Cambridge Journal of Education
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713410698

Curriculum Evaluation Revisited

Nigel Norris [a]
[a] Centre for Applied Research in Education, University of East Anglia.

Online Publication Date: 01 June 1998
To cite this Article: Norris, Nigel , (1998) 'Curriculum Evaluation Revisited', Cambridge Journal of Education, 28:2, 207 - 219
To link to this article: DOI: 10.1080/0305764980280206
URL: http://dx.doi.org/10.1080/0305764980280206

Curriculum Evaluation Revisited

NIGEL NORRIS
Centre for Applied Research in Education, University of East Anglia

CONTEXT

Teach Your Children Well

In many countries the organisation of school systems is subject to intense pressure. There is talk of reforming, restructuring or even re-inventing schooling (Crowson *et al.*, 1996). Concern about the effectiveness and efficiency of schooling and its contribution to economic well-being has created a climate of intense scrutiny. In a number of societies a powerful mixture of budget deficits, managerialism, business ideology and populism has provided the political impetus for various kinds of 'reform'.

In some cases reform embraces the deregulation of the curriculum or the decentralisation of curriculum decision making and the devolution of power to the school or community (Lawton, S.B., 1992; Norris *et al.*, 1996). Under these circumstances evaluation is thought to provide the information needed to maintain government 'control and steering of the quality of education' (Laukkanen, 1995, p. 17). In other cases reform involves what Michael Fullan (1991, p. 16) has called 'intensification', which he described as the

> Increased definition of curriculum, mandated textbooks, standardized tests aligned with curriculum, specification of teaching and administrative methods backed up by evaluation and monitoring ...

It is not just the organisation of schooling that has become a hotly contested issue. What the curriculum should be and how it should be constructed and assessed is in a state of flux. Notwithstanding the widespread commitment to nationalisation of the curriculum, the content of formal education is under pressure from globalisation, rapid technological change and changes in the structure of labour markets as well as the forces of tradition and nostalgia. The organisation of schooling and its academic and moral content is a major site of conflict over social futures (Paquette, 1991; Goodson, 1997). Where the school curriculum is conceived as an expression of national or cultural interests the power of teachers to select from culture what is meaningful for pupils and relevant to their needs is considerably reduced. Consequently, education is less immediate, less tractable, less responsive to the motivation of the child, less likely to reflect local needs and circumstances and curriculum evaluation is

0305-764X/98/020207-13

much more likely to be defined as something that is done to schools not by schools.

Ubiquity

Curriculum evaluation emerged as an organised and developing body of experience in the context of educational innovation. It was investment in planned change that prompted a concern for curriculum evaluation. New curricula lacked the warrant of history and, unlike established practice, were more likely to raise questions about quality, the value of innovation and the essential features for its success. Beginning in America in the 1930s and Britain in the 1950s, the demand for curriculum evaluation increased rapidly as the pace of social change accelerated and the political complexity of educational innovation became apparent.

Barry MacDonald (1982) once observed that evaluators were the camp followers of education reform movements. Educational evaluation was initially seen as an aspect of curriculum development. Before long, however, it became associated with accountability, quality control, systems maintenance and resource allocation. In the USA evaluation has been a mandated requirement associated with federal and state programmes in education for some time. Internationally there has been massive investments in the development of performance indicators to monitor social and educational performance.

While still a young discipline, evaluation has grown beyond measure (Cronbach, 1981; House, 1993). According to Pawson & Tilley (1997, p. 2) evaluation has become 'a mantra of modernity'. We live in what has been described as an 'evaluative state' (Neave, 1988; Henkel, 1991), where evaluation procedures and processes are a routine part of management and administration and a mandatory requirement of government investment in public services (Norris, 1995).

In these developments the focus of attention, the lodestar of evaluation, has been the measurement of congruence between objectives and performance, goals and outcomes. The emphasis on objectives achievement and accountability arises from a number of political interests and concerns: Are people doing what they are supposed to be doing?; Is public money being used for the purposes for which it was intended?; What is the pay-off from public investment in education, health, welfare, criminal justice?; Are public services meeting their targets?; how can the efficiency and effectiveness of services be improved? Like other areas of evaluation, the practice of educational evaluation is about social planning and control and the key value is that of order (Norris, 1990).

Definitions

Put simply, curriculum evaluation is about describing the meaning, values and impact of a curriculum to inform curriculum decision making. But behind this statement there are a multitude of political perspectives and methodological

complications. What curriculum is, what it ought to be and how it should be organised are not matters about which there is agreement, far from it. Similarly, there are many different approaches to curriculum evaluation and marked differences between its practice and theory.

What it is to evaluate the curriculum partly depends on how curriculum is conceived. Denis Lawton (1984, 1989) argues that curriculum is a selection from the culture of a society. This general definition of curriculum raises a number of issues. First, who should select those aspects or elements of culture for the school curriculum. Second, how should such a selection be made and justified. Third, how should this knowledge be organised. Fourth, how is the culture of a society to be conceived, especially given the contradictory trends in many societies towards cultural pluralism, cultural maintenance and the globalisation of culture. Fifth, how should the effects of particular expressions of culture become known and evaluated. If curriculum is a selection from the culture, then how that selection is made, the values it promotes and how it is mediated by the school become important evaluative questions.

Others have conceived the curriculum simply as a course of study to be followed or a written prescription of what it is intended should happen in schools. If a curriculum is a statement of intentions, it follows that curriculum evaluation should be about the extent to which intention is realised in practice. Curriculum is also sometimes thought of as all that happens in the school. If this is so, then to evaluate the curriculum is evaluate the school and curriculum and school evaluation are the same thing. To evaluate the totality of school experiences is a very substantial undertaking. Nonetheless, there are ways of deepening understanding of what happens in schools and improving the quality of educational experiences. To be affordable, formative and supportive of development such efforts have to be local and largely based on the internal evaluative capacity of the school itself. This is what Barry MacDonald (1978) had in mind when he proposed a process model of school accountability based on the idea of the self-report [1]. In many societies teachers are held responsible for what happens in schools yet have no power to change the curriculum and few resources to develop or evaluate their practice. Even where power over the curriculum is devolved, school-based self-evaluation lacks credibility both as a process of improvement and as an instrument of accountability. Instead of a detailed knowledge of the particular processes and circumstances of education forming the basis of evaluation, schools are largely known through summary judgements of their performance.

PRACTICE

Broadly speaking there are six commonly used and routine approaches to curriculum evaluation: experimentalism, the objectives achievement model, performance indicators, self-study, expert or peer review and inspection. Taken together, these six approaches represent much of the daily practice of evaluation.

That Which Purifies Us Is Trial and Trial Is by What Is Contrary

Some methodologists have argued that the only way to settle disputes regarding educational practice and verify educational improvements is to use the procedural rules and controlled conditions of experimental and quasi-experimental designs (Campbell & Stanley, 1966; Cook & Campbell, 1979; Rossi *et al.*, 1979; Saxe & Fine 1981; Boruch, 1997). Experimentalism appeals to those who see education and its control as susceptible to the production and use of reliable, accurate, valid and, most importantly, generalisable scientific knowledge. By ruling out alternative explanations and threats to validity, properly conducted and evaluated experimental trials distil the truth about curricular effects and effectiveness.

For some politicians and administrators the experimental comparative validation of alternative curricula promises facts on which to adjudicate between competing claims for public investment or to support budget requests. The appeal of seemingly definitive scientific judgement on the merits of particular curricula is moderated by the costs and problems of using experimental designs to evaluate education. Very few innovations are designed as experiments and educational provision is not organised in such a way as to support rigorous scientific experimentation. There is little political commitment to the values or practices necessary for an experimental society characterised by evidence-based policy making and an openness and willingness to change. Nonetheless experimentalism and quasi-experimental designs are very influential in the way people think about evaluation and in the implicit standards they hold for determining certain kinds of value questions.

Let Me Know Mine Ends and Measure Them

The view that evaluation is essentially the process of determining if objectives are achieved remains the mainstay of evaluation practice. Routine evaluative activity takes this form in many areas of public life. In education, the objectives achievement model has a long history of development and application. It was originally conceived and refined by Ralph Tyler as an integral part of curriculum development in the 1930s.

For Ralph Tyler (1949) education was essentially about changing the behaviour of pupils, thus the curriculum could be constructed through the specification of desirable behaviours, which were stated as objectives to be achieved. Evaluation then became a relatively simple matter of finding out to what extent the stated objectives were realised by the programme of curriculum and instruction. This approach to curriculum evaluation has great common sense appeal. The curriculum objectives serve as a consensual source of standards and criteria for evaluation, thus, to a certain extent, avoiding potential conflict about the value of a curriculum. The criterion of effectiveness is transparent, since the congruence between stated goals and actual outcomes is the measure of programme success. Because the curriculum is evaluated in

terms of student achievements, assessment data is the main source of evaluative evidence, making evaluation economical and permitting comparisons to be made over time. Moreover, the logic of the objectives achievement model can, at face value, be applied at any level of the system. It can be used by teachers as a tool to help plan the curriculum and make instructional decisions. It can be used by the school as a whole to monitor overall performance and identify potential problems. It can be used by curriculum developers to test their ideas in practice and the same logic can be used by local and national education authorities to monitor the performance of the system as a whole.

The simplicity and directness of the objectives achievement model of evaluation has been enormously persuasive and pervasive. It has serious limitations, however. Not all educational intentions can be described with reference to visible behaviour. Many of the outcomes of schooling are multi-dimensional and cannot be adequately represented as measurable learning objectives. The prescriptive nature of detailed and tightly defined learning objectives is antithetical to the development of personal autonomy and expression (Stenhouse, 1975). A significant problem with the objectives achievement model is that unintended consequences tend to get ignored. It also represents an extreme relativisation of evaluation, since it contains no procedures for judging the worth or value of goals (Scriven, 1967). Mostly such criticisms have not been heeded. In his original conception of curriculum evaluation Ralph Tyler saw it as a tool for the teacher. Others saw it as a way to by-pass teacher judgement and as a tool for surveillance and control.

Every Breath You Take, Every Move You Make, I'll Be Watching You

A development closely connected with the objectives achievement model is the use of performance indicators as evaluation instruments. Performance indicators are usually constructed to reflect and record change over time across significant dimensions of an organisation or system. For example, the Audit Commission collects annual performance indicator information on English and Welsh local council performance across an array of services including education, libraries, social services and housing (CIPFA, 1997).

Central to the construction and use of performance indicators are decisions about goals and what is of most importance and value. Performance indicators are associated with the political imperative for more effective and efficient public services. The concern for effectiveness is manifest in the specification of objectives, the measurement of progress towards them and consideration of alternative courses of action to reach the same end. The concern for efficiency usually reflects an interest in minimising inputs to achieve the same level of output or maximising outputs with the same level of input.

The most commonly used performance indicators for education are student test results of one kind or another; external public examinations and national or local standardised tests. Although test data can be used to judge individual progress towards learning goals, it is also used to make comparisons between

pupils, teachers, subjects or departments, schools, school districts and nation states. In most education systems it is the teachers who are seen as responsible for changes in test results. The methodological inadequacies of standardised tests for evaluating the effectiveness of teachers or schools have been known for some time (Glass, 1974). Nonetheless, testing for the purposes of comparative teacher and school judgement has continued unabated.

The consequences of the use of large-scale national testing to evaluate the impact of curriculum and the performance of schools and teachers are becoming all too clear (Haney & Madaus, 1986; Darling-Hammond, 1991; Madaus, 1992; Torrance, 1997). As measurement and assessment specialists could have predicted, there are severe problems with the validity of comparative inferences that can be drawn from indicators. The measurement and assessment issues associated with large-scale testing for comparative purposes are probably poorly understood by potential users, many of whom are likely to have more confidence in the results and their straightforward interpretation than is warranted by the methodology. If important decisions are presumed to be related to test results, then teachers are likely to teach to the test and the curriculum will be correspondingly narrow. When teachers teach to the test, then the assumption that student performance on the knowledge sampled by the test fairly represents the broader domain of curriculum goals is violated and inferences of this kind are invalidated (Haney & Madaus, 1986). Perhaps most corrosive of educational values is where test scores are used for making decisions about school and teacher rewards and sanctions, thus encouraging schools to exempt poor students from the curriculum and assessment or help them take the tests or to cheat. High stakes indicators are subject to corruption and are likely to distort the social processes they are intended to monitor (Campbell, 1976).

Know and Better Thyself

The development of an internal curriculum evaluation resource offers the school (and possibly teachers) significant advantages. In Britain throughout the 1970s and 1980s ideas such as the 'teacher as researcher' (Stenhouse, 1975), the 'self-monitoring teacher' (Elliott, 1978), 'institutional self-evaluation' (Adelman & Alexander, 1982; Mitchell, 1984; Nixon, 1992) and the 'teacher as evaluator' (Simons, 1987) were seen as ways of improving the curriculum in action. Institutional self-study, or more accurately review, was directly supported by a number of local education authorities (LEAs) producing self-evaluation guidelines. An early and influential example of an LEA-sponsored school self-valuation initiative was the Inner London Education Authority's *Keeping the School Under Review* (Inner London Education Authority, 1977). This was a method of self-evaluation for schools devised by the authority's inspectorate. The method consisted of a series of questions that prompted the school to review its organisation, policies and practices; in effect a check-list. Other initiatives, such as the Schools Council's GRIDS (Guidelines for Review and Internal Development) project produced advice for schools on how to conduct

a school review and implement improvements. The GRIDS project recommended a five step process for school improvement—getting started, initial review, specific reviews, action for development, overview, restart. Each of these steps was further broken down into its component parts. Helen Simons (1987, p. 220) found that by the end of the 1980s 'approximately three quarters of the LEAs in England and Wales had initiated discussions on school self-evaluation and one fifth had issued guidelines to schools'.

Underlying the development of institutional self-study is the belief that teachers working together in their schools are the most effective means and focus for improving education. Early approaches to self-evaluation tended to emphasise teachers researching their own classroom practice. In the USA Waples & Tyler (1930) were advocating that teachers do classroom investigations on their own problems as early as the 1930s. In Britain Lawrence Stenhouse (1975, 1983) argued that curriculum research and development ought to belong to the teacher and that teacher research should provide the basis for teaching. Teachers researching their own practice has the advantage that the producer and consumer of evaluative information is the same person. Yet it is very hard to sustain a practical commitment to classroom research as the basis for curriculum evaluation and professional development in the context of central curriculum control and external demands for accountability. Self-evaluation has low credibility and authorities that are suspicious of professional power are unlikely to support self-reporting as the basis of accountability. As the locus of curriculum decision making shifts from the classroom and school to local and national authorities the audiences for curriculum evaluation have similarly changed and formative evaluation for local development has given way to summative evaluation for surveillance and monitoring.

Experto Crede

External expert consultants are frequently asked to evaluate the curriculum. This approach to evaluation is based on the use of professional judgement and expertise to review the quality of curriculum, teaching and learning. Sometimes the focus of evaluation is on learning products and the experienced and expert eye of an outside peer is brought to bear upon the outcomes of an education to make judgements about its qualities. Other examples of this approach include: (i) professional associations reviewing and accrediting the curriculum plans of institutions providing professional training; (ii) subject associations or a panel of subject experts, usually drawn from the university, reviewing the content and organisation of school subjects such as physics, geography or literature.

This form of evaluation emphasises the values of a discipline and the professional wisdom of peers. Its purpose may be to validate or accredit the curriculum or it may be to challenge the perceptions of developers and teachers. The use of expert panels in curriculum evaluation obviously raises questions about the values that are brought to bear in the process of judgement. The

power of expert panels rests on the authority and expertise of the individuals concerned rather than on a particular methodology.

The Real Inspector Hounds

Inspection is a form of evaluation. Typically it involves an external small team of inspectors (experts, connoisseurs, senior members of the profession or trained personnel and sometimes including 'lay' representatives) visiting institutions, scrutinising and assessing written evidence, using performance indicators assembled specially for the visit, interviewing or interrogating staff, talking to clients, consumers or other partners and stakeholders, observing practices and reporting on individual and institutional strengths, weaknesses and various aspects of quality. Judgement lies at the heart of inspection. It may follow largely private norms and values grounded in accumulated professional wisdom and feel, or it may, and increasingly does, follow explicit procedures, processes and criteria. For reasons of economy inspections are usually brief affairs by comparison with the time that would be needed to adequately understand an institution. The real costs of inspection are also shared between the inspection or commissioning agency and the institution. The costs to the institution of preparing for an inspection can be considerable.

Inspection might be regarded as a form of expert review, although when it is based on following a handbook, manual or strict set of guidelines the claim to special expertise is undermined. In education inspection is meant to provide both feedback to the school staff, information to parents and accountability for the expenditure of public money. Inspection tends to emphasise independent summary judgement at the expense of advice (Thomas, 1996) or negotiated outcomes. Where inspection is bureaucratic, judgement often rests on conventional criteria (Gilroy & Wilcox, 1997) embodying a standardised model of good practice.

Not surprisingly, teachers can feel threatened by inspections and they are occasions of stress and anguish (Brimblecombe *et al.*, 1995; Dean, 1995). Inspections take up a large amount of teacher time and the true cost of inspections appear to be largely unknown, as are its benefits, although there is some evidence that the process of review it prompts might well be beneficial (Ouston *et al.*, 1997). Given the bureaucratic and only quasi-independent nature of much inspection, it is not surprising that concerns about curriculum tend to focus on issues of fidelity and implementation.

THEORY

Mind the Gap

Over the past 20 years the gap between evaluation practice and evaluation theory has widened. The contemporary theory of evaluation highlights the specificity of programmes and projects, the importance of multiple perspectives,

the interests and information needs of different audiences and the methodological strengths of case study. The emphasis on specificity reflects two important lessons from the evaluation of curriculum development initiatives over the last 30 years. First, that curriculum development occurs within a structure of opportunity and constraint. To understand why a curriculum is as it is, one must first describe in detail the circumstances and processes of its production. Second, that the salience of school context and teacher interpretations in determining curriculum outcomes is usually underestimated by developers and policy makers alike.

Multiple perspectives have assumed greater significance as evaluators have realised that curricula mean different things to different people and that varied sources of measurement, interpretation, values and judgements are essential to sound evaluation. The school and the teacher are part of a web of relationships which influence performance. This means that for evaluation to be maximally useful it must serve a range of different interests and audiences that make up this web of relationships. Most importantly, evaluations that are limited to the immediate needs of policy makers or curriculum developers are unlikely to be helpful to teachers, who carry the main burden of curriculum implementation.

As Bob Stake (1967) argued over 30 years ago, curriculum evaluation should reflect the fullness, complexity and importance of educational programmes. Curriculum evaluation should go beyond the measurement of outcome data as reflected in conventional achievement tests and focus on antecedent conditions and classroom transactions, paying more attention to both description and judgement. Yet the routine administration of curriculum evaluation remains stubbornly resistant to the methodological development needed to understand and improve education. In contrast to the particular and heterogeneous nature of curriculum in classrooms and schools, evaluation is treated as a standardised process of description, measurement and reporting which largely ignores difference and context. Instead of multiple perspectives and valuings there is a tendency towards the use of a limited range of performance indicators and test data as the major evidential source for judgement. These forms of evaluation, largely derived from an interpretation of the objectives achievement model, rely on a restricted understanding of the contingencies on which successful schooling depends. They are apt to distort the very processes they are intended to measure and support the attribution of blame rather than the processes of development.

If the curriculum and the work of teachers are to be judged fairly and if judgements about their value are to have some claim to validity, then they must be adequately and accurately represented, given their constraints and opportunities. This means that proper attention must be given to the particularity of curriculum. Since the purpose of evaluation is to help policy makers, schools, teachers and others make decisions about how to implement or improve curriculum, then evaluation must include an account of the particular contexts of the curriculum, how it is practised in specific situations and what are the salient features of success and failure.

Even national curricula are particular things, particular in two important senses. First, they are particular expressions of the ways in which knowledge can be organised for the purposes of education. Second, when they are enacted in the practice of schools and classrooms they are particular interpretations of a national curriculum in action. To evaluate a national curriculum is to study a particular expression of knowledge and instances of its translation into practice.

THE ART

Designing educational evaluations is an art. As Lee J. Cronbach (1983, p. 1) observed, evaluations need to be designed afresh in each new undertaking. There is no recipe or procedure for evaluation which if followed will produce reliable and valid judgements about the quality of education. Neither education nor evaluation are so mechanistic. Despite efforts to develop one, there is no unitary and uniform system of measurement that can encompass the processes and outcomes of education, let alone improve its practice. The art of curriculum evaluation is much more difficult than the technologies of evaluation admit. They are crude instruments by comparison with the complexities of learning, teaching, classrooms and schools.

The art of curriculum evaluation is to conceive, obtain, construct and distribute information that can be used to improve educational practice. Let us look briefly at each of these activities. Conceiving of information that can be used to improve practice is to have a theory of development relevant to the context of educational action. A crude distinction can here be made between evaluations that support centralised or decentralised curriculum decision making. In evaluation obtaining information is to do more than use the methods of research, it is also to create the conditions of access by which it is possible to collect comprehensive, reliable and useful data. Evaluations are frequently faced with competing interests and values. Issues of territory, competition, hierarchy and privacy impinge on the work of evaluators to a greater extent than is usual in research. The political circumstances of programmes can undermine the methodology of evaluation and affect the kind and quality of evidence available to it. As a consequence, specific measures must be taken to foster the trust and openness necessary for the conduct of evaluation. When constructing evaluative information decisions have to be made about the form and presentation of reports that are most likely to support the development of practice. Finally, the distribution of evaluative reports requires that the evaluator attends to the conditions necessary for the fair exchange and use of information to support development.

It is teachers who translate the curriculum into educational action in classrooms and schools. It is teachers who interpret and give life to the curriculum specifications of governments and ministries. Curriculum evaluation must be related to and supportive of the work of teachers if it is to develop educational practice.

Often it is not clear what people need to know and do to improve the quality of education. Part of the reason for this is that it is only the least powerful in a school or education system, the pupils, who have any sense of what the experience of the curriculum is like as a whole. Teachers do not encounter the whole curriculum; they are usually responsible for some limited aspect of it, a year group or a subject. Headteachers or principals are even further removed from the experience of curriculum. They may see the timetable and deal with the problems of curriculum management and they may, by wandering about, gain some idea of what happens in and around the school, nevertheless, the curriculum in action is beyond their compass and command.

While teachers make the curriculum happen in their own classrooms and schools, they do so not in circumstances of their own choosing or design. Curriculum evaluation must take account of both the context of schooling and its immediate environment and the policy framework in which teachers have to teach. Outside the ambit of the school the reality of the curriculum becomes more remote and idealised. As we move away from those who have direct and sustained contact with schools to those responsible for planning and policy the curriculum becomes a generalised statement of intent, a specification of what should happen in the name of education. Curriculum planners may see the rational construction of balance, coherence and progression in the curriculum on paper, but they are far removed from what it is to put this into practice and what it is to experience on a day-to-day, year-to-year basis the plans of others. From ministry to classroom, from minister to child is a long way. This is the space that evaluation must fill if it is to develop educational understanding.

Correspondence: Nigel Norris, Centre for Applied Research in Education, School of Education and Professional Development, University of East Anglia, Norwich NR4 7TJ, UK.

NOTES

[1] MacDonald (1978) argued that 'the performance of the school is in part a function of its circumstances and cannot fairly be assessed without detailed knowledge of those circumstances. It is the duty of the school to provide the best possible opportunities for learning consistent with its circumstances. This should be the basis of a school accountability model—a process rather than a product model.... A process model of school accountability could be brought about', says MacDonald, 'by the initiation and development of school self-reports for the local community'. As an instrument of school accountability he notes that the self-report has the following merits: (i) it testifies to the autonomy and responsibility of the school and its professional status; (ii) it locates the development of school accountability firmly in the hands of those most vulnerable to its consequences; (iii) it lets the schools themselves define what they would accept as informed criticism (though they will never have a sole right to define the terms by which they are to be judged); (iv) it offers the best possibility of coordinating information gathering for routine internal purposes with information gathering for accountability; (v) in the absence of models of institutional competence or effective instructional behaviour, it gives schools the opportunity to provide the descriptive basis from which such models might be derived; (vi) it gives schools the right and the

opportunity to define the accountability of their co-actors in the system—those who make policy, provide resources and services and give advice.

REFERENCES

ADELMAN, C. & ALEXANDER, R. (1982) *The Self-evaluating Institution* (London, Methuen).

BORUCH, R.F. (1997) *Experiments for Planning and Evaluation: a practical guide* (Berkeley, CA, University of California Press).

BRIMBLECOMBE, N., ORMSTON, M. & SHAW, M. (1995) Teachers' perceptions of school inspection: a stressful experience, *Cambridge Journal of Education*, 25, pp. 53–62.

CAMPBELL, D.T. (1976) *Assessing the Impact of Planned Social Change*, Evaluation Centre Occasional Paper no. 8 (Kalamazoo, MI, College of Education, Western Michigan University).

CAMPBELL, D.T. & STANLEY, J. (1966) *Experimental and Quasi-experimental Designs for Research*, (Chicago, IL, Rand McNally).

CIPFA (1997) A guide to performance indicators, 1995/96, *Public Finance*, 7 (Suppl.).

COOK, T.D. & CAMPBELL, D.T. (1979) *Quasi-experimentation Designs and Analysis Issues for Field Settings* (Boston, MA, Houghton Mifflin).

CRONBACH, L.J. (1981) *Towards Reform of Program Evaluation* (San Francisco, CA, Jossey-Bass).

CRONBACH, L.J. (1983) *Designing Evaluations of Educational and Social Programs* (London, Jossey-Bass).

CROWSON, R., BOYD, W.L. & MAWHINNEY, H.B. (Eds) (1996) *The Politics of Education and the New Institutionalism* (London, Falmer Press).

DARLING-HAMMOND, L. (1991) Policy uses of indicators, paper prepared as part of an international indicators project developed by the OECD, Teachers College, Columbia University.

DEAN, J. (1995) What teachers and headteachers think about inspection, *Cambridge Journal of Education*, 25, pp. 45–52.

ELLIOTT, J. (1978) Classroom accountability and the self-monitoring teacher, in: W. HARLEN (Ed.) *Evaluation and the Teacher's Role* (Basingstoke, Macmillan).

FULLAN, M. (1991) *The New Meaning of Educational Change* (New York, NY, Teachers College Press).

GILROY, P. & WILCOX, B. (1997) OFSTED, criteria and the nature of social understanding: a Wittgensteinian critique of the practice of educational judgement, *British Journal of Educational Studies*, 45(1), pp. 22–38.

GLASS, G. (1974) Teacher effectiveness, in: H. WALBERG (Ed.) *Evaluating Educational Performance*, pp. 11–32 (Berkeley, CA, McCutchan).

GOODSON, I. (1997) *The Changing Curriculum* (New York, NY, Peter Lang).

HANEY, W. & MADAUS, G. (1986) Effects of standardized testing and the future of the national assessment of educational progress, a working paper for the NAEP study group, Centre for the Study of Testing, Evaluation and Educational Policy, Boston College.

HENKEL, M. (1991) The new evaluative state, *Public Administration*, 69, pp. 121–136.

HOUSE, E. (1993) *Professional Evaluation* (London, Sage).

INNER LONDON EDUCATION AUTHORITY (1977) *Keeping the School Under Review—a method of self assessment devised by the ILEA inspectorate* (London, ILEA).

LAUKKANEN, R. (1995) The formation of evaluation policies in education in Finland, in: Y. YRJONSUURI (Ed.) *Evaluating Education in Finland* (Helsinki, National Board of Education).

LAWTON, D. (1984) Cultural analysis and curriculum evaluation, in: M. SKILBECK (Ed.) *Evaluating the Curriculum in the Eighties*, pp. 15–31 (London, Hodder & Stoughton).

LAWTON, D. (1989) *Education Culture and the National Curriculum* (London, Hodder & Stoughton).

LAWTON, S.B. (1992) Why restructure?: an international survey of the roots of reform, *Journal of Education Policy*, 7, pp. 139–154.

MACDONALD, B. (1978) Accountability, standards and the process of schooling, in: T. BECHER & S. MACLURE (Eds) *Accountability in Education*, pp. 127–151 (Slough, NFER).

MACDONALD, B. (1982) Educational evaluation in the contemporary world, invited presentation to the *Symposium International De Didactica General y Didacticas Especiales*, La Manga del Mar Menor, Spain, 27 September–2 October.

MADAUS, G.F. (1992) A national testing system: manna from above?, Centre for the Study of Testing, Evaluation and Educational Policy, Boston College (mimeo).

MITCHELL, P. (1984) Institutional evaluation: the process within a school, in: M. SKILBECK (Ed.) *Evaluating the Curriculum in the Eighties*, pp. 64–70 (London, Hodder & Stoughton).

NEAVE, G. (1988) On the cultivation of quality, efficiency and enterprise: an overview of recent trends in higher education in western Europe, 1986–1988, *European Journal of Education*, 23, pp. 1–2.

NIXON, J. (1992) *Evaluating the Whole Curriculum* (Buckingham, Open University Press).

NORRIS, N. (1990) *Understanding Educational Evaluation*, (London, Kogan Page).

NORRIS, N. (1995) Contracts, control and evaluation, *Journal of Education Policy*, 10, pp. 271–285.

NORRIS, N., ASPLAND, R., MACDONALD, B., SCHOSTAK, J. & ZAMORSKI, B. (1996) *An Independent Evaluation of Comprehensive Curriculum Reform in Finland* (Helsinki, National Board of Education).

OUSTON, J., FIDLER, B. & EARLEY, P. (1997) What do schools do after OFSTED—or before, *School Leadership and Management*, 17, pp. 95–104.

PAQUETTE, J. (1991) *Social Purpose and Schooling: alternatives, agendas and issues* (London, Falmer Press).

PAWSON, R. & TILLEY, N. (1997) *Realistic Evaluation* (London, Sage).

ROSSI, P.H., FREEMAN, H.E. & WRIGHT, S.R. (1979) *Evaluation: a systematic approach* (Beverly Hills, CA, Sage).

SAXE, L. & FINE, M. (1981) *Social Experiments: methods for design and evaluation* (Beverly Hills, CA, Sage).

SCRIVEN, M. (1967) The methodology of evaluation, in: R.W. TYLER, R.M. GAGNE & M. SCRIVEN (Eds) *Perspectives of Curriculum Evaluation*, AERA Monograph on Curriculum Evaluation no. 1. pp. 39–83 (Chicago, IL, Rand McNally).

SIMONS, H. (1987) *Getting to Know Schools in a Democracy—the politics and process of evaluation* (London, Falmer Press).

STAKE, R.E. (1967) The countenance of educational evaluation, *Teachers College Record*, 68(7), pp. 523–540.

STENHOUSE, L. (1975) *An Introduction to Curriculum Research and Development* (London, Heinemann).

STENHOUSE, L. (1983) *Authority, Education and Emancipation* (London, Heinemann).

THOMAS, G. (1996) The new schools' inspection system, *Educational Management and Administration*, 24, pp. 355–369.

TORRANCE, H. (1997) Can testing raise educational standards?, in: K. WATSON, C. MODGIL & S. MODGIL (Eds) *Educational Dilemmas: debate and diversity*, Vol. 4: *Quality in Education* (London, Cassell).

TYLER, R. (1949) *Basic Principles of Curriculum and Instruction* (Chicago, IL, University of Chicago Press).

WAPLES, D. & TYLER, R. (1930) *Research Methods and Teachers' Problems: a manual for systematic studies of classroom procedure* (New York, NY, Macmillan).