self evaluation. The emphasis on sharing learning intentions and success criteria reveals the influence of the development consultant, Shirley Clarke, who was involved in live presentations and whose books were provided (e.g. Clarke, 2001); 80% of teachers chose these ideas as their point of departure.

Perhaps this was a good place to start because, for half the teachers in the pilots, the AfL strategies were entirely novel, yet, after a short time positive changes were found. Pupils were described as more confident, persevering and strategic. Teachers were more focused on pupils' needs; they planned for AfL; they were more reflective and had changed their pedagogy. Nevertheless there were concerns over intentions-practice gaps, equity issues i.e. whether the approaches were suitable for all pupils, and the involvement of parents. Moreover there were substantial

implications for support for professional development from senior management of schools. The need to share experience with other teachers and to develop practice over time was thought to be important for sustainability. In other words there appeared to be some resistance to these innovations, which is not surprising since Northern Ireland teachers had been embedded in a summative testing culture for so long.

2.3.4 England

In England, ideas associated with AfL were first taken up by the Qualifications and Curriculum Authority: a quasi-autonomous organization set up with public funds to advise government and to implement aspects of policy on curriculum and assessment. Soon after the publication and distribution by ARG of its *AfL: 10 Principles* poster, QCA requested permission to publish the poster on a book mark and on its website. The Association for Achievement and Improvement through Assessment (AAIA), an association created largely by and for assessment inspectors and advisers in local authorities, also took an early interest and developed materials to support AfL development work with schools (See http://www.aaia.org.uk/afl Accessed 22 January 2011). AfL also became established as an element in the Labour Government's Primary National Strategy (PNS) and the Secondary National Strategy (SNS), which were key components of national policy focused on the development of pedagogy. The National Strategies were managed directly by the Department for Children, Schools and Families (DCSF), the name given in 2008 to the government department concerned with schools, formerly the Department for Education and Skills. These national strategies had the status of guidance – they were not mandatory – but the pressure to comply was considerable, not least through the Ofsted inspection process that expected to see the strategies in operation in schools, or very good reasons why they were not.

By 2009 the Primary National Strategy concentrated on literacy and mathematics, and its 2008 renewed materials had a substantial section on assessment for learning. This section had three sub-sections that shifted the focus progressively outwards from teaching and learning interactions in the classroom, to supportive conditions for learning and then to leadership and management and support. This built on an earlier publication (DfES, 2007a) that revealed a shift towards working more with school leaders in supporting improvement in order to embed AfL practice in classrooms because "the greatest prizes are still to be won".

The Key Stage 3 (lower secondary) section on the DCSF Standards Site contained a suite of Assessment for Learning materials for download, including resources on: AfL in everyday lessons, formative use of summative tests, objective led lessons, oral and written feedback, peer and self-assessment, target setting, securing progression, and questioning and dialogue. A report of an action research project with eight secondary schools (DfES, 2007b) also engaged with the challenges of implementing and embedding AfL practice in schools, acknowledging

that the results of previous efforts had been equivocal, in terms of raising standards as judged by national test and examination results. This SNS report focused on the impact of AfL on pupil learning and standards, and on the leadership and management of change. The findings indicated that "fundamental to developing AfL in the classroom is developing the independent learner and, fundamental to developing the leadership and management of whole school change is developing distributed leadership" (see Chapter 2, Section 2.2, author's emphasis). Curiously, however, given this work sponsored by the DCSF, the renewed SNS Frameworks published in May 2008 made scant use of AfL ideas although they incorporated some reference into an initiative called Assessing Pupils' Progress (APP). The emphasis throughout the new frameworks was on teachers regularly assessing pupils against target levels, and tracking progress.

Both the PNS and the earlier SNS materials made reference to definitions of AfL and research-base accounts of good practice from the Assessment Reform Group, Paul Black and colleagues at King's College London, and the Learning How to Learn project (James et al., 2007). However, the text of these materials, and especially the newer Secondary Frameworks, also revealed tensions with researchers' definitions of AfL (Personal communication suggests that some of the authors were aware of these tensions). For example, the more recent PNS materials referred to "day-to-day assessment" and the SNS materials to "everyday" assessment. This had two contrasting implications. The first was to reinforce the idea that AfL is part of continual interaction and reflection in classrooms, and integral to teaching and learning. But the second implication was that by changing the descriptor to "day-to-day" or "everyday", AfL can be formative, or summative, or both. Politically this was probably unavoidable because the Labour Government in England had invested a great deal in the development of pupil tracking and planning tools, to help teachers and headteachers use the results of statutory national tests for monitoring, prediction and target setting (see for example: http://www.raiseonline.org/About.aspx. Accessed 22 January 2011). However, one can also argue that what was being promoted was no longer formative assessment, as part of pedagogy, for the purpose of enhancing real and lasting learning, but frequent mini-summative assessment to secure higher performance on tests to meet prescribed targets.

The distinction between learning and performance is a subtle one and not well understood. Measured performance should indeed be an indicator of underlying learning (or what Dweck, 2000, calls "mastery") but debates about the validity and reliability of assessments underscore the difficulties of making such assumptions. It is quite possible to drill pupils to perform well on tests without enhancing their learning and, given the high stakes consequences for schools that perform badly, there is increasing evidence that this has happened in England (ARG, 2002b). One possible explanation for the mixed messages that appeared in DCSF documents is that the authors tried to finesse competing claims between those who were convinced by research that formative assessment is the key to improved learning and achievement, and those who still believed that the pressure of regular testing raises standards. Or the mixed messages may simply be indicative of some confused thinking that has elided "learning" and "performance". A publication from the

DCSF (2008), which launched the Labour Government's Assessment for Learning Strategy, raised such questions.

This new AfL Strategy was a very significant new initiative backed by £150 m of government money over 3 years for the professional development of teachers in AfL. It was supported by the DCSF, QCA (which had become the QCDA), the National Strategies and the Chartered Institute of Educational Assessors. The document began by quoting the ARG's definition of Assessment for Learning and incorporated the ARG's ten principles, albeit unattributed and with a different graphic design (Ibid, p. 5). Much of the rest of the text developed ideas based on the ARG definition that: "Assessment for learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (ARG, 2002a). However it also built on the DCSF's own Assessing Pupils' Progress work and its Making Good Progress Pilot (http://www.teachernet.gov.uk/teachingandlearning/schoolstandards/mgppilot/. Accessed 22 January 2011). The idea behind the "Making Good Progress" pilot was to introduce single level tests (SLTs) for teachers to use to check their own judgments, twice a year. On the surface this might look like the Scottish system of banking assessments for teachers to use when they think pupils are ready, but the expectation in England was that standard tests would still be administered to all pupils and the results aggregated, with, at least in the case of the pilot, financial rewards to schools that could show progress of two levels for pupils over a Key Stage. The consultation on the proposed pilot generated concerns that even more testing, which this implies, is not the way forward and more serious effort should be given to developing AfL. However, in its response, the government argued that, by putting AfL together with pupil tracking, formative and summative assessment could be made to work together more effectively. Thus deciding "where the learners are in their learning, where they need to go and how best to get there" had come to mean assessing pupils frequently according to national curriculum levels and sub-levels to track progress, setting new levels as targets and then working (somehow) to attain them. This was not an interpretation that the ARG had in mind when it wrote the definition of AfL although, if "learning" is interpreted as "performance", then the definition is sufficiently ambiguous to make such a reading understandable. A less generous explanation might that the DCSF chose deliberately to appropriate an idea that had gained considerable professional support from teachers in order to take the next pull of the testing lever to meet its performance targets.

Until October 2008, the Labour Government in England seemed unwilling to relinquish any aspect of its testing system as its primary instrument of change. Despite the disastrous experience with missing national test results and poor marking quality in the summer of 2008, it had no obvious plans to rethink the system that was breaking under the strain. In August, Alice Miles (2008), a *Times* columnist, offered her explanation:

> ... in the face of scepticism about the achievements of their Government, these multi-coloured graphs have become the only measure by which they can trumpet their success. This explains the obsession with testing: it has become not a tool of policy, but policy itself.

Then on 14 October, at the height of the financial crisis (a day to bury controversial news – or to save some money?), the Secretary of State for Education, Ed Balls, announced an end to compulsory national tests for 14 year olds (http://www.dcsf. gov.uk/pns/DisplayPN.cgi?pn_id=2008_0229. Accessed 22 January 2011). The proposals were that Key Stage 3 tests would be replaced by improved classroom assessment by teachers and frequent reporting to parents in years 7, 8 and 9. Teacher assessments would be published at national and local authority level only. School Report Cards, along the lines of those used in New York City, would become the mechanism for reporting a wide range of achievements at school level. Standards at age 14, across the education system, would be monitored by national sampling. Assessment for learning was seen to have a role "to help schools use ongoing assessment to drive up pupils' progress". According to the DCSF's Press Notice this involved: "gaining clear evidence about how to drive up an individual pupil's attainment; an understanding between teachers and pupils on what they need to improve; and an agreement on the steps needed to promote individual progress". Given the discussion above, it is worth noting that there was only one reference to pupils' "learning" in the whole four page document.

The announcement of the end of Key Stage 3 tests was widely welcomed by teachers, parents and politicians, including the Opposition parties, and some argued that Key Stage 2 tests should be abolished too. On this matter the Government insisted: "externally marked Key Stage 2 tests were critically important and would continue". However, at a teachers' union conference in April 2009, the then Secretary of State for Education, Ed Balls, said that these tests for 11 year olds were "not set in stone" and that he would work with teachers and parents to reform them, if necessary, after the report, scheduled for May 2009, of an "expert group" on testing made up of headteachers and educational professionals.

Despite this change in policy in England there was little evidence that the underlying rationale had altered significantly. The "drive" was still to raise standards as measured by national curriculum assessment levels; and assessment for learning continued to be seen as an instrument for this purpose. More subtle ideas, about the role of AfL in pedagogy to enhance the learning of capable, resourceful and autonomous citizens in a changing world, seem almost entirely absent.

Recent history of education in England has been characterized by constant change. This has continued. In May 2010, one of the first acts of the incoming Conservative/ Liberal Democrat Coalition Government was to change the name of the DCSF to the Department for Education (DfE). Shortly afterwards it promised what has been popularly called a "bonfire of the quangos" to reduce the levels of government bureaucracy – and to save money in order to reduce the level of national debt. The QCDA was the first quango to be served notice. Within the Department of Education, the National Strategies were also axed. At the time of writing, the future policy profile of AfL is unclear; probably existing material will be placed in an accessible archive for schools to use if they wish. What is more certain is that some form of national testing will continue at the end of the primary phase or at the beginning of secondary schooling and that these results will be published for

accountability purposes. The Coalition Government has asked Lord Bew to conduct an independent review of the effectiveness of the current Key Stage 2 tests, and to report to Ministers in June 2011.

2.4 The Extent of Policy Borrowing

The four countries of the UK are constantly looking at one another's policies to see what they might borrow or adapt, or what they should definitely reject. They each regard their particular context as in some way unique, as indeed they are, so they seek to tailor any "borrowed" policy to their own circumstances. Some countries are more inclined to acknowledge the influence of their neighbours than others. For example Northern Ireland gives Scotland's AifL website as a "useful link".

Exchanges between Scotland, Northern Ireland and Wales, and between researchers and policy-makers there, seem more open than with, and within, England. An illustration of this might be the fact that the ARG felt obliged to hold separate seminars for the dissemination of its projects: always a separate one for England and one or more for the other three countries. Unless meetings are held in or near Westminster it is often difficult to meet key policy-makers in England whereas politicians, civil servants, advisers and researchers meet relatively frequently in the other countries. Of course, England has a much bigger and diverse population, and a much more complex bureaucracy with many layers of decision making, even within the area of education. Differences in opportunities for networking might explain, in part, why policy has diverged between England and the Celtic Fringe. Most significantly for AfL, the government in England, in contrast to Scotland, Wales and Northern Ireland, continues to be committed to an accountability system based on published results of summative tests and assessments. Changes in the political party in power have not changed this overall direction. AfL is therefore pressed into the service of this overarching goal, rather than fulfilling a fundamental purpose of its own. No doubt many policy-makers in England see this hard-headedness as a virtue. After all, there is still no strong evidence yet that when AfL (understood as formative assessment as part of pedagogy) moves to the centre of policy it raises performance standards across the system as a whole. For example, results from the Progress in International Reading Literacy Study (PIRLS) 2006 placed England 19th and Scotland 26th in its distribution table of reading achievement in 40 countries (Mullis, Martin, Kennedy, & Foy, 2007, p. 37). This is unlikely to encourage England's policy makers to adopt Scotland's approach. Although school performance tables, based on tests at Key Stage 3, will no longer be compiled in England, any commitment to national sampling for system monitoring is still a long way from policy in Scotland.

It certainly seems that policy makers in Scotland, Wales and Northern Ireland have been more convinced than those in England by the arguments of researchers that a choice does not have to be made between raising performance and good

learning because with AfL you can have both. These countries have been convinced enough to reject all whole cohort national testing and the publication of league tables in favour of summative teacher assessment combined with formative assessment. One reason for coming to this common view may be that, rather than directly borrowing from one another, they have each drawn directly on the same pool of evidence, mainly from Black, Wiliam and the ARG. England, as we have seen, has made reference to these sources, but the three other countries have made more use of the researchers themselves as *partners* in the work of development and policy formation. In terms of the impact of educational research on education policy, this must one of the most remarkable examples in recent times.

2.5 Challenges for Educational Change

The story of how AfL has emerged as a focus for policy and practice in the UK, and how it has been variously interpreted and implemented, is a story of relationships between educational researchers and policy-makers. Although some researchers in education are, quite legitimately, content to work in the contributory disciplines of education to produce new knowledge for its own sake, other "educational researchers" seek to use the insights, theories and tools of research to illuminate issues of policy and practice in the hope and expectation that such knowledge will be utilized in the policy context. This is not straightforward because the two different communities of practice – research and policy formation – are characterized by:

- Different time pressures and workflows
- Different priorities and responsibilities
- Different conceptual frameworks and discourse
- Different accountability and incentive systems
- Different cultures and structural positions in society
- Different career structures and pathways

Any attempt to work productively together creates tensions for both groups. The press for quick policy change – to show results before the next election – puts pressure on researchers to deliver ideas and results in a timescale that they find difficult, if not impossible. But sometimes it works the other way and policy-makers are only just getting to grips with certain ideas before researchers have moved on in their thinking, the subtleties of which may be lost on policy-makers who have the task of trying to engineer complex ideas into relatively simple frameworks for action. This may indeed be the case with AfL.

It is notable, from the accounts given above, that a key element of AfL policies in UK countries has been a strategy to introduce teachers to the five or six clusters of AfL practices that were identified in *Inside the black box* and KMOFAP i.e. sharing learning objectives and criteria of quality, feedback, rich questioning, peer

and self-assessment, and formative use of summative tests. The common approach thereafter has been to encourage teachers to test and adapt these ideas in their own circumstances and to evaluate the results with colleagues. However, as researchers have found in subsequent projects, such as in the Learning How to Learn project (James et al., 2007) which investigated the conditions in schools that would enable the embedding and spreading of sustainable AfL practice, and as government agencies have discovered in developing AfL projects for national roll out, a number of problems have arisen which were not well dealt with by the early research, notably:

- How to avoid AfL practices becoming mechanistic, ritualized and ultimately meaningless and boring to pupils.
- How to integrate them meaningfully in the flow of lesson activity, not simply added on.
- How to establish, in teachers' minds, a relationship between the practices and the theoretical ideas that underpin them so that they have the intellectual resources to "know what to do when they don't know what to do".
- How to shift AfL from being a teacher-led activity to it being a learner-led activity.
- How to convince teachers that they have the power and responsibility (agency) to make AfL work in contexts of accountability where they feel constrained by mandatory demands for summative assessments and curriculum coverage that tend to take priority.
- How to manage opportunities for teachers to work together to plan, try out, observe, reflect, discuss and revise their AfL ideas and practices.
- How to encourage school managers to become committed to AfL and to accept responsibility for the professional learning of their staff.

The research relating to these questions was not readily available when policy initiatives in the UK countries were put in place so the policies themselves have become experiments from which lessons have be learned. No doubt they will lead to further adjustments in the future. Inevitably, this constant policy tinkering is irritating to teachers and can make them cynical or passive: a state of affairs inimical to the kind of active engagement that successful AfL demands.

Two issues of particular importance have now become a focus of recent work by researchers in the field. The first concerns the theoretical underpinning of AfL practice and how formative assessment might relate more broadly to learning and pedagogy. Chapters in an edited collection from the ARG published in 2006 (Gardner, 2006) began to examine this relationship. Also, in March 2009, a Third International Conference on Assessment for Learning, involving 30 educational researchers and developers from Australia, Canada, Continental Europe, New Zealand, the USA and the UK wrote a position paper on AfL which acknowledged the difficulties of articulating and communicating this relationship (This can be downloaded from: http://www.annedavies.com/assessment_for_learning_ar_a010.html. Accessed 22 January 2011. It is also reproduced in Klenowski, (2009)). They began by noting the widespread use of the phrases *assessment for learning* and

formative assessment in educational discourse but expressed concern about some of the ways in which the words are interpreted and made manifest in educational policy. They then attempted to clarify the relationship between assessment and learning by emphasising that the primary aim of assessment for learning is to contribute to learning itself. This follows from the logic that when true learning has occurred, it will manifest itself in performance. The converse does not hold: mere performance on a test does not necessarily mean that learning has occurred. Learners can be taught how to score well on tests without much underlying learning.

AfL is therefore the process of identifying aspects of learning as it is developing, using whatever informal and formal processes best help that identification, primarily so that learning itself can be enhanced. This focuses directly on the learner's developing capabilities, while these are in the process of being developed. AfL seeks out, analyses and reflects on information from students themselves, teachers and the learner's peers as it is expressed in dialogue, learner responses to tasks and questions, and observation. AfL is part of everyday teaching, in everyday classrooms. A great deal of it occurs in real time, but some of it is derived through more formal assessment events or episodes. What is distinctive about AfL is not the form of the information or the circumstances in which it is generated, but the positive effect it has for the learner. Properly embedded into teaching-learning contexts, AfL sets learners up for wide, lifelong learning. These ideas were summed up in a short second-generation definition of assessment for learning generated by the Conference:

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

The second issue from research, with implications for policy, concerns the continuing attention that needs to be given to support for teachers' own learning if AfL is to be effective in classrooms, embedded in schools, spread across schools, and sustained over time. This was the main focus of the Learning How to Learn Project and has also been a theme for another project from the ARG, which has used desk research, interviews and expert seminars to analyze insights emerging from twelve projects, many of which are mentioned in this chapter. The Analysis and Review of Innovations in Assessment (ARIA) project began disseminating its findings in a pamphlet, *Changing Assessment Practice: process, principles and standards* (Gardner, Harlen, Hayward, & Stobart, 2008). The chapter by Gardner and colleagues in this volume is another output.

All these developments indicate that research and policy development are not sequential activities (one before the other) but are necessarily pursued alongside each other. For this reason, the channels of communication have to be open, and deliberate efforts have to be made to engage in dialogue, in order to understand the pressures on, and the possibilities for, evidence informed policy making for effective change. The UK's Teaching and Learning Research Programme (see http://www.tlrp.org), with which the author of this chapter has been involved as deputy director, has been centrally involved in "creative mediation" between the 700+ researchers

and 100+ projects in its portfolio and all the potential "users" who might be interested in and benefit from their findings. Andrew Pollard, the director of TLRP from 2002, characterizes its role as a form of "reflexive activism":

> We are trying to build the social capital of educational research – developing relationships and networks, sharing perspectives and building alliances with present and future stakeholders both within and beyond the research community. We are trying to promote collective, open and reflexive debate and action in respect of the changes which need to be faced. We are working on politically engaged impact and dissemination strategies with a view to making a difference. And finally, we are attempting to position ourselves strategically in respect of long-term issues. (Pollard, 2005, p. 4)

This statement could equally characterize the disposition and orientation of the UK's Assessment Reform Group and its belief that researchers, policy-makers and practitioners have to work together in the middle ground of "creative mediation" if research and policy are to contribute to effective educational change.

References

Assessment Reform Group. (2002a). Assessment for Learning: 10 principles. Assessment Reform Group. http://www.assessment-reform-group.org. Accessed 22 January 2011.

Assessment Reform Group. (2002b). Testing, motivation and learning. Assessment Reform Group. http://www.assessment-reform-group.org. Accessed 22 January 2011.

Black, P. J. (1997). Whatever happened to TGAT? In C. Cullingford (Ed.), *Assessment versus evaluation* (pp. 24–50). London: Cassell.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Books.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, policy & practice*, 5(1), 7–75.

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

Council for the Curriculum, Examinations and Assessment (CCEA). (2006). *Assessment for learning report*. Belfast: Council for the Curriculum, Examinations and Assessment.

DARG. (2004). *Learning pathways through statutory assessment: Key stages 2 and 3. Final report*. Cardiff: Welsh Assembly Government.

Daugherty, R. (2007). Mediating academic research: The Assessment Reform Group experience. *Research Papers in Education*, 22(2), 139–153.

Department for Children, Education, Lifelong Learning and Skills (DCELLS). (2008). *Curriculum and assessment update,* Issue 4, Spring. Cardiff: DCELLS, Welsh Assembly Government.

Department for Children, Schools and Families (DCSF). (2008). *The Assessment for Learning strategy*. London: DCSF re: 00341-2008DOM-EN. www.teachernet.gov.uk/publications. Accessed 22 January 2011.

Department for Education and Skills (DfES). (2007a). *Leading improvement using the Primary Framework: Primary National Strategy Guidance for primary headteachers and school leaders*. London: DfES re: 00484-2007BKT-EN. www.teachernet.gov.uk/publications. Accessed 22 January 2011.

Department for Education and Skills (DfES). (2007b). *Assessment for Learning – 8 schools project report: Secondary National Strategy Guidance for School leaders, teachers, school improvement advisers and national policy makers*. London: DfES ref: 00067-2007BKT-EN. www.teachernet.gov.uk/publications. Accessed 22 January 2011.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality and development*. Philadephia: Psychology Press.

Elliott, J. (1990). Educational research in crisis. *British Educational Research Journal*, *16*(1), 3–18.

Gardner, J. (Ed.). (2006). *Assessment and learning*. London: Sage.

Gardner, J., & Cowan, P. (2000). *Testing the test: A study of the reliability and validity of the Northern Ireland transfer procedure test in enabling the selection of pupils for grammar school places*. Belfast: School of Education, Queen's University Belfast.

Gardner, J., & Cowan, P. (2005). The fallibility of high stakes '11-plus' testing in Northern Ireland. *Assessment in Education*, *12*(2), 145–165.

Gardner, J., Harlen, W., Hayward, L., & Stobart, G. (2008). *Changing Assessment Practice: Process, principles and standards*. ISBN 9780853899297. http://www.aria.qub.ac.uk. Accessed 22 January 2011.

Hargreaves, D. (1996). *Teaching as a research-based profession: Possibilities and prospects*. London: Teacher Training Agency Annual Lecture.

Hayward, L., Spencer, E., & Simpson, M. (2006). *Assessment is for learning: Exploring programme success. The Aifl formative assessment project*. Glasgow: University of Glasgow.

Hillage, J., Pearson, R., Anderson, A., & Tamkin, P. (1998). *Excellence in research on schools*. London: Department for Education and Employment.

James, M., McCormick, R., Black, P., Carmichael, P., Drummond, M. -J., & Fox, A., et al. (2007). *Improving learning how to learn: Classrooms, schools and networks*. London: Routledge.

Klenowski, V. (2009). Assessment for learning revisited: An Asia-Pacific perspective. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 263–268.

Miles, A. (2008). Today's SATs result: Government 0 Pupils 0, *Times Online*. 6 August. http://www.timesonline.co.uk/tol/comment/columnists/alice_miles. Accessed 22 January 2011.

Mullis, I., Martin, M., Kennedy, A., & Foy, P. (2007). *PIRLS 2006 international report: IEA's Progress in International Reading Literacy in Primary Schools in 40 countries*. Boston: Boston College. http://timss.bc.edu/pirls2006/intl_rpt.html. Accessed 22 January 2011.

Pollard, A. (2005). *Taking the initiative? TLRP and educational research*. Educational Review Guest Lecture, 12th October 2005, School of Education, University of Birmingham. http://www.tlrp.org. Accessed 22 January 2011.

Scottish Government. (2005). *AifL – Assessment is for Learning information sheet*. http://www.scotland.gov.uk/Publications/2005/09/20105413/54156. Accessed 22 January 2011.

Senge, P., & Scharmer, O. (2001). Community action research: Learning as a community of practitioners, consultants, and researchers. In P. Reason & H. Bradbury (Eds.), *Handbook of action research: Participative inquiry and practice*. Thousand Oaks, CA: Sage.