Draft chapter
Not for quotation

# Protecting the innocent – the need for ethical frameworks within mass educational innovation

Having spent 20 years immersed in the policy and development work associated with large scale innovation and transformation of the UK education and training system, I have in recent years become increasingly concerned about the three things: the rapidity of change and turnover in major innovations; the lack of transparency in the origins of innovations (mitigated only to a degree by the Freedom of Information Act); and the lack of safeguards for those learners caught up in major system innovation. Rather than these being three separate things, my involvement in policy development on mass innovations – as head of research first at the National Council for Vocational Qualifications and then the Qualifications and Curriculum Authority – forced a stark recognition that they are intimately linked. This chapter examines these linkages but places a spotlight on the ethical issues relating to safeguards for learners.

From the increasingly vitriolic exchange between 'resisters' (Oakley A, 2006) and 'militant empiricists' (Lather P, 2004) I wish to extract two concepts. From Oakley I will extract the notion of 'harm' (Oakley A, 2000). From Hammersley I will extract the notion of 'experimentation' (Hammersley M, 1997). In extracting these I am not ignoring the highly specific commitments of (i) the discourse in which they are embedded nor (ii) the specific theoretical assumptions which are associated with them within this discourse. But I DO deliberately wish to depart on a tangent – a line of enquiry which will hopefully be of interest to both sets of warring parties. Whilst some may feel I am simply engaging in a minor skirmish on the periphery of the conflict, my contention is that the new evidence I present opens up a whole new front.

Firstly, Hammersley: in education (and training) '….strict experimentation is often ruled out for practical or ethical reasons.' (Hammersley M, 1997 p145). The possibility of 'experimentation' has been long contested in educational research (McCall WA, 1923; Miles M, 1964; Davies H et al, 2000). Hammersley's use of the construct in his critique legitimates its application to enquiry in social and natural science by virtue of his claim that its use is limited through 'practical and ethical' limitations, rather than on absolute (ontological) grounds. As a result, I will extract the concept of 'experimentation' and examine mass innovations in education and training (such as the introduction, in England, of the National Curriculum) against criteria which are, or can be, utilised in the regulation of experimentation. Put simply, if we can describe some actions in education as 'experiments', then we can examine their characteristics and the outcomes of different approaches.

Secondly, Oakley:

> Paradigm wars will not bring about social justice or justify public expenditure:
>
> The goal of an emancipatory (social) science calls for us to abandon sterile word-games and concentrate on the business in hand, which is how to develop the most reliable and democratic ways of knowing, both in order to bridge the gap between ourselves and others, and to ensure that those who intervene in other people's lives do so with the most benefit and the least harm.
>
> (Oakley, A. 2000 p.3)

From this I extract the key notion of 'harm'. This evokes a linkage with medical ethics. I make this central to the analysis in this chapter. Building on the notions of 'experimentation' and 'harm', I will map a line of argument which covers:

- the legitimacy of a partial and precise analogy between education and medicine in respect of the positioning of ethical regulation of practice, asserting that 'do no harm' should be central to educational evaluation and policy-making

- the role of realist ontology and epistemology in explaining problematic elements of the relationships between educational researchers and policy-makers, and in reforming these relationships through a more realistic and functional set of mutual expectations

- the severe ethical defects of major evaluations of fundamental reforms within the English education and training system. Four case studies will be used to exemplify different forms of these defects

I shall argue that there has been chronic and serious neglect of key ethical dimensions in major reforms over the past 20 years. I conclude by presenting an argument for a radical look at new arrangements for ethical regulation of both educational evaluation and of innovation in education and training arrangements.

## Is 'do no harm' relevant to mass innovation in education and training systems?

The principal focus of this chapter is: how should the principle of 'do no harm' be interpreted at educational system level? But is it vital to first ask – is 'do no harm' relevant to education and training? Here, I wish to argue that an analogy can be drawn between the decisions made by teachers and educationalists on the one hand, and medical professionals on the other.

I will be precise about my focus:

I am not dealing with the ethics of educational research and evaluation

I am not dealing with the ethics of educational practice (eg ethical codes for teachers)

I am dealing with the ethics of mass innovation – policy development and the implementation of reform

Whatever the precise origin and etymological history of 'do no harm' within medical ethics (Friedson E 1970), it has emerged as a primary driver of medical ethics. Yet, what can explain one startling difference between education and medicine – the prominence of ethical regulation of medical practice and implementation of new therapies, and the lack of prominence of ethical regulation for educational innovation? The answer cannot lie in the structure of competence in the two domains, since researchers increasingly use cross-analysis of educational practice and medical practice in order to push forward understanding of the structure of competence (Boreham N 2002; Fischer M 2004).

| | Examples from the practice of medicine | Examples from the practice of education |
|---|---|---|
| Practice as 'treatment' | Decisions regarding therapeutic regimes, drugs administration etc; | Decisions regarding learning environment, adapting learning approaches and activities in order to optimise learning |
| Formalised, conscious procedures | Diagnostic protocols; initiating appropriate tests; operating consent and confidentiality protocols; optimising outcomes for individuals; working to targets | Deployment of teaching and learning approaches such as objective-based learning, the Literacy and Numeracy strategies; cognitive acceleration; within-class pupil grouping; working to targets |
| Underlying unconscious procedures, implicit professional rules | Adapting communication strategies to the needs of different patients; optimising personal performance in different team settings | Adapting teaching and learning strategies to the needs of different learners in different contexts; optimising personal performance in different team settings |
| Examples of confounding variables in securing effective outcomes | Family wealth, social background, self-medication, prior conditions | Gender, date of birth, family wealth, social background, early educational experience |

If systematic application of ethical controls are a principal driver of practice, policy formation and innovation in one domain, why should it not figure in the same structural way in another domain which has many fundamental features in common?

However, I wish to draw an analogy between education and medicine in respect of on key dimension only. I am NOT arguing that the structure of knowledge in education is the same as in the medical arena (Sokal A and Bricmont J 1998) nor am I arguing that methods in educational research can or should emulate those in medical research (Cook T 2003). However, I wish to analogise between education and medicine in respect of *intervention*. This follows the use of the term 'intervention' to describe educational practice in educational research work which falls into an experimental paradigm (eg Fitzgibbon C & Defty N, undated). However, it is clear that the concept of 'intervention' extends beyond this, into areas of programme-based educational practice, such as the English literacy, numeracy and standards programmes (see DfES 2006).

Note that I am not implying that the term 'intervention' applies only to policy of a certain form or that all interventions follow a top-down approach – interventions can follow highly collaborative and negotiated forms (Stenhouse L 1985; Eliot J 1999). However, I am arguing that policy oriented towards change (in systems, in pedagogy, in pupil outcomes) can be characterised as 'intervention'; that is, an attempt to affect the trajectory of development in systems and individuals.

Is there a seamless join between a national policy directed at change and the guidance of the teacher, supporting a learner in the day-to-day exchange of a learning setting, or are they different in type? I would argue that – in one key respect – they are the same. Even if viewed from a hardline constructivist position, the exchange between teacher and learner can be theorised as a common attempt to facilitate learning – that is, to extend learners' conceptual frameworks and schemata; to change the condition of the learner (most simply, from unknowing to knowing; unskilled to skilled; non-competent to competent) (Whitebread D 2003). Learning necessitates change (Geber B 1977; Mayer R 1977) – and change can be construed, with validity, as involving *intervention* in things as they are, in order to make them something else. In supporting learning, a teacher or instructor faces decisions about what they say and do to support the learner and when; they exercise judgement (Walter GA 1981). This can be construed as 'intervention' – inputs/actions which cause learning to occur more effectively than if that support was absent. It can also be construed as 'treatment' – not only are there decisions about whether or not to support a learner directly, there are decisions about *appropriate* actions – there are choices in respect of what to do. Ignoring the connotations of passivity and one-way power relationship in the idea of a 'therapeutic regime', teachers and instructors *are* concerned with setting up 'learning mileu' – where all elements of the learning context are configured to produce beneficial learning (Stenhouse L 1983). I am suggesting that this is analogous to 'medical treatment'. In this I am not arguing that a state of 'unknowing'– eg not knowing a set of mathematical operations, or not knowing the distinction between 'sex' and 'gender' – is an illness or pathology and that 'knowing' is a state of health. What I am arguing is that managing learning settings such that a learner can understand these things, can acquire knowledge and skill etc, involves complex choices (Walter GA 1981) and these can be characterised as a choice of 'treatment'.

> Treatment:
>
> care by procedures or applications that are intended to relieve illness or injury
>
> the management of someone or something; "the handling of prisoners"; "the treatment of water sewage"; "the right to equal treatment in the criminal justice system"
>
> http://wordnet.princeton.edu/perl/webwn?s=treatment
>
> A treatment is a specific combination of factor levels whose effect is to be compared with other treatments.
>
> http://www.itl.nist.gov/div898/handbook/pri/section7/pri7.htm

Given the power relationships which DO obtain in early years education and compulsory schooling (Biken SK and Pollard D 1993) – and the power relations implicit in the requirements of a national curriculum (Aldrich R and White J 1998; Tapper T and Salter B 1978; Tapper T and Salter B 1981) and its assessment – there is a moral imperative associated with decision-making relating to choice of 'educational treatment' (Peters RS 1970; Bottley M 2000). This invokes the analogy with medical treatment – an analogy strengthened by the conceptual commitments of the increasing number of teachers' codes of ethical practice which are emerging around the world. From New Zealand:

1. Commitment to learners

The primary professional obligation of registered teachers is to those they teach. Teachers nurture the capacities of all learners to think and act with developing independence, and strive to encourage an informed appreciation of the fundamental values of a democratic society.

Teachers will strive to:

a) develop and maintain professional relationships with learners based upon the best interests of those learners,
b) base their professional practice on continuous professional learning, the best knowledge available about curriculum content and pedagogy, together with their knowledge about those they teach,
c) present subject matter from an informed and balanced viewpoint,
d) encourage learners to think critically about significant social issues,
e) cater for the varied learning needs of diverse learners,
f) promote the physical, emotional, social, intellectual and spiritual wellbeing of learners,
g) protect the confidentiality of information about learners obtained in the course of professional service, consistent with legal requirements.

http://www.teacherscouncil.govt.nz/ethics/code.stm

In England, the General teaching Council has produced an analogus statement:

GTC statement of professional values and practice for teachers

Teachers place the learning and well-being of young people at the centre of their professional practice.

Teachers respond sensitively to the differences in the home backgrounds and circumstances of young people, recognising the key role that parents and carers play in children's education.

Teachers see themselves as part of a team, in which fellow teachers, other professional colleagues and governors are partners in securing the learning and well-being of young people.

Teachers entering the teaching profession in England have met a common professional standard.

http://www.gtce.org.uk/shared/contentlibs/92511/92572statement_of_values.pdf

Such codes assert the ethical basis of practice – ie that ethics are not a contingent consideration but are intrinsic to practice. This mirrors the work on the centrality of ethics within competent professional practice (Boreham N 2002; Eraut M 1994;Oates T 2004). A key to my analysis here is the commitment to b) above: '… base their professional practice on continuous professional learning, the best knowledge available about curriculum content and pedagogy, together with their knowledge about those they teach…'.

These ethical considerations relate to transformation of practice – with a focus on the relationship between individual teachers and groups of learners. But in terms of magnitude of responsibility and power of effect, if ethical considerations apply at this level they surely must, by implication operate with even greater force at the level of system review and revision. 'Do no harm' is thus justified as a constraint on policy and mass innovation. But before examining a number of initiatives against this principle, an excursion into the philosophy of science is necessary.

## Understanding the status of knowledge which is generated through educational research and evaluation – putting in place more realistic expectations of 'useful knowledge'

The extended debate regarding the putative reductionism in commitments to implement education policy based on 'what works' (Chalmers I 2005) has skirted only loosely around the issue of the nature of knowledge in social science. Yet this is fundamental. If policy-makers and politicians are encouraged to rely on knowledge generated by educational research (Davies H et al 2000; Sebba J 2004) they have a right to know the status of that knowledge – and I contend that this issue has been grossly neglected, leading to frustration on the parts of both researchers and of those in the policy community.

The US National Research Council's influential *Scientific Research in Education* (ref) contains a crucial error. The authors assert strongly the credentials of educational research, and that accumulation is both desirable and possible (Shavelson R and Towne L 2002). Unfortunately they argue for the status of social enquiry on the basis that there is no distinction between the knowledge-creation processes of natural and social science (p28ff). Ironically, this infringes the very principle of accumulation that they avow and to which they aspire. They have overlooked the key contribution of realist theory to social science – namely that there is an absolute distinction between natural and social science (Bhaskar R 1979), which will be a principle underpinning for this chapter. In reaching their (flawed) conclusion regarding the commonality of natural and social science, they failed to include realist theory in their 'accumulations'.

Bhaskar's realist ontology (Bhaskar R 1975; Bhaskar R 1979) is essential to adequate understanding of the status of knowledge generated in social science in general, and educational research in particular. I will contend that it helps to re-cast the relationship between the research and policy communities. Ironically, it better cements relationships by showing the necessarily conditional nature of the knowledge emerging from enquiry into social/educational systems – ie that this knowledge is more provisional and less certain than the knowledge yielded by natural science. Realist perspectives thus hold the potential for placing this relationship on a more *realistic* footing – rather than one side (policy-makers) having over-ambitious expectations of the knowledge created by the other side (the researchers). In this chapter I wish to redefine what we mean by 'useful knowledge' (Davies H op cit) and try to lay the foundation for a redefined relationship between policy makers and researchers. But I also wish to explore the extent to which a failure to characterise accurately the nature of educational research and of policy activity is accompanied by a serious omission of procedure: the failure to erect a set of ethical principles regulating innovation and experimentation, and the allied need for careful and impartial evaluation. I will offer four case studies which illuminate different aspects of this failure, and argue that innocents have been placed at risk.

Bhaskar's original texts are dense (Bhaskar R 1975; Bhaskar R 1979) – but the implications can be sketched simply: in physics, no matter what you feel about gravity, things will carry on falling at the same rate. You can run different experiments on gravitational forces, get angry that some don't work as you wanted them to, feel a warm glow as the results come up as expected. A given theory may prove right or wrong - in other words, our view of the world may be correct or defective - but meanwhile gravity will simply carry on doing what it always does; keeping my coffee cup neatly on the table and terrifying me when I go rock climbing.

But social science is *fundamentally* different. It's a science which has a different relationship with the world. The theories we have about the social world actually affect how that world operates. If I have a theory that certain social groups are better at certain activities, or that females are better at certain activities than men, then that affects how my relations with those groups actually operate. In other words, the social theories I have affect the social world in a way in which theories about the natural world do not affect the processes which operate in that world. Social theories are a part of the social world - they affect the way the social world operates.

Roy Bhaskar gives us an excellent example of this important perspective on social theory: the one pound coin. It's a round piece of metal which costs a great deal less than one pound. But it's worth one pound. Why? Because a group of people share a common belief that it's worth a pound. It is of course really useful that these shared beliefs operate in the social world. It enables the whole banking system, indeed the whole economy, to work. It shows us that beliefs play an important role in the operation of important social systems. And it tells us that these systems do not operate independently of human action, as suggested by phrases such as '..natural operation of the market…'.

But while social theory and social research can be very good at explaining things – why certain social groups behave in certain ways, what's going on in a specific community – Bhaskar's analysis asserts that social science will of necessity lack the capacity to predict events. High on explanation and low on prediction again shows a

critical difference between social and natural science. Natural science has high predictive power – such as predicting the temperature at which water will boil when taken to 6000 metres, or the gauge of copper electrical cable which is needed to safely run a large piece of industrial plant of a certain power consumption.

The fact that social theory is part of the system which it seeks to explain means that social science falls over repeatedly when it comes to prediction. When you come to the implementation of social policy it just gets worse. This is exemplified in the model of 'planned failure'. At the Third International Conference of Learning at Work in Milan, June 1994, Frank Achtenhagen outlined a model of 'planned failure' in social policy. This linked the ontology of realist theory to social policy. He stated: if you fail to adequately understand the nature of the problem you are tackling, you formulate policy which half-engages with the problem – but at the same time, putting the policy in place changes the nature of the system you are dealing with, giving you a whole new set of problems which you no longer understand at all. This is the principal causal mechanism behind the increasingly-mentioned 'unintended consequences' of policy.

It is also the mechanism through which Goodhart's Law (Goodhart C 1984) and the 'Texas Test Effect' operate (Wiliam D 2001). In line with Bhaskar, there is reflexivity between measures and the system – the system adapts to the indicators. The weak correlation between performance indicators and performance itself is not evidence of the distance between them, but of the way in which they are entwined.

Inflated expectations of the predictive power of knowledge from educational research and evaluation poses problems for a well-grounded relationship between the research and policy communities. This is a key problem: the provision of knowledge which has adequate power – knowledge in which decision-makers at national level can have confidence. In other words, knowledge which makes clear what is going on, why it is happening, and whether revised arrangements offer an improvement over existing arrangements (particularly in terms of better use of public money). Despite their limitations, this issue was confronted in both the Tooley Report (ref) and the Hillage Report (Hillage J, Pearson R, Anderson A. and Tamkin P 1998). Both highlighted acute and chronic problems in educational research and the research-policy relation in respect of: a failure to generate/fund research knowledge which engaged sufficiently with practice; paradigm conflicts within research which militated against knowledge accumulation; and contradictory findings from broadly parallel studies – often derived from defects in method (Goldstein H, 1991; Goldstein H and Blatchford P 1998). These problems formed a starting agenda for the ill-fated National Educational Research Forum (NERF) (NERF undated). These were also confronted in the US National Research Council's (USNRC) influential reference text on educational research (op cit).

But two distinct arguments were conflated in these discussions: (i) the contingent argument that researchers tended to focus on esoteric areas unrelated to practice; and (ii) the more fundamental issue of confidence which could be placed in the findings and thus their utility for policy formation. The former has, notably, been addressed in the revised ESRC arrangements for contracting and directing research within the TLRP programme (TLRP undated). I will not deal with this area. Rather I am focussing on the second issue, and examining the relationship between the structural form of knowledge in enquiry in education, the nature of the confidence which we can have in such knowledge, and the effect that inappropriate understandings of this have had on the researcher-policymaker relation.

Far from pouring oil on troubled waters, the implied centralism in the NERF arrangements fuelled oppositions in the research community (Ball SJ 2001; Hodkinson P 2001). In 2004, Phil Hodkinson re-entered the still-raging debate on paradigm wars in educational research and the research-policy relation by asserting that the necessary multiplicities of theory and discourse in educational research suggest absolute relativism in knowledge of complex social systems (Hodkinson P 2004). In other words, Hodkinson's ontological and epistemic commitments imply that knowledge accumulation – indeed meaningful communication – is not possible in educational research, nor can knowledge accumulation support more effective policy-making in education. The plea for 'useful knowledge' which emerged in the wake of Hillage looked as though it was being ruled out of court on principle. But the burgeoning internal contradictions of Hodkinson's argument causes it to collapse under its own weight. Not least, just as Samuel Beckett's plays heralding the death of meaningful human communication relied on absolute precision in staging in order to communicate this message (Worth K 2001), Hodkinson relied on assertion of explanation through conventional discourse in order to dispute the explanatory power of conventional discourse. This contradiction did not escape either this author or Martyn Hammersley (Hammersley M 2005).

Bhaskar's realist position suggests that whilst research enquiry in social systems cannot possess high levels of predictive power and thus is to be differentiated from natural science, it does possess high levels of causal and explanatory power – and qualifies as science. This position on the status of the object of enquiry also suggests that some theories and enquiries will yield higher quality knowledge than others (using causal power and

explanatory power as criteria). Contrary to Hodkinson's relativist position, knowledge located within different paradigms can be related one with another, since they relate to the same (real) object – social systems and their operation. All theories are not equal, and discussion, dispute and accumulation is thus not ruled out. Hodkinson's implied 'get your tanks off my lawn' position in respect of policy-makers, is thus replaced with a shared, common concern – deriving the most powerful understanding of the mechanisms at play in education and training systems.

But is worthwhile spending a little more time on the implications, for policy and research, of realist theory. Bhaskar argues that the power of explanation in social science are at their highest under conditions of *ceteris paribus* (Bhaskar R 1979). But this argument in respect of *ceteris paribus* – all other things being equal - is of critical importance, not least in that in respect of complex systems such as educational systems they usually aren't. That is, the educational attainments of individuals are determined by a range of factors – some of which are outside the control of the formal educational system, such as family wealth (or lack of it), culture, economic drivers, etc. Improvement in educational outcomes may derive from changes in these factors rather than from policy impacting on the form of education and training. Often neglected in explanations of educational performance, the linking of social, economic and educational policy is seldom effected in system reform and development (recent notable exceptions being SureStart (Sylva et al 2004) and Excellence in Cities (DfES 1999). Under this form of fragmented policy management, ceteris paribus is startling in its absence – that is, the additional variables which will determine the success of a focussed policy are not attended to – all things are not equal, and there is no attempt to affect the additional variables which make this the case. In addition, complex social systems such as education systems evolve unpredictably as individual human actions shift the underlying relations which make up those systems – actions motivated by subtle shifts in individuals' operational theories of the world ('…I'll try this slightly different approach to managing my class…' which shifts the nature of the social dynamic in the class, the behaviour of pupils and thus the teacher's perceptions of the pupils, which demands another shift in class management approaches). As stated above, unlike in natural systems, theory in the social domain *is part of that domain*.

All this impacts adversely on the ability of educational researchers to generate 'certain' knowledge – ie knowledge and theory with high explanatory and causal power. As outlined above, this contingent lack of 'all things being equal', combines with an absolute limit on definitive predictive power. Bad news for the researcher who is asked an apparently clear and simple question by a minister – '…but will it definitely improve standards…'? Researchers with Hodkinson's ontological and epistemic commitments would be bound by principal to say '…this is not a reasonable (or even recognisable) question…'. For funder's of research and evaluation, for policy-makers and developers, indeed, for teachers interested in innovation, this is not a response which is welcomed. By contrast, the Realist educational researcher would say '…all other thing being equal, this innovation is likely to produce a particular set of tendencies in the system (for these particular learners, and possibly with these unintended consequences)…'.

This hints at the nature of a key difference between policy-makers' and (Realist) researchers' views of knowledge of educational systems. Policy makers require knowledge which gives clear direction and clear analysis of outcomes (explaining what has happened and what has caused it to be the case). A Realist researcher would assert that the necessary limits on prediction in social (educational) systems lead us to conclude that it is possible to:

- generate explanations for events and outcomes
- generate causal explanations (x caused y)
- generate knowledge of likelihoods and tendencies

and, despite falling necessarily short of the forms of knowledge that can be created in natural science, this is an adequate base for the formulation and evaluation of innovation in education and training systems. Under these conditions, knowledge from different communities of researchers can be accumulated (even though different paradigms may be in play, they are both being deployed on the same, real object of study – ie features of complex social systems), and can both comment on the effectiveness of innovation and on promising ways forward. But researchers and policy-makers should recognise that research should not try to meet too high a burden of proof (Minister: '…will this innovation *definitely* improve standards?…'). Rather, we should characterise the application/utilisation in policy of the outcomes of educational research as 'the integration of adequate knowledge of tendencies'. Minister: '…will this innovation definitely improve standards?...'. Researcher '…on the basis of these five qualitative and quantitative studies it is the most likely initiative to raise standards, given that factors x y and z in the surrounding context do not change radically…'. This is not an ineluctable descent into relativism, nor is it an assertion of infallible knowledge. It is a realistic statement of the kind of 'useful knowledge' that educational research can contribute to policy formation and policy evaluation (Pawson R and Tilley N 1997). Too high a

burden of proof is not justified – it leads to the paradox of 'the descent into the irrational due to the rejection of 'acceptable uncertainty' and the consequent formation of 'half-right assumptions' (Hodgson et all 2006). That is: policy maker: '…if you can't give me a definitive and absolute answer to whether this innovation will definitely improve standards then I'll just go with my gut feeling and the fact that my son/daughter/nephew/niece/neighbour enjoyed this new qualification…'. Avoiding this paradox is vital: research and evaluation must generate useful knowledge; policy makers must have realistic expectations of the status of this knowledge (its true limits as well as its true power).

The case studies which now follow show how the relation between research/evaluation and policy/innovation has been both defective and dysfunctional. I am suggesting that the ethical dimension is vital both for understanding the nature of these defects and for considering the way in which an ethical framework for regulating innovation might contribute to (i) avoiding such serious problems in the future, and (ii) better protecting learners.

Case study #1

## Ethics and 'ownership of innovation' - the handling of adverse evaluation findings from Curriculum 2000 (C2K)

The problem of 'ownership' of innovation was the subject of extended discussion following the author's presentation at the dissemination conference for the ESRC project ESRC-funded research project - *Education and Youth Transitions in England, Wales and Scotland 1984-2000.* From the audience of senior researchers and educationalists there emerged a strong sense of unease with current evaluation arrangements relating to mass innovation. Critically, the timing of evaluations of major innovations and programmes are most usually out of synch with the political cycles created by (i) general elections, (ii) the duration of ministerial office. The timetable for review, development and effective evaluation of a new or revised qualification or learning programme will extend well beyond a typical 4-5 electoral period, and certainly well beyond the period of office of a set of education ministers (Lea R 2002). The revision of A levels and GNVQs (the latter into Applied A level) following the Dearing Review of 14-19 Qualifications is typical:

| | |
|---|---|
| 1994 | Commissioning of the Review of 16-19 Qualifications |
| 1996 | Report of the Review of 16-19 Qualifications |
| 1997 | Consultation on 14-19 reforms |
| 1999 | new form of qualifications developed for introduction from Sept 2000 |
| 2000 | revised qualifications implemented |
| 2001 | report of QCA Research Team interim first year evaluation of revised two-year qualification |
| 2002 | first full report of QCA Research Team review of operation of two-year qualification |
| 2003 | second full report of QCA Research Team review of operation of two-year qualification |

Restricted reporting by QCA to Government in 2001 was the first indicative feedback on the effects of the changes (based on a 100% sample of post-16 centres). The first report of evaluation of the full operation of the two-year qualifications was possible only in late 2002. Learning programmes and assessment arrangements only begin to settle to some degree after one two-year cycle of operation (Kingdon M and Stobart G 1988). Teachers may adapt programmes and refine administrative arrangements before this, in response to emerging problems – implementing these in the intake year for the next cycle (ie the start of year two of implementation), but the outcomes of a full implementation of a two-year qualification are obvious only after the completion of the full two years of operation. In the light of changes which practitioners are likely to implement, it is also probable that further significant shifts in practice or evidence of stability in practice will only emerge after three or four years of operation of a two-year qualification. In the case of GNVQ and A level revisions under Curriculum 2000, the initial proposal to review preceded the emergence of comprehensive evaluation findings by 8 years. John Patten (Con 06.07.92-20.07.94), Gillian Shephard (Con 20.07.94-01.05.97), David Blunkett (Lab 02.05.97-07.06.01), Estelle Morris (Lab 08.06.01-23.10.02); Charles Clarke (Lab 24.10.02-16.12.04).

It is clear from this that a minister in receipt of problematic evaluation findings from an innovation may have had no responsibility for the implementation of the innovation, let alone for the commissioning of the review which led to the proposal for, and design of, the innovation. Eight years can easily span three governments, with similar issues of continuity and discontinuity in strategy and 'ownership'.

This is not a claim that there is no continuity in policy direction over considerable periods of time. Far from it, some elements of policy remain highly enduring – for example, (i) the commitment to a National Curriculum, (ii) the commitment to accountability arrangements based on assessment of every child at the end of each key stage (nominally 7, 11 and 14). These have remained commitments of successive Conservative and Labour

administrations over two decades. Rather, this is a claim that the use of evaluation data for the purpose of policy revision is out of synch with the timings of political cycles.

A ministerial team can thus be faced suddenly with a report of adverse findings for a major element of the education and training system which was neither conceptualised nor initially implemented under their 'watch'. I will not here analyse the full range of problems which this can give rise to (Hodgson A and Spours K 2004) – which include obvious effects such as premature rejection of innovations on the basis of 'not invented here' and inappropriate modification of programmes which distort original aims and objectives; but also include counter-intuitive results such as re-invention of defective programme structures due to poor 'policy memory' (Higham J and Yeomans D 2005). Instead, I will examine the events around the reaction in December 2002 to emerging adverse evaluation findings on C2K, approximately eighteen months after the initial implementation of the C2K programme. Note that this was some six months before the so-called 'A level crisis' or 'A level meltdown' of summer 2002, which featured unprecedented press scrutiny of both A levels and the C2K changes, led to the rapid resignation of the Chair of QCA Sir William Stubbs (on 27 09 2002), and was a contributing factor in the subsequent resignation of the Secretary of State for Education and Skills, Estelle Morris (on 23 10 2002).

Again, it is important to note my purpose of examining this case study: that is, to establish the extent to which the ethical basis of reforms was attended to during key periods of decision-making and handling. In the case of C2K and its evaluation, a serious breakdown occurred in the winter of 2001. Government had requested a report from QCA on the progress of C2K; concern had been growing as a result of the problems during awarding of the first AS qualifications in the summer of 2001 – particularly in Maths (Advanced Subsidiary – the qualification typically taken at the end of the first year of advanced study). The QCA report submitted to DfES in December 2001 took the line that problems were superficial – essentially 'bedding in' issues; there was confidence that with persistence and support, the innovation would achieve its objectives, and quickly (QCA 2001). But behind the scenes, QCA researchers were deeply troubled that the report had failed to include a series of observations from a comprehensive and on-going survey of all post-16 centres. This survey started prior to the implementation of C2K and allowed sensitive analysis of curriculum and assessment issues (UCAS/QCA 2003) emerging in the first three years of C2K implementation – elements of the survey have continued to 2007. It involved both centre and student analysis, and included systematic combination of data from a variety of agencies, using a data-sharing protocol (Oates T 2001). The observations from the survey contained indications of serious and structural weaknesses and problems. In essence, the reports from the evaluation suggested that the fundamental aims of C2K were not being met. This was fully supported by evidence. After submission of the Christmas summary report, concerned researchers approached the Chairman directly, in March. After this closed briefing, a summary report was taken by the Chairman direct to senior DfES officials and the findings aired at the Education Select Committee.

In essence, the report suggested that of the three main aims of C2K, 'breadth', 'depth' and 'progression', both 'depth' and 'progression' were heavily compromised. Indeed, following years showed that they remained so. In particular, the bulk of HE continued to emphasise 'three good A level grades' as the basis for entry. Very seriously, enrichment activities remained a casualty of C2K implementation, and a growing number of voices could be heard to say 'assessment treadmill', 'assessment overload' etc (BBC 2001; LSDA 2001; ATL 2002; Tomlinson M 2004; Hodgson A, Spours K and Waring M 2005)

### Breadth and progression

Positive outcome: The UCAS/QCA national survey showed 58% of year 12 students taking 4 AS qualifications, with 2.8% taking 5 or more. Using matched candidate data from exam entries, 105,067 (53.4%) out of a total of 196,570 Year 12 students in 2001 took 4 AS subjects.

This can be compared with around 25% taking combinations of 4 or more A level and/or old AS qualifications prior to C2K. Therefore, as a result of Curriculum 2000 there been a substantial increase in students taking 4 subjects in their first year of study.

Underlying issue: however, when researchers looked in detail at the subjects which students were studying, they tended to take for their fourth AS a subject related to, rather than contrasting with, their three 'main' AS qualifications. Ofsted characterised this as 'bulk rather than breadth'. There were exceptional centres which have implemented a curriculum philosophy based on breadth but they were indeed exceptions. The numbers taking vocational qualifications at Advanced level (AVCE) remained static, at around 12% of the cohort taking advanced level qualifications. On the basis of the national survey of centres, the numbers of students mixing

academic and vocational qualifications (eg A level students doing the 3 and 6 unit ACVEs) remained more or less static (at around 22% of those doing advanced qualifications).

Significant problem: There is a near universal view from schools and colleges contacted in the national evaluations that HE institutions in general have not changed admissions policy and practice in the wake of Curriculum 2000. The messages from the majority of Higher Education institutions to students, schools and colleges continued to emphasise 'three good A2 grades' as being the principal target for students.

### Breadth – attainment of Key Skills

Significant problem: Key Skills remained very unpopular amongst students; the percentage of students which centres expected to enter for certification in three key skills dropped from 56.2% in Autumn 2000 (the number entered for the Key Skills qualification) to 20.2% in Autumn 2001 (the number entered for 3 Key Skills).

This headline figure for entry contrasted with much lower completion figures. Amongst case study centres committed to key skills and offering coherent provision, they typically had found it hard to keep students attendance to reasonable levels, and amongst students who have attended provision only low percentages have completed the tests; fewer still completed their portfolios.

Admission tutors in Higher Education sent messages to schools and colleges that key skills were far less significant than high grades in three main subjects - when pressure increased in their programmes, students understandably placed less emphasis on key skills.

### Progression from level 2 and within the vocational segment of the system

Underlying issue: Prior to C2K, progression from level 2 GNVQ post-16 provision to level 3 GNVQ was emerging as an established progression route for young people, often those with prior attainments other than 5 or more GCSEs at Grade C or above. Following the introduction of AVCE, schools and colleges emphasised to the national evaluation teams that this progression route had been adversely affected by: change of assessment regime and backwash into learning styles; increasing entry requirements for VCE in the wake of the changes to the qualifications; and centres expressing concern that vocational GCSE would be perceived by post-16 students who have done less well in GCSEs at 16 more as a 'GCSE resit' than the motivating vocational alternative provided by Intermediate GNVQ.

It was clear that looking only at the impact of Curriculum 2000 on those taking advanced qualifications obscured the importance of the progression routes needed for those working at level 2 post-16.

### Breadth - curriculum enrichment activities

Underlying issue: through the UCAS/QCA survey, schools and colleges reported a decrease in enrichment activities in the first two years of Curriculum 2000. This category included a wide range of activities, including those which were extra-curricular in a formal sense and those which broadened teaching and learning through such means as field-trips and visits, visiting speakers, debates and many others.

Whilst it was clear that resource utilisation (eg teaching room usage) had improved as a result of C2K, survey work indicated that enrichment activities had decreased in over 30% of schools and colleges for two years in a row.

| Response | Percentage | |
|---|---|---|
| | Nov 2000 Y12 | Nov 2001 Y12 |
| Increased | 7.2% | 10.3% |
| Decreased | 38.4% | 32.6% |
| Remained the same | 50.2% | 57.1% |

### 'Bedding down' versus fundamental change

Underlying issue: there are indications throughout the UCAS/QCA survey data that centres are learning from the first year's experiences and adopting more robust approaches to timetabling, teaching etc. However, with HE strongly reinforcing 'three good grades at advanced level', movement away from Key Skills; and static numbers in vocational options, there is little to suggest that there will be any significant movement from the status quo. There are no strong drivers or incentives to stimulate substantial change.

(Evidence base for the analysis: 20 case study centres regularly visited by the QCA Research Team; 100 schools regularly visited by Ofsted; 40 case study centres regularly visited by team from Institute of Education; Bi-annual questionnaire sent to all post-16 centres by UCAS & QCA; Lasda evaluation reports, information collected through focus groups & centre visits; AoC evaluation reports, information collected through AoC members; Data from the FORVUS matched candidate data set; DfES Analytical Services; QCA Stats & Information Team)

By December 2001 it was clear from triangulated evidence that major structural problems were apparent in the implementation of C2K and that the key policy objectives were unlikely to be met. Interestingly, the feed-in from the QCA research team included warnings regarding grade aggregation for AS-A2 (A2 being the qualification typically taken at the end of the two years of advanced level study, the outcomes in the A2 being combined with the AS outcomes to yield an overall A level grade). This was prescient, and the potential problems highlighted in March 2002 became a stark reality in the summer awarding session.

But what of the relation between these policy events and ethical frameworks. Although the research group took corrective action which sidestepped the blockages in 'normal channels', I contend that the failure to incorporate the evaluation findings in the initial report to Government constituted an ethical breakdown. 'Do no harm' – or do not introduce innovation which is retrograde in outcomes - was contravened in a number of respects: there was a failure to report an innovation which was not and would not be likely to meet its objectives; failure to quantify the true relationship between benefit and effort/cost; a lack of analysis of the opportunity costs and educational impact of the severe erosion of enrichment activities and of the significant problems with key skills; and a failure to put in place strategies to deal with assessment and awarding problems. By ignoring the research, 'bedding in' became the dominant message (QCA 2001), which understated dramatically the risks to the (educational) welfare of the learners exposed to the innovation. Crucially, the failure to anticipate and ameliorate problems through piloting was all too evident, and would breach an educational analogue of medical ethics.

I contend that using an ethical framework to guide the evaluation of the initiative would have increased the likelihood of avoiding this breakdown, by placing the welfare of learners exposed to innovation in educational 'treatment' at the heart of the policy process.

The seriousness of the exposure of young people to untrialled 'treatment' was subsequently highlighted by the education select committee:

Government admits failing to pilot new A-level
Polly Curtis
Wednesday July 23, 2003
Guardian Unlimited

The government has owned up for the first time to failing to sufficiently pilot Curriculum 2000, a mistake which has been widely blamed for last year's A-level crisis.

In written evidence submitted to the education select committee, published today, the Department for Education and Skills says: "The AS was piloted before its introduction as QCA recognised that it was a new qualification and that a new standard would need to be established. The possibility of piloting the A2 exam was not contemplated at the time. We accept that this would have been desirable."

The admission comes after the select committee took the unusual move earlier this month of rejecting the government's first submission as "unsatisfactory" after they failed to respond to the charge that failure to properly pilot Curriculum 2000 led to last year's A-level crisis.

Barry Sheerman, chairman of the committee, said the new evidence was "eight of ten" for effort.

"The original was four out of ten – we said it was so poor that we sent it back and told the secretary of state he had to stay in after school and write a new one.

> "They have come back and said there will never be another exam without proper piloting. That puts our minds at rest," he said.
>
> The failure to pilot the A2 exams in Curriculum 2000 is understood to have masked the fact that better results at AS would inevitably lead to grade inflation in the overall A-level and the problems of last summer.
>
> (Guardian 2003)

But a key question remains – a reassurance may have been given during a period of political exigency – but where in public institutions around and within education are the formal ethical safeguards which regulate policy development and implementation effort?

Case study #2

## Ethics and 'concealed continuity' - the renaming fallacy: successive versions of vocational training for young people

Windscale did not become safer by virtue of being renamed Sellafield, although it did remove the immediate association with the 1957 accident which sent radioactive material into the UK atmosphere (Crabtree J 1959). Youth training 1980-2006 appears to have undergone the same process of superficial change of identity but not of underlying form.

The twenty-year period saw:

WEEP (Work Experience on Employers' Premises) (1976)
UVP (Unified Vocational Preparation) (1977)
YOP (Youth Opportunities Programme) (1978)
YTS (Youth Training Scheme) (1983)
YT (Youth Training) (1989)
Modern Apprenticeship (1994)

The initial schemes attracted favourable comment from evaluators, welcoming the structure and curriculum focus of the programmes (Evans K et al 1983). However, from late 1980's, commentary became more vocal and critical (Hyland T 1994; Winch C, 2002; Wolf A, 1993; Wolf A, 1995; Steedman H, Gospel H and Ryan P 1998) concerned at: the failure of successive governments to erect a high-quality, high-volume vocational route of reasonable esteem (Hodgson A and Spours D 1997); the outcomes of the schemes for individuals (Robinson P 1997); and the economic insignificance of the output for the economy (Wolf A 2004). The constant changing of name and re-casting of programme content concealed an underlying continuity in the fundamental structural forms of the programmes:

- short-duration training oriented towards attaining minimum competence for labour market entry
- an emphasis on outcomes-based learning rather than professional 'formation'
- no tie-in to any adjuvant policy (eg licence to practice) which shifted labour market incentives and drivers; both for young people and for employers (Oates T 2005)

In particular, without an attempt to engage in the last of these, it was inevitable that volumes would be likely to remain constant, and little improvement would be obtained in esteem, outcome and progression. The introduction of Advanced Modern Apprenticeship heralded a break with this – and was welcomed by critical commentators such as CEP researchers at the LSE (Steedman H et al op cit). Advanced Modern Apprenticeships began to gather momentum (25,000 in 1996 to 135,000 in 2000 'in training') (source: DfES) and included more robust curriculum requirements and a commitment to longer duration, formation-oriented work-based learning (Oates T op cit; Unwin L and Fuller A 2003, DfES and LSC 2004). However, as numbers at Advanced level began to plateau in 2000, the 'renaming effect' kicked in once again. Official participation figures celebrated the dramatic increase in Foundation Modern Apprenticeship – 'in-training' figures up from 28,000 in 1999 to 90,000 in 2001. But while the increase in Advanced Apprenticeship represented a substantial improvement in flows into youth training, the Foundation Apprenticeship figures only increased as a result of the apparent reduction in numbers on 'other' government-funded training. In fact, lower level (level 1 and 2) programmes were being designated as 'Foundation Apprenticeship' – a substitution effect.

This case study is intended to exemplify another breakdown in ethical regulation of policy. Learners and the public were being told that fundamental transformation of the vocational route was underway – in fact the lower level route was simply replicating the structural form of previous programmes. The chances of substantial enhancement of progression for those participating in the programmes (return to individuals, in particular), and improvements in participation (returns to the state and the economy, in particular) were therefore remote. A form of educational 'treatment' was thus perpetuated, participants were *not* informed of the fact that innovation had in fact *not* occurred. In one sense, 'do no harm' had been upheld – there had been no substantial improvement in provision, but no significant withdrawal or deterioration either. But by analogy, it could be considered to be a successive placebo effect – change had been championed but in fact no change had taken place. The key ethical issue here would seem to be consent and information. The 'renaming effect' was misleading – policy innovation promised change in outcomes and effects but with no proven mechanism for yielding such changes ('false hope'?); in addition, learners were not informed of the true nature of the educational treatment which the 'revised' programmes actually offered.

Case study #3

## Ethics and protection against risk - the implementation of the National Curriculum

This case study is brief – the ethical issues I wish to engage with are stark and reasonably simple to analyse, and this despite the scale and gravity of the National Curriculum in England (Colwill I undated; Aldrich R and White J 1998) – and indeed the scale of comment upon its form and effects (Chitty C ed 1993; Osborn M, McNess and Broadfoot P 2000).

Ethical argument for 'no pilots' version 1: During the mid-late 1980s, DES officials repeatedly stated that to test out a full prototype national curriculum offer with a selected group of pupils would be ethically unjustified. Their reasoning was that education is so fundamental to an individual's life chances that it is unjustified to 'experiment' with a small group. This sentiment/rationale remained in place for a considerable period, and was evident in the input of DES officials to the discussions of GNVQ piloting referred to elsewhere in this chapter (mid 1990s).

Ethical argument for 'no pilots' version 2: However, a researcher interviewed by the author who was at that time at the NCC (National Curriculum Council) stated that other senior officials offered a very different rationale for 'no pilots'. They stated that '…the National Curriculum is clearly of such benefit to learners that its deployment should not be delayed by piloting…large cohorts of children could lose out as a result of these delays…'.

The first ethically-based argument outlined above is deeply flawed and contradictory. The practical implication of the position is to bring about a situation where rather than experimenting on a small group, ALL pupils become part of a grand national 'experiment'. Experiment is of course an inappropriate term, since the introduction of the National Curriculum was not structured as an experiment – it was an innovation which was treated as 'must succeed' – despite the high level of innovation involved in specifying subjects as outcomes, developing assessment approaches, etc. The ethical principle enshrined in the argument that pupils should not be exposed to undue risk since 'education is so important to each individual' is quite simply grossly contravened by exposing all pupils to risk. Not only are individuals potentially compromised, but the whole system was reconfigured to deliver the national curriculum, which without piloting, was of unknown benefit and possessed unknown effects.

The second ethically-based argument is flawed to a similar degree, but in a different way. Without piloting, the balance of benefits of the national curriculum was simply not known.

After more than a decade of operation, the benefits of the National Curriculum have been comprehensively catalogued (Chitty C op cit; Colwill I op cit) – and include help with pupil transfer in the system (Dobson and Polley C 2004; Ewans D 2005); balanced subject coverage in Primary phase education (Sammons P et al 2003); content which encourages enhanced performance by girls in areas such as maths (Elwood J and Comber C 1996); enhanced development of skills (SCAA 1997). However, the balance sheet of benefit and deficit also includes a long list of serious matters such as acute overload (remedied in part by Dearing's first review); overbearing assessment (Wiliam D 2001); adverse impact of assessment on teaching and learning approaches (Wiliam op cit; Osborn M, McNEss and Broadfoot P op cit); reduction of curriculum innovation (White R 1997); marginalisation of certain subjects in the 'battle for room' (Rawling E 1999).

The urgent and successive revisions to content and assessment (Guardian 2004) hint at the problems which were caused by a failure to pilot the National Curriculum. Changes were hastily formulated in the face of teacher boycott of national tests in 1993 (Alexander R, Willcocks J and Nelson N 1996). Duncan Graham's autobiographical

account of the development and implementation of the National Curriculum provides the sole reference for the decay of public accountability mechanisms around the development work, is itself devoid of explicit consideration of the ethical dimension of mass innovation, yet includes tantalising glimpses of concern at the direction and pace of change: '….it can be argued that many of the 1988 reforms were already in place without some of the prescription and detail which was to plague the introduction of the national curriculum. There is, therefore, a valid case for saying that in 1988 the need for a national curriculum was less urgent than it has been in the early 1980s…' (Graham D & Tytler D 1993 p5).

From an ethical standpoint, the National Curriculum stands as a mass innovation with inadequate ethical controls in relation to risk to students, and indeed in respect of the risks to the integrity of the education system as a whole. The seemingly ethical argument deployed against piloting thus has the form of legitimation deriving from utility frameworks rather than ethical frameworks.

Case study #4

## Ethics and temporal discontinuity between evaluation and policy formation - the peculiar case of the GNVQ, a 10-year-old qualification which never really made it out of it pilot phases

The history of the GNVQ is extremely peculiar. It was designed by six people, including the author, as a result of policy analysis which suggested that a vocational qualification designed for delivery in full-time educational settings was needed both to provide a government-sanctioned, high-status vocational route, and to protect the NVQ from an increasing tendency towards delivery in FE colleges rather than workplaces. The original formulation in fact predated the government invitation to develop the qualification. The design process at NCVQ (the National Council for Vocational Qualifications) was well underway before an interested political administration sought to embrace the initiative as a government-sponsored innovation.

The design initially included some highly unusual features.

Firstly, the qualification was fully modular/unit-based, in line with NVQ criteria. However, the number of units was determined by an intention to articulate with A-levels. The 12-unit Advanced GNVQ was designed to occupy two-thirds of a full-time programme, allowing a single 6-module A level (the dominant UCLES model for A levels) to be taken with the GNVQ, or a cluster of GCSE re-sits. This was the first general vocational qualification designed explicitly to articulate with A levels.

Secondly, the original assessment model centred on teacher-based assessment, assembled on a continuous basis in portfolios.

Thirdly, the assessment was mastery-based; the qualification was ungraded.

Fourthly, the qualification required attainment in 6 key skills as a condition of completion.

Finally, the qualification was to be available only at Advanced level (equivalent to A level) and Intermediate level (equivalent to GCSEs). It was felt by the designers that lower level qualifications would not provide sufficient competence to allow progression to technically-demanding employment, nor allow progression to advanced FE and HE; a key requirement. It was also felt that GNVQs would never be delivered in Key Stage 4 (14-16 education) due to the then-overbearing size of the National Curriculum requirements.

Of the key features, only the 6-unit building block structure survived the first months of complex negotiation. With government commitment lined up behind the GNVQ – necessary to release adequate funding for the development – political expectations and scrutiny increased. In order to ensure sufficient system-wide support, governance arrangements were set up with wide representation: the National Curriculum Council, the Schools Examination and Assessment Council, the Department for Education and Skills, the University and Colleges Admissions Council, the Further Education Unit, Her Majesty's Inspectorate for Schools, the Further Education Funding Council and representatives from agencies for Northern Ireland, Wales and Scotland. The power relations in the policy discussions led to constant erosion of the innovative elements of the assessment model and qualification specifications. Reluctant to let these elements be diluted and the original commitments abandoned, NCVQ developers elaborated the assessment model rather than transformed it (QCA undated).

The reluctance to totally abandon innovation in assessment and simply move to a model which replicated A level and GCSE assessment meant that no parties in the development process felt comfortable with the emerging

models (Wolf A 1997). This led to constant, incremental change in both assessment and the qualification specifications (NCVQ 1995; Wolf op cit) – an issue which dogged the entire history of GNVQs.

This case study becomes relevant to the issue of ethical regulation of innovation when the details of the initial pilot are laid out.

As outlined above, the original model was anathema to key members of the first formal governance groups – particularly the DES, HMI, and SEAC. They demanded immediate changes in the assessment model. This was implemented extremely rapidly, since pilot schemes loomed. NCVQ stated that a two-year pilot phase was necessary. This was immediately rejected by ministers and senior DES officials. They stated that 'the most senior levels of Government' – assumed to be the Prime Minister – had instructed that a one year pilot would be entirely adequate. NCVQ developers were appalled at the decision, aware of the implications. The reality of the timeframe was as follows.

To ensure that specifications would be re-written in the light of the pilots and be ready for full implementation in September of the following year, they would need to be presented for accreditation by Easter. They would then be sent to schools at the start of the summer term, to allow adequate preparation for September starts. To allow re-accreditation by Easter, the units would need to be re-drafted in January-March. This allowed ONE SCHOOL TERM for pilot work. Remember that this is a two-year qualification. Remember also that this is a unit-based qualification where the units may be taken in any order. This meant that not only was there only one term in which to review the realities of implementation of the units, a term in which schools would be coming to terms with a highly innovative assessment (and learning) model, but some units might in fact not be taught in ANY schools by the end of the first few months.

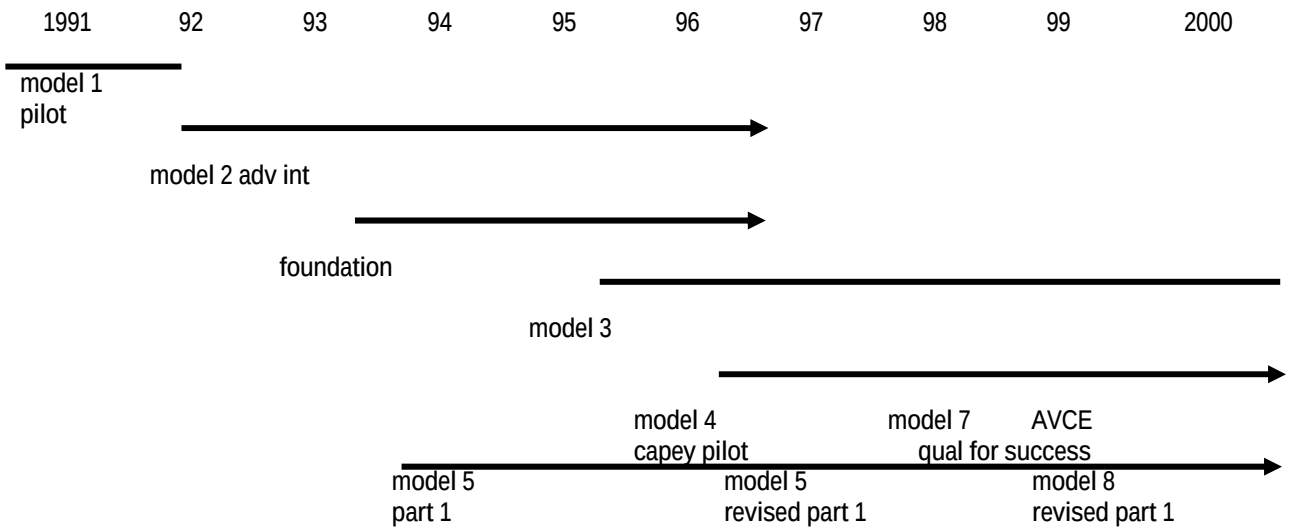
NCVQ were aware that this was likely to lead to ill-conceived alterations to the unit specifications and the assessment approaches – in effect replicating Achtenhagen's 'cycle of planned failure'. However, the degree of political attention and general public interest in GNVQs (fanned not least by the running-down of the Technical and Vocational Education Initiative and its extension (Merson M 1992; Yeomans D undated, Stanley G 2005)) made them realise that they would have to conform with Government requirements. This might not have been a disastrous situation. However, the 'battle for the heart of GNVQ' ensured that it was. As outlined above, NCVQ developers were deeply committed to maintaining the original functions and models at the centre of the GNVQ. Other key agencies within the governance of the qualifications were as equally as committed to abandoning these functions and models and adopting more conventional approaches. NCVQ developers gave ground only slowly and with great reluctance. The effect of this was to condemn the qualification to constant modification. Each modification was undertaken in the same compressed time schedule as the original pilot. On each iteration, solutions to half-understood problems led to changes which themselves only partially addressed the original problems (such as consistency of teacher judgement, management of student portfolios, consistency of awarding body practice, etc) and gave rise to new, unanticipated problems. What is striking is that the nature of the central, confrontational political dynamic meant that the cycle of short-term change became entrenched.

This had very serious consequences. There could be no accumulation of evaluation information, since re-design invalidated prior information and invalidated precise findings. Evaluation data was always grossly incomplete due to the compressed evaluation timescales. Despite the best intentions of evaluators, they simply could not keep up with the frequency and scale of changes. Successive changes were over-determined by sectional interests rather than valid evaluation information. In addition, the nature of the contested territory (learning and teaching, qualification structure, and assessment models) led to an increasingly complex qualification. One thing became increasingly clear as the evaluations rolled on: teachers were increasingly frustrated and confused by the changes. In this tortured development cycle, four remarkable things emerged: the GNVQ candidature continued to increase (albeit driven to a significant degree by FEFC funding arrangements) (Hodgson A and Spours K 1997), distinctive learning styles were established (driven by the grading criteria) and were favoured by candidates (Fitzgibbon C and Meagher N 1995); a highly effective progression route into HE was opened up from Advanced GNVQ (Oates T and Hillier J 1997); and a third 'technical route' of reasonable volume and status opened up in the system (Hodgson A and Spours K op cit).

But despite the gains, the history of the development of GNVQs shows acute and repeated failure at an ethical level. The evaluation-development cycle was essentially out of control, yet persisted for nearly ten years. At no point could learners be assured of evidence-based modification of the qualification. The versions are shown in figure 1.

Fig 1

The development of GNVQs – a qualification always in pilot?



At one point in the life of the qualification (1996) three versions of the same qualification (eg Advanced Level GNVQ in Business) were being taught by teachers to different groups in centres, with some of the version differences being substantial (QCA undated).

Two significant failings at an ethical level are clearly present. Firstly, the original pilot was incredibly innovative in form – yet no consents or compensatory mechanisms were in place. Secondly, as a result of the almost complete and extended breakdown in the evaluation/development cycle, the effects of each revised 'educational treatment' deriving from the modified qualifications were essentially unknown.

As a footnote to the history of GNVQs, one can only wonder at what might have been achieved if developers had been given ten years to move from design to a fully operational mass qualification - including access to all the resources which were in practice mobilised for the successive revisions which took place. In stark contrast to the National Curriculum and C2K – mass innovations without pilot – the GNVQ was destined to forever struggle to escape continuous piloting and revision.

## Where to now?

The DfES recently has strengthened policy in respect of ethical approvals and regulation of research. I am arguing for a gross extension of this: ie for consideration of the ethical dimension of mass innovation. In implementing mass innovation, it would appear incumbent to meet one or more of the following criteria (and avoid infringing any of them):

Criterion 1
Offer an improved performance/experience for a significant number of learners, without significant reduction of the performance/experience in any individuals or learner groups

Criterion 2
Reduce inequalities by raising the achievement/improving the experiences of disadvantaged individuals and groups

Criterion 3
Encourage increased participation which results in a greater supply of knowledge and skills to the individual, society and economy – ie offers individual, social and economic benefit

Criterion 4
Replace existing systems with systems which offer at least the same quality of learning experience and outcomes but which offer improved system performance – eg at a reduced cost

Performance/experience is here construed as including: the immediate experiences within learning programmes; attainments/outcomes; outcomes with currency for progression.

Medical professionals would infringe professional ethics if they changed practice knowing that it reduced beneficial outcomes (Bonell C, Bennet R and Oakley A 2003; Friedson E 1970) or replaced a known therapeutic regime with a regime with unknown characteristics or outcomes. I am arguing here that educational policy-makers possess the same duty of care towards learners (particularly those individuals of age 5-16 who have had their liberty withdrawn by law and are required to attend compulsory schooling (Tapper T and Slater B op cit; Lawton D 2006)) and should not replace elements of existing systems of a known level of performance with new practices which give a lower level of performance or have unknown characteristics or outcomes. This is the hub of the case for ethical regulation of practice, policy-formation and system development.

This then raises the question of how to undertake experimentation and development in education. Again, practice in the medical arena provides a critical lead. There are three key elements:

1
obtain consents and provide means for individuals to withdraw from experimentation and development work

2
put in place adequate safeguards to protect participants in experimental work, including cessation of experimental work if unduly poor outcomes become evident

3
provide compensatory arrangements if participation in experimental work compromises the outcomes of individuals participating in the study

This chapter is not intended to state what a framework of ethics for regulating educational policy and its implementation in practice – and particularly the management of innovation – might contain in detail. What it attempts to provide is a strong argument for the existence of such a framework and its pre-eminence in controlling innovation – ie to lay down a legitimation for the central role of such a framework in policy-making in education. I believe that the problems which I have highlighted through case study are serious, and have resulted in exposing large numbers of learners to substantial risk. Innocents have not been protected. I believe that my arguments have answered the first of the following questions with a robust 'yes'. The others remain open for further public discussion.

1
Is there a need for an ethical framework regulating mass innovation in education and training?

2
Who decides on the content?

3
What are the mechanisms by which such a framework would be enforced?

4
How can innovation be managed in an education and training system with powerful accountability measures – particularly when participation in innovation might compromise individual practitioners' and education/training providers' standing in the performance tables etc?

And finally….the case studies have revealed that the temporal problems of the evaluation-policy cycle are acute. One implication of my analysis is that the timing of political cycles and the timing of ethically-regulated and fully-evaluated mass innovation are irreconcilable. In answer to 3 above, it may be that bold moves to take mass educational innovation out of single-term politics may be required – perhaps an educational version of the Bank of England's Monetary Policy Committee?

Tim Oates
Cambridge
February 2007

(vers 16 02 07)

References


Aldrich R and White J 1998 National Curriculum: beyond 2000 the QCA and the aims of education Institute of Education, University of London

Alexander RJ, Willcocks J and Nelson N 1996 Discourse, pedagogy and the National Curriculum Research papers in education 11 (1) pp81-120

ATL 2002 Work, work, work; students' perceptions of work/life balance under Curriculum 2000 Association of Teachers and Lecturers

Ball SJ 2001 "You've been NERFed!" Dumbing down the academy: National Education Research Forum: "a national strategy – consultation paper: a brief and bilious response" Journal of education policy 16 (3) 265-271

BBC 2001 New exams 'stress out' sixth formers BBC News 11 04 2001 British Broadcasting Corporation http://news.bbc.co.uk/1/hi/education/1272225.stm

Biken SK and Pollard D 1993 Gender and education National Society for the Study of Education

Bhaskar R 1975 A realist theory of science Harvester

Bhaskar R 1979 The possibility of naturalism: a philosophical critique of the contemporary human sciences Harvester

Bonell C, Bennet R and Oakley A 2003 Sexual health interventions should be subject to experimental evaluation in Stephenson JM, Imrie J and Bonell C (eds) Effective health interventions: issues in experimental evaluation Oxford University Press

Boreham N 2002 Work process knowledge Routledge

Bottley M 2000 Education, policy and ethics Continuum

Chalmers I 2005 If evidence-based policy works in practice, does it matter if it doesn't work in theory? Evidence and policy 1 (2) pp227-242

Chitty C 1993 The National Curriculum: is it working? Longman

Colwill I undated What has the national curriculum ever done for us? Qualifications and Curriculum Authority

Cook TD 2003 Why have educational evaluators chosen not to do randomized experiments? Annals of the American Academy of Political and Social Science 589 pp114-49

Crabtree J 1959The travel and diffusion of radioactive material emitted during the Windscale accident Quart. J. Royal Meteorological Society 85, 362

Davies H et al 2000 What works? Evidence-based policy and practice in public services Bristol: The Policy Press

DfES 1999 http://www.standards.dfes.gov.uk/sie/eic/

DfES and LSC 2004 21st Century apprenticeships, end to end evaluation of the delivery of Modern Apprenticeship Department for Education and Skills, Learning and Skills Council

DfES 2006 NLS intervention programme http://www.standards.dfes.gov.uk/primary/faqs/literacy/702497/?subject=S_899187

Dobson J and Pooley C 2004 Mobility, equality, diversity: a study of pupil mobility in the secondary school system University College London

Elliot J 1999 Action research for educational change Open University Press

Elwood J and Comber C 1996 Gender differences in examinations at 18+Institute of Education, University of London

Eraut M 1994 Developing professional knowledge and competence Falmer Press

Evans K et al 1983 Innovation in continuing education and training in the United Kingdom: a report prepared for the European Centre for the Development of Vocational Training University of Surrey Department of Educational Studies

Ewans D 2005 Moving home and changing school – widening the analysis of pupil mobility briefing 2005/32 Greater London Authority

Fitzgibbon C & Defty N, undated How effective are interventions designed to help under-aspiring pupils? http://www.pipsproject.org/renderpage.asp?linkid=30325012

Fitzgibbon C and Meagher N 1995 Analysis of learning styles in GNVQs – ALIS data National Council for Vocational Qualifications

Fischer M 2004 European perspectives of workplace knowledge Office for Official Publications of the European Commission

Friedson E 1970 Profession of medicine: a study in the sociology of applied knowledge Dodd Mead and Co

Geber B 1977 Paiget and Knowing Routledge and Kegan Paul

Goldstein H1991 Better ways to compare schools? Journal of Educational. Statistics 16 (2) 89-91

Goldstein H and Blatchford P 1998 Class Size and Educational Achievement: a review of methodology with particular reference to study design. British Educational Research Journal 24 ( 3)

Goodhart C 1984 Monetary theory and practice: the UK experience Macmillan

Graham D & Tytler D 1993 A lesson for us all – the making of the National Curriculum Routledge

Guardian 2003 Government admits failure to pilot new A level 23 07 2003 http://education.guardian.co.uk/alevels2003/story/0,,1004337,00.html
Guardian 2004 The SATs story http://education .guardian.co.uk/sats/story/0,,1289880.html
Hammersley M 2005 Countering the 'new orthodoxy' in educational research: a response to Phil Hodkinson British Educational Research Journal 31 (2) pp139-155
Higham J and Yeomans D Policy memory and policy amnesia in 14-19 education: learning from the past? Nuffield review of 14-19 education and training Nuffield
Hammersley M 1997 Educational research and teaching: a response to David Hargreaves' TTA lecture British Educational Research Journal 23(2) pp141-61
Hillage J, Pearson R, Anderson A and Tamkin P 1998 Excellence in research on schools Research Report RR74 Department for Education and Employment
Hodgson A and Spours K 1997 Dearing and beyond Kogan Page
Hodgson A and Spours K 2004 14-19 education and training in England: a historical and systems approach to policy analysis Institute of Education; University of London
Hodgson A, Spours K and Waring M 2005 Higher Education, Curriculum 2000 and the future reform of 14-19 qualifications in England Oxford review of education 31 (4) dec 2005 pp475-495
Hodgson A, Steer R, Spours K, Edward S, Coffield F, Finlay I, and Gregson M 2006 Learners in the Learning and Skills Sector – the implications of half-right policy assumptions. Research Report 3 July 2006 TLRP London University Institute of Education
Hodkinson P 2001 NERF strategy proposals: a major threat to acadmic freedom Research Intelligence no74 pp20-2
Hodkinson P 2004 Research as a form of work: expertise, community and methodological objectivity British Educational Research Journal 30 (1) pp9-26
Hyland T 1994 Competence, education and NVQs dissenting perspectives Cassell Education
Kingdon M and Stobart G 1988 GCSE examined Falmer Press
Lather P 2004 The disciplining of education: New languages of power and resistance Trentham Books pp21-36
Lawton D 2006 Education and Labour Party ideologies Routledge Falmer
Lea R 2002 Education and training – a business blueprint for reform Institute of Directors
LSDA 2001 Curriculum 2000+1 Learning and Skills Development Agency
McCall WA 1923 How to experiment in education Macmillan
Merson M 1992 The four stages of TVEI: a review of policy British Journal of Education and Work 5 (2) pp5-18
Miles MB 1964 Innovation in education Bureau of Publications Teachers College Columbia University
NCVQ 1995 GNVQ assessment review (Capey Report) National Council for Vocational Qualifications
NERF undated http://www.nerf_uk.org/bulletin/ (National Education Research Forum)
Oakley A 2006 Resistances to 'new' technologies of evaluation: education research in the UK as a case study Evidence and policy 2 (1) 63-87
Oates T and Hiller J Parity of esteem – both cause and effect Journal of Curriculum Studies
Oates T 2004 The role of outcomes-based qualifications in the development of an effective vocational education and training (VET) system' in Policy Futures in Education ISSN 1478-2103 2 (1)
Oates T 2001 Data-sharing protocol for inter-agency Curriculum 2000 evaluation Qualifications and Curriculum Authority
Osborn M, McNess and Broadfoot P 2000 What teachers do Continuum
Pawson R and Tilley N 1997 Realistic evaluation Sage Publications
Peters RS 1970 Ethics and education Allen and Unwin
Osborn M, McNess and Broadfoot P 2000 (with Pollard A and Triggs P) What teachers do – changing policy and practice in primary education – findings from the PACE project Continuum
QCA undated the story of GNVQs Qualifications and Curriculum Authority http://www.qca.org/610_1807.html
QCA 2004 Press notice: Curriculum 2000 review: QCA's report on phase 2 Qualifications and Curriuculum Authority http://www.qca.org.uk/2586_1841_html
Rawling E 1999 Geography in England 1988-98 – costs and benefits of national curriculum change International research in geography and environmental education 8 (3)
Robinson P 1997 The mtyh of parity of esteem: earnings and qualifications Discussion paper no.354 Centre for Economic Performance London School of Economics
Stenhouse L 1975 An introduction to curriculum research and development Heinemann Education
Stenhouse L 1983 Authority, education and emancipation: a collection of papers Heinemann Education
Stenhouse L 1985 Research as a basis for teaching: readings from the work of Lawrence Stenhouse Heinemann Education
Salter B and Tapper S 1981 Education, politics and the State: the theory and practice of educational change McIntyre

Sammons P et al 2003 The early years transition and special needs project (EYTSEN) Department for Education and Skills
Sebba J 2004 Developing evidence-informed policy and practice in education in Thomas G and Pring R (eds) Evidence-based practice in education Open University Press
Shavelson RJ and Towne L (eds) 2002 Scientific Research in Education National Research Council
Sokal A and Bricmont J 1998 Intellectual impostures: postmodern philosophers' abuse of science Profile Books
Stanley G 2005 Operational selection policy OSP31 Post 16 Education 1974-1988 The National Archives
Steedman H, Gospel H and Ryan P 1998 Apprenticeship A strategy for growth Centre for Economic Performance London School of Economics
Sylva K et al 2004 The effective provision of pre-school education (EPPE) Technical paper no12 The Final Report DfES Institute of Education, University of London
Tapper T and Salter B 1978 Education and the political order: changing patterns of class control Macmillan
TLRP http://www.tlrp.org/ (Teaching and Learning Research Programme)
Tomlinson M 2004 14-19 Curriculum and Qualifications Reform DfES
Tooley J and Darby D 1998 Educational research: A critique – a survey of published education research OFSTED
UCAS/QCA 2003 UCAS/QCA report on Curriculum 2000 QCA http://www.qca.org.uk/7496.html
Unwin L & Fuller A, 2004 Learning as Apprentices in the Contemporary UK Workplace: creating and managing expansive and restrictive participation SKOPE Research Paper, Universities of Oxford and Warwick
White R 1997 Curriculum innovation: a celebration of classroom practice Open University Press
Wiliam D 2001 Level Best? Levels of attainment in national curriculum assessment Association of teachers and lecturers
Walter GA 1981 Experiential learning and change: theory, design and practice Wiley
Whitebread D 2003 Teaching and learning in the Early Years Routledge Falmer
Winch C, 2002 Work, well-being and vocational education, paper given to the Oxford Easter Conference of The Philosophy of Education Society of Great Britain
Wolf A, 1993 Assessment issues and problems in a criterion-referenced system Further Education Unit
Wolf A, 1995 Competence-based assessment Open University Press
Wolf A 1997 GNVQs 1993-1997 A national survey report Further Education Development Agency
Wolf, A. 2002. Does Education Matter? Myths about education and economic growth. London: Penguin Press.
Worth K 2001 Samuel Beckett's Theatre Oxford University Press
Yeomans D undated Constructing vocational education from TVEI to GNVQ Post 14 Research Group Research Paper University of Leeds