



National Foundation for Educational Research

**EVALUATION OF CURRICULUM REFORM
IN SLOVENIA:
TOWARDS A FRAMEWORK**

FIRST DRAFT

Lesley Saunders

March 1999

1. INTRODUCTION

This document is a supplement to the report *Evaluation of Curriculum Reform in Slovenia: Towards a Framework (First Draft)*. It has been produced in response to the request from the President of the National Council for the Evaluation of Curriculum Reform (NCECR) that the following points be addressed:

- ◆ description of the main alternatives in planning the methodology of evaluation;
- ◆ description of the evaluation procedures, methods and techniques which evaluators of curriculum reforms have found effective in the past (in England and/or elsewhere);
- ◆ specification of the ‘success criteria for evidence of impact of curriculum reform’ as outlined in the first document.

This document provides some discussion of the first two points in the following sections. So far as the third point is concerned, the NFER understands that this is of great interest and importance to the NCECR; but we consider it to be crucial that criteria for the success of reforms are discussed, negotiated and agreed by representatives of the ‘key players’ – it is an essential part of the process of building in evaluation to educational reform that the principles and criteria can be ‘owned’ by those being evaluated, and not just imposed from outside or above. Instead, therefore, we have provided just one or two examples in order to show what is meant by ‘success criteria’.

Several of the suggestions made or implied here would be dependent on some at least of the recommendations made by OECD (Reid *et al.*, 1998) having been put into practice or at any rate taken account of in implementing the reforms. This document is accordingly referred to at several points in the text and a list of the recommendations, excerpted from the text, is given at **Appendix A**.

2. ALTERNATIVE APPROACHES TO EVALUATION

There is no obvious or best way to evaluate curriculum reform and its impact – the problem is one which many countries are grappling with. The key questions for deciding on the most appropriate approach to evaluation in any given instance must be, ‘what is it we most want to evaluate?’ and ‘what do we want to do with the results?’ The main choices can be grouped as follows:

2.1 Formal Consultations and Consultative Reviews

Governments may wish to conduct consultations and consultative reviews to canvass the views and experiences of key stakeholders about proposed reforms, whether specific or wholesale. In doing so, they usually take due account of what has happened or is happening, so as to learn the lessons for the future. Consultation exercises are often contracted out to market research or other appropriate organisations.

2.2 Measurement of the Impact of Reforms

This form of evaluation is the Holy Grail of evaluations – on the one hand, it should provide the answer to the all-important question ‘has the reform worked?’ On the other hand, it is quite difficult in reality to identify the impact of central government initiatives – particularly wholesale reform – even when the desired outcomes of the reform have been specified in advance and are known to be quantifiable and collectable (which they often are not!).

This is because the number of ‘input’ variables which must be taken into account is considerable, ranging from, say, differences in leadership skills between different headteachers/principals which affect how well the reforms are implemented, or differences in the competence of teachers which affect how well the reforms are delivered, to differences in socio-economic background between (groups of) pupils which affect their performance and aspirations. To isolate the impact of the reforms themselves on, say, young people’s examination results or transition into the labour market, as distinct from the impact of any/all of these other variables is a time-consuming and technically demanding task. Furthermore, such evaluations cannot be accomplished unless there is already sufficient ‘baseline’ or ‘time zero’ information, so as to measure any ‘before and after’ changes. A useful paper, ‘Measuring the effects of educational interventions’ (Brooks *et al.*, 1997), is available from NFER.

Given these constraints, there is a range of techniques which can be used to provide measures of impact. The following are summarised, in descending order of rigour, from Holtermann, 1998:

- ♦ *randomised controlled trials*: these offer clear-cut results with firm attribution of impact/effectiveness
- ♦ *before and after studies with control group*: these offer clear-cut results if the control group is well-matched
- ♦ *before and after studies with no control group*: this method offers an uncertain basis for the attribution of impact
- ♦ *longitudinal studies of population samples*: this method must control statistically for all other influences on outcomes before being able to provide an attribution of impact
- ♦ *cross-section studies*: this method relies on assumptions about the transferability of findings from one population to another
- ♦ *opinion and attitude surveys*: these offer an insufficient basis for an assessment of impact/effectiveness, although they are of interest in their own right
- ♦ *studies of services and users (processes and practices)*: insights from these methods are vital in designing workable and acceptable services, but without a control group they do not provide data on impact/effectiveness.

It needs to be acknowledged that – for very good political reasons – most educational initiatives, especially large-scale reforms, are usually not set up and designed in such a way as to permit the possibility of control groups or ‘matched’ populations. This is certainly true of the national curriculum and associated reforms in England and Wales. So evaluations usually need to be carried out by less certain methods.

Most forms of evaluation of educational systems or initiatives fall, in practice, into one or more of the categories listed below. Please note that a critique of these approaches so far as they have been deployed in the UK is to be found in Section 3.

2.3 Reviews of Existing Research

Governments or their agencies may commission an independent identification, analysis and assessment of existing research studies – government-funded and independent; published and unpublished – into the implementation and impact of curriculum reform. Such an exercise is normally undertaken at critical points in time only (such as when revisions to legislation are being considered) in order to provide a broad base of different kinds of information for the government’s monitoring

programme and in particular to update government on issues of concern and areas where further developments or modification are needed.

2.4 Ongoing Monitoring of System's Effectiveness

2.4.1 Ongoing structured and unstructured monitoring by government departments/agencies, including through activities like questionnaire surveys and extensive fieldwork (which may be contracted out to independent organisations), visits to schools, focus group discussions, liaison and cooperation with other official bodies – teacher associations, subject organisations, awarding bodies – and informal analysis of unsolicited correspondence and contacts.

2.4.2 Measurement of performance of students in key skill/knowledge areas, including evidence of improvement or otherwise over time and of relative differences, if any, between different student groups. This may be done by central government/education ministry statistics sections and/or a government agency, or may be contracted out to independent expert institutions. If the systems have been set up prior to the reform, it may be possible provisionally to attribute changes in performance to the impact of reform, provided other relevant variables – such as demographic change – can be taken account of. Measurement of performance can be done by sampling or on a universal basis, depending on purpose (see Section 3).

2.4.3 A variant of the above is found in **international studies of attainment**, like the Third International Mathematics and Science Study (TIMSS) and the International Assessment of Educational Progress (IAEP) in which Slovene children took part (see, for example, Šetinc, 1999); see also Brown (1998) for a critical view of international studies.

2.4.4 Measurement of performance of institutions on key indicators, including evidence of improvement or otherwise over time. Again, this may be done by central government or its agency and/or further contracted out to relevant expert institutions for parts of the data collection and/or analysis. This approach requires a universal system of assessing student performance. If the system can permit the collection of the requisite data, evidence may include 'value added' measurement of relative progress made with given cohorts of students between entry and exit.

2.4.5 Independent inspection of schools on quantitative and qualitative indicators, carried out by government agency and/or independent teams according to explicit criteria and possibly, although not necessarily, accompanied by advice and guidance to schools on how to improve. This may give broad indications of change over time, but it may be hard to ensure precise comparability of the application of inspection criteria in different circumstances/by different teams.

2.5 Specific, Local and/or Contextual Studies

The above forms of monitoring and evaluation are often complemented/supplemented by the following.

2.5.1 **Value for money studies** of particular aspects of the system, carried out by independent auditors; these studies would be done by the national body/bodies with such a remit and/or commissioned by central government from market consultants, etc. In practice, there are more and less precise ways of defining, and assessing, 'value for money'. It is a relatively new area for educational evaluation and the methodology cannot be said to be well developed.

2.5.2 **Programme evaluations** of particular programmes or initiatives, normally carried out by an independent research institution under contract to the central government education ministry. These aim to provide evidence – both quantitative and qualitative – of the impact of specific funded programmes; but this has usually been within their own terms i.e. without measuring the impact of one programme compared with another.

2.5.3 **'Occasional' reports by regulatory bodies**, like the schools inspectorate, on particular themes of current interest – for example, the development and use of information and communications technology in schools.

2.5.4 **Qualitative studies** of the process (often including management of change) and/or perceived outcomes/experiences of selected aspects of reform, in order to explore issues raised by the performance evidence.

2.5.5 **Local inspections**, by appointed municipality/local education authority teams, on agreed performance indicators of particular interest to local communities;

2.5.6 **Independent studies**, i.e. funded otherwise than by central government, undertaken by university teams or other research bodies to investigate aspects of reform from different perspectives.

2.5.7 **Internal, informal or self- evaluation by schools**, perhaps to explore process and experiential indicators.

Examples of these approaches are given in the following section.

3. EVALUATION APPROACHES FOUND TO BE EFFECTIVE

3.1 Evaluation Methods in the UK

This section lists the main evaluation methods recently or currently used in the UK (or rather England and Wales) and, without making a definitive assessment of their effectiveness, attempts to summarise their key strengths and weaknesses as commented on by others (the author is not competent personally to judge some of the technical aspects associated with test development, for example).

Given that ‘evaluation is a political activity’ (Holt, 1981), it needs to be recognised throughout this section that not only the legislation enacted by successive governments but also the creation of the various evaluation bodies, and of the methodological approaches they represent(ed) or sponsor(ed), has typically been premised on or shaped by ideologies of varying shades, strengths and degree of sophistication. Such ideologies have directly competed with or succeeded each other in the course of changing political administrations and intellectual fashions. It is not possible to represent the intricate social and political history of educational evaluation, even over the last decade (which has been typified by contention and conflict), in this document. Many books and articles have been written on this theme, of course, and one or two selected items appear in the References list at the end of the text.

3.1.1 Government Monitoring and Evaluation of the National Curriculum and its Assessment

The agencies responsible for the national curriculum and its assessment (NCC and SEAC, later SCAA, now QCA) have had a remit for ongoing monitoring and review; this has been fulfilled in a variety of ways, including large-scale independent evaluations of the new assessment regimes, questionnaire surveys and interviews with schools as part of the Schools Sampling Project, and a commissioned review of existing research on the implementation and impact of the national curriculum (see Section 3.1.8 below). This monitoring programme has given rise to a great many reports on different topics, several of which have, however, remained unpublished. Evaluations of the implementation of the national curriculum were funded by the then National Curriculum Council (see, for example, NCC, 1991) and also undertaken by Her Majesty’s Inspectorate (see Section 3.1.4 below). Broad-based evaluation strategies were also built into the brief, phased over a number of years, for developing assessment instruments for measuring students’ achievements in core subjects at the

ends of ‘key stages’ of the curriculum (see Section 3.1.3 below). Formal reviews of the curriculum, undertaken through wide consultation exercises, have also been carried out from time to time (see Section 3.1.6 below). The NFER would be willing to make relevant documentation available after discussions with the Slovenian NCECR have sharpened the focus.

Key strengths: the fact that these major reforms were monitored and evaluated from the outset by government is an important principle to have established; a variety of different evaluative approaches were used, although – as noted in Section 2 above – none was able to measure the impact using a ‘before and after’ model.

Potential weaknesses: in retrospect, it has been hard to uncover (i) a coherent view summarising the findings of the various evaluations, not helped by the fact that much evaluative material was not published; (ii) to what extent the results of the evaluations were fed back and systematically used to assist implementation.

3.1.2 Measurement of Student Performance

If what is wanted is an overview of what students achieve nationally in key areas of the curriculum, performance may be surveyed by testing a *sample* of the student population in order to limit both expenditure of resources and unwanted ‘backwash’ effect on the curriculum. This was done in the UK from the mid-’70s to 1990 by the Assessment of Performance Unit (APU), which was set up by the Department of Education and Science to ‘promote the development of assessing and monitoring the achievement of children at school and to seek to identify the incidence of under-achievement’ (Foxman *et al.*, 1991). The APU commissioned research teams in five curriculum areas – language, mathematics, science, foreign languages and design and technology – to carry out the tasks of developing and applying appropriate methods and instruments to the assessment of children’s performance and identifying ‘significant differences related to the circumstances in which children learn’ (Foxman *et al.*). Some forty-three large-scale surveys were mounted by APU in a rolling programme covering schools in England, Wales and Northern Ireland over the period 1978-1988. Assessment methods went beyond the highly formalised procedures for testing student performance previously in use, developing interactive, aural/visual and untimed modes as well as traditional timed, written, memory-based modes.

The Unit was disbanded in 1990, and its function temporarily taken over by the Evaluation and Monitoring Unit (EMU), housed within the School Examination and Assessment Council (SEAC).

Key strengths: ‘Light sampling’ kept burden on schools to a minimum but still enabled extensive coverage of students’ performance at different ages and points in time, a detailed picture of curriculum-related performance in key subject areas, and identification of school and student factors associated with performance. Appropriate sampling procedures and effective working relationships with local education authorities and schools were developed.

Potential weaknesses: Questions of reliability and validity were raised by, for example, Holt, 1981, although he took the extreme view that ‘national monitoring has a meretricious attraction... the APU has no educational value in its own right’. In particular, the work of APU caused an international debate over the so-called Rasch model. This model was designed as a solution to the problem that, because test items may become out-of-date over time as culture and society change, longitudinal studies of performance can be severely compromised; but it was strenuously criticised by Goldstein (1979), among others.

If, on the other hand, what is wanted is a systematic picture of what each and every student achieves and how s/he progresses, then a universal system of testing and assessment is required. In principle, the way this has been done in England and Wales since 1988 is to set out in the national curriculum what students ought to be able to do and then gather evidence of whether they can do it. Arrangements were made for students’ performance in the specified attainment targets of the national curriculum to be assessed and reported on at ages 7, 11, 14 and 16 (i.e. at the end of each ‘key stage’ of the national curriculum). Assessment has been by a combination of externally set tests (standard assessment tasks/test or SATs) and teachers’ assessments (TAs) in the core subjects. The development, piloting and refinement of this regime was a massive undertaking, requiring large investment of funding and other resources.

The range of compromises which need to be managed in developing a national system – for example, between representing ideal standards and reflecting actual practice, between precision and coherence, between comprehensiveness and manageability, or between reliability and congruence with real classroom learning – is formidable; the acreage of reporting and debate about these issues, and the modifications to the regime which ensued, is far too extensive and important to even attempt a summary here. Examples of valuation studies are Sainsbury *et al.*, 1992 and Ashby *et al.*, 1995. Useful overviews of issues include Black (1998) and Whetton (1997). An insightful discussion of some key theoretical areas is presented in Sainsbury and Sizmur (1998). At the same time, the proposals for this system – which were put originally together by the Task Group on Assessment and Testing (GB.DES/WO 1987, 1988) – included provision for using national curriculum assessments for assessing and comparing the performance of *schools*. This was one of the means adopted by government of

making schools publicly accountable for their performance in step with delegating considerable financial and managerial responsibilities to them (see Section 3.1.3 below). It was further argued (or assumed) that parents would make use of results to choose a school and that this exercise of the educational market would help to drive up standards. Assessment of student performance in England and Wales has thus in the most recent period of reform been more or less inextricably conflated with both the accountability and the marketisation agendas – in retrospect, a perhaps unhelpful association.

Results in the statutory tests for individual students are not published, but reported by the school to the parents (and the students themselves); results aggregated to school level are published annually by the DfEE in performance tables. Results aggregated to local education authority (LEA) level are also published by DfEE and used to judge LEA performance. Results aggregated to national level are published by DfEE and used to monitor progress towards national targets.

Key strengths: Because of the confusion or combination of several different agendas, it is hard to isolate the real strengths and weaknesses of the national system of assessment. But it may fairly be claimed that the extensive programme of national test development and evaluation has actively involved many teachers and other professionals in creating a collective understanding of the purposes and significance of assessment, whatever the perceived limitations of the system as it now exists. It has also changed the default: it would be hard to argue convincingly these days that students' achievements and progress should not be assessed and underachievement identified.

Potential weaknesses: As well as questions of validity and reliability pertinent to all assessment regimes, there are other weaknesses associated with the English model. Since the system is universally applied and not done by 'light sampling', and since students' results are also used to scrutinise schools' and, increasingly, teachers' performance, there is a real risk of 'teaching to the test'. This means that performance in tests becomes what, in the reality of most classrooms, drives the curriculum and its teaching instead of the other way round. There has been a proliferation of tests for the non-statutory year groups, as teachers strive to 'predict' future performance from current performance. The ultimate danger, or fear, is that the curriculum becomes impoverished, although this has yet to be conclusively demonstrated. Furthermore, because the assessment system is intended to serve multiple purposes which have turned out to be in mutual conflict, some commentators argue that it serves none of the intended purposes well.

3.1.3 Measurement of Institutional Performance

See Section 3.1.2 above for a summary of how the measurement of institutional performance came to be instituted. This has become a notorious issue at least partly

because schools' results, in the form of ranked performance or 'league tables', have been disseminated with much publicity – 'naming and shaming' – in the local and national press. But see also an interesting article (Downes, 1998), in which the author, a headteacher, argues that 'the opening up of schools to public interest is a welcome and long overdue development'.

The original proposals (made by TGAT) for using student performance to assess schools were criticised early on by Goldstein and Cuttance (1988) on the grounds that this system would do an injustice both to schools as would-be providers and to parents as would-be selectors of educational quality. They argued that if the system made comparisons between schools on the basis of pupils' attainment in circumstances where there was variation, either social and/or academic, in student intakes to schools – and it could be assumed that in the real world there *would* normally be such variation – this would mask the true extent of the progress made by students in different schools, as distinct from the standards reached by them. This argument, since elaborated on in many articles and reports, has been the essential counter to the then government's espousal of 'league tables' of performance based on 'raw' results.

There have been two broad kinds of response to league tables from the education profession. One is to abolish them altogether – although proponents of this view are surely impelled to say what they would substitute to satisfy legitimate requirements for accountability. The other is that 'value added' measures of performance should be developed, i.e. measures which take due statistical account of background variables known to correlate with performance. This is the route latterly adopted by the previous Conservative government and continued by the Labour administration. It is, however, fraught with technical and operational difficulties, especially with regard to constructing indicators which can be readily understood by a lay audience, i.e. the very people for whom performance tables were designed – parents and the public. For a review of developments and issues in 'value added', see Foxman (1997) and Saunders (1999).

This area is one which is still under development by DfEE, QCA and OFSTED; acceptance of the latest system of performance indicators – called 'the Autumn Package' – by school teachers and senior managers varies enormously; and it has yet to be shown that the effort involved in collecting, analysing, reporting and disseminating the data is resulting in school improvement.

Key strengths: Performance tables have opened up schools to public scrutiny as never before, and much of the debate this has occasioned within and outside the profession

has been well-grounded and about fundamental principles as well as technicalities, trivialities and ‘defended territory’. A spin-off has been the greater understanding by school staff of the benefits, uses and limits of statistical data, to an extent not imaginable a decade ago.

Potential weaknesses: There is a risk that the focus on measurement of performance and value added analyses will constitute a ‘displacement activity’ for school staff, at a time when there is a huge educational agenda for them to address in the wake of the Labour government legislation. There is a more fundamental danger, that a great deal of effort is being expended in the policy, research and practitioner arenas for an activity which may end up not being able to reflect to the public the complex reality of schools’ performance on the one hand nor be a sure way of raising standards in all schools on the other.

3.1.4 Independent Inspection of Institutions

Her Majesty’s Inspectorate (HMI) was established over 150 years ago, with the remit of providing government with information and inspection-based advice about the education service. By the late 1970s, as the presumed link between education and the state of the economy became the big idea, central government had a keener interest than previously in standards being achieved in schools, in the relevance of what was taught and learned, in the sound management of schools and LEAs, and in funding, resources and ‘value for money’ – and therefore in the kinds of detailed information inspections could in theory provide (Bolton, 1998). Inspections were carried out by LEAs of the schools within their control but these were not done according to common and rigorous criteria; HMI was thus pivotal to providing detailed up-to-date information on the outcomes and workings of education system.

From 1983, reports of inspections of individual schools were required to be published; and there grew up an expectation that all HMI reports – including those on specific government initiatives, teacher training, youth training and expenditure by LEAs – would be in the public domain. From 1988, annual reports summarising the school inspections were published (see, for example, GB.DES.HMI, 1989, 1990, 1991, 1992). Importantly, for the purposes of this paper, each of these focused to a large extent on issues connected with the implementation of the national curriculum.

As the need for inspection evidence grew in line with the demand for increased accountability in the system, it was inevitable that the structure of the inspectorate would have to be changed to meet this need. Of the four viable options available (Bolton, 1998), the model chosen in 1992 was one in which independent inspection agencies undertake the actual work of inspection, carried out according to established

national criteria and protocols, and quality-controlled by HMI from within a completely independent office, the Office for Standards in Education (OFSTED).

As before, the outputs consist, on the one hand, of individual reports on schools, whose findings must now be responded to by the school governing bodies with a published action plan and timeline; and, on the other, of annual reports by Her Majesty's Chief Inspector (HMCI). The problem now is that advice and guidance are no longer part of the remit for the former, which leaves some schools exposed as having 'serious weaknesses' but without effective support in addressing them; however, the annual reports – which do offer general pronouncements and advice – are not, and cannot be, written by the people who have carried out the inspections on which they are based.

Key strengths: The current system is based on a clear framework of criteria which cover a great deal of the curricular, pastoral and social provision which schools ought to be delivering, and are concerned with efficiency and effectiveness in the use of public money. The notion of publicising both positive and negative aspects of schools (most school reports are covered in the local press) so as to leave 'no hiding place' for poor practice is one which is attractive to many parents.

Potential weaknesses: So much controversy has surrounded OFSTED in its current manifestation – the perceived variability of inspection teams, the questioned accuracy of their findings and the polemical pronouncements and abrasive personality of the current HMCI – that it is quite hard to disentangle these from more underlying weaknesses. Bolton (1998) suggests that in order to manage the variability of teams, the inspection framework has had to be made tightly prescriptive – and thus incapable of being used as an instrument to capture the real diversity and complexity of schools. There is a strong feeling among educationists that the inspectorate remit for advice and guidance needs to be reinstated. There is some question over lines of accountability for HMCI *vis à vis* the DfEE.

3.1.5 Programme Evaluations

Besides the national curriculum and its assessment, many individual policy programmes aimed at reforming provision in other ways have been launched by successive central government administrations from the early 1980s onwards. Many of these have been concerned with making education – especially vocational education and training (VET) – more directly relevant to the needs of the British economy. Examples of these are the Youth Training Scheme, the Technical and Vocational Education Initiative (TVEI) and Compacts. Later programmes, such as the national literacy and numeracy programmes, have been targeted on early years.

Such programmes, not least because they are supported by large-scale central funding and other resources, are expected to make a tangible impact on the achievements of young people, and evaluations are funded by central government with the need to assess their impacts to the fore. Some of the early programme evaluations were instrumental in developing performance indicators by which to judge impact. But because the programmes are often phased in over several years, governments usually want formative input as well as summative evaluations; it is therefore important that good working relationships are established between government and evaluators, who need to understand and respect each other's professional priorities. It is also the case that, within a broad national strategy and guidance framework, responsibility for programme structure, content and delivery rests with local managers, on the grounds of responsiveness to local economic and educational need. Evaluations must take account of different models of implementation and process in order to shed light on potentially differential outcomes, so that the knowledge and skills required of evaluators go beyond those of traditional psychometric research (see Parlett and Hamilton, 1987). Some commentators have argued that since programme evaluators are funded by programme initiators they tend to become progressively identified, by themselves in an unacknowledged way or by others (particularly those being evaluated), with the policy perspective; and that this is to the detriment of the impartiality of the evaluation (see Kushner and MacDonald, 1987).

Key strengths: Programme evaluations are immensely diverse, but as a generalisation it is fair to say that they have contributed to the discussion and development of appropriate local and national performance indicators, both quantitative and qualitative. Many of them also have continued to demonstrate the importance of understanding the management of change; and of the interactive nature of educational innovations, such that concepts like 'added value' and 'multiplier effect' have come to replace or at least supplement the notion of 'impact'.

Potential weaknesses: A common problem is that programme evaluations are bounded by their individual terms of reference, which do not usually include (i) an evaluation of the policy itself, only of its implementation; (ii) any comparative information which could throw light on relative cost-effectiveness as between one programme and another with similar remits.

3.1.6 Formal Consultations and Consultative Reviews

It is fair to say that the national curriculum and associated reforms met, especially in the early years, with resistance and even hostility from many in the educational profession, at least partly because of the speed, and lack of consultation, with which they were introduced as well as the huge demands they made on school managers and

classroom teachers. As was noted above, it was not evident that the evaluation and monitoring programmes taking place were being used to modify the reforms in an open, coherent and consistent way. It became clear to later central government administrations that, to maintain the viability and credibility of the reforms in the public mind, a wide-ranging review of the curriculum and its supporting arrangements was necessary. A series of consultation exercises was undertaken by the then School Curriculum and Assessment Authority, leading to the publication of high-profile, easy-to-read reports accompanied by recommendations: these reviews have come to be known by the name of the Chairman, Sir Ron Dearing (see Dearing, 1993, 1994). Broadly speaking, they resulted in successive rationalisations and simplifications of the curriculum and its assessment. While these have been welcomed, they must be balanced against the need to avoid continual ‘tinkering’ and disruption.

The latest in these reviews was initiated in early 1998 by the Labour government, as a new administration with a commitment to education as a central strand of policy but a somewhat different view of the curriculum from the Conservative government. The intention is to have a revised curriculum in place in all schools from September 2000, but one which will minimise disruption and focus on raising standards. The Secretary of State’s proposals for a revised national curriculum framework encompass not only further simplifications but also greater flexibility and a more explicit values base (see the earlier paper produced by NFER for Slovenije). The body with responsibility for providing advice and practical guidance is the Qualification and Curriculum Authority, and another consultation exercise by QCA – to be contracted out to a research organisation – is about to begin. According to the specification for the consultation exercise, those to be sent the proposals include all schools and local education authorities, professional organisations and faith groups, teacher associations and business and industry. The consultation questionnaire itself will be sent to a sample of schools and available to others on request, and a report on the analysis of the survey will be made.

3.1.7 Independent Studies

There are many bodies, apart from central government departments and agencies, which fund educational research in England and Wales, either regularly or occasionally; and a great many institutions, ranging from university departments of education to private management consultants, which carry out studies within a range of different and conflicting perspectives, from the post-modern to the managerial. It is therefore quite impossible to give an overview of independently funded research

into the curriculum reforms of the 1980s and 1990s, although a provisional starting point is the review noted in 3.1.9 below. An example of one useful wide-ranging empirical study, not cited in that review, is Weston *et al.*, 1992, which investigated the management of the whole curriculum for students from five to 16, and evaluated the requirements of the Education Reform Act in the light of schools' and LEAs' efforts to maintain, recreate or develop coherence in the curriculum.

3.1.8 Institutional Self-Evaluation

As was noted in NFER's first report to the Pedagoški Inštitut, much of the recent work of the Department for Education and Employment in England has focused on developing a systematic approach to self-evaluation by schools as part of the government's raising attainment agenda. The White Paper *Excellence in Schools* and the related documents on school improvement, target-setting and benchmarking have together provided guidance for a five-stage cycle of review and target-setting by schools. Clearly, self-evaluation is an important aspect of quality assurance in education, although – as the NFER's earlier paper suggested – there are a number of issues which have yet to be resolved about its context and conduct.

New developments to assist LEAs with supporting systematic school self-evaluation are being initiated by the Audit Commission and OFSTED as well as the DfEE.

Key strengths: Capable of reaching the parts of institutional life and practice not reached by most other forms of evaluation; may, if handled appropriately, engage staff and students in 'creative' and professionally/personally enriching ways of understanding themselves and their role.

Potential weaknesses: May be perceived as a 'soft option' by some policy-makers, who would rather see self-evaluation as, ideally, a form of internalised inspection. It is also hard to make an overall assessment of the impact of self-evaluation on standards, although the DfEE commissioned a large-scale qualitative study from NFER in 1997-98 with this in mind.

3.1.9 Comprehensive Reviews of Existing Research

The only example of a comprehensive review – i.e. intended to cover all government-funded and credible independent research, whether published or unpublished – known to the author in the field of national curriculum evaluation is a report commissioned by the then School Curriculum and Assessment Authority from an English university department of education, but subsequently not published (Ford *et al.*, 1998). This fact makes the report's status dubious and the material excerpted from it which follows

should be treated as confidential. A search of the list of references to this report suggests that the review is not in fact comprehensive. However, because of its synoptic approach it is a unique document with key points of relevance to Slovenije. These can be highlighted as follows:

Educational purposes

- ♦ ‘The concept of citizenship, defined as preparing children for life as active members of a democratic community, emerges consistently in the literature with a view that this should be the pivotal principle in any revised curriculum.’¹

Curriculum structure

- ♦ ‘Systems for curriculum evaluation in some countries take into account a broader range of curriculum objectives, and depend less heavily on universal testing... than in England and Wales. Several countries employ sampling methods similar to the now defunct Assessment of Performance Unit’ [see above].
- ♦ ‘There has not been enough research into the impact of the curriculum, and the different ways of organising it, on pupils’ learning and achievement.’
- ♦ ‘The increasing knowledge about pupils’ cognitive development has not yet informed the structure of the curriculum or of subjects. More work is needed in this area.’
- ♦ ‘There is a tension in the literature between the concept of “flexibility” or local determination, and central prescription, with concern that flexibility, unless regulated, may lead to even less balance in the curriculum than currently exists.’

Cross curricular issues

- ♦ ‘There is a growing interest in defining learning outcomes in relation to cross curricular issues [such as society studies, citizenship education and personal and social development], and of linking these to an effective system of assessment so that this area of the curriculum might gain status.’

Lifelong learning

- ♦ ‘“Lifelong learning” and the “learning society” are highly contested concepts and a number of different versions of each are currently competing for attention, with some emphasising economic competitiveness, others social cohesion and still others learning to learn, or a combination of all three.’
- ♦ ‘There is overwhelming support in the literature that more emphasis needs to be placed on understanding how learning takes place, and on how the knowledge, skills and attitudes required for developing as lifelong learners are acquired. Pedagogy remains a seriously under-researched topic.’

Standards

- ♦ ‘There are many outstanding problems with the [perceived] validity and reliability of tests so that they have still to prove their worth as an appropriate measure of standards.’

¹ This recommendation forms part of the Secretary of State’s proposals for a revised curriculum framework, as noted in NFER’s previous report and above.

- ◆ ‘It is not possible to state unequivocally that standards of performance have risen or fallen as a result of national curriculum reforms, and the debate about standards, both internationally and in England and Wales, remains an area of disagreement.’

Assessment

- ◆ ‘There is evidence of improvement in classroom practice in Key Stage 1 [Years 1 to 3] prompted by assessment requirements.’
- ◆ ‘There is some evidence that for low achieving pupils in particular self-esteem and disposition to learn are being damaged by some assessment practices.’
- ◆ ‘Statutory assessment and its associated reporting arrangements are generally accepted by parents. However this does not entirely satisfy their desire for more detailed information about their children’s learning and development.’

Managing implementation

- ◆ ‘Insufficient attention has been given to the process of turning statutory guidance into curriculum content as delivered by teachers and experienced by pupils. This has led, in places, to an over-reliance on textbooks and published schemes, an emphasis on delivering content and insufficient attention being given to the process of teaching and learning.’
- ◆ ‘An incalculable amount of teacher time and energy has been dissipated by the frequent changes to the national curriculum. This has eroded their goodwill and has implications for future developments.’

Classroom processes

- ◆ ‘Teachers have difficulty in implementing problem-solving, enquiry and investigative aspects of the curriculum. This appears to be due partly to limitations in teachers’ conceptual understanding of subjects, to the dominance of certain teaching styles and to a limited range of classroom interactions which they find it difficult to relinquish.’
- ◆ ‘Important parts of the national curriculum Orders are not being implemented because insufficient attention has been given to supporting teachers in changing their classroom practices, especially where these are concerned with helping pupils to better understand their own learning.’

Teacher development

- ◆ ‘Many studies show that teachers’ beliefs about, and images of, teaching are powerful mediating factors in the implementation of the national curriculum. The research suggests that insufficient attention has been given to changing or to utilising teachers’ beliefs about teaching and learning, through small-scale action research and reflection.’
- ◆ ‘Teacher development as a key factor in effective change has been largely ignored, and curriculum change has been treated as though it were a technical issue rather than a human one.’

The NFER has not been able to ascertain the reasons for non-publication; but even if the findings are felt not to be a full and accurate view of the situation in England and

Wales the points made above deserve to be taken seriously as ‘early warnings’ for other systems.

3.2 New Generic Developments in Evaluation in the UK

As a result of new pressures on all government departments to account more closely for public sector spending, research and evaluation in education are now being challenged to meet more stringent policy requirements in respect of value for money and to provide better, more systematic and synthesised evidence of ‘what works’.

3.2.1 Assessment of Value for Money

Discussion about the ‘economics of education’ is usually concerned with the costs of education, understood broadly as a percentage of gross national product, and with the outcomes of education, understood broadly as skills, personal wealth and productivity. But in fact ‘value for money’ in education is not a self-evident principle or criterion: what sort of value and for which stakeholder groups, is a question which springs immediately to mind.

Sometimes the terms ‘cost effectiveness’, ‘value for money’, ‘best value’, ‘cost-benefit analysis’, ‘rate of return analysis’, ‘economy, efficiency and effectiveness’ are used interchangeably despite the fact that they mean different things. Holtermann (1998) makes a helpful distinction between cost effectiveness and cost benefit analysis: ‘cost effectiveness studies usually measure outcomes in quantitative terms, but do not always measure all outcomes and do not attempt to place monetary value on outcomes that are not already in monetary terms’. Cost benefit analysis, by contrast, ‘measures all benefits and costs, and places monetary values on all the benefits as well as all the costs’. She goes on, ‘There are, however, few occasions where it is possible and acceptable to measure all the benefits in monetary terms, and so a cost benefit analysis in this formal sense is something of a rarity.’

Holtermann says that ‘cost benefit analysis aims to answer the question “what works best in relation to cost?” and it cannot be done unless the question “what works best?” is first answered.’ This issue is very much linked with the current Government concern to establish a sounder evidence-base for policy (see Section 3.2.2 below) and has far-reaching consequences for the design of initiatives and their evaluations.

In England and Wales, the Audit Commission was set up in 1983 with a remit of ‘assessing expenditure, not just for probity and regularity, but for value for money as well’ in local (education) authorities. The Commission published a series of reports on schools’ management of finances earlier this decade, *Adding Up the Sums* (1993a and b, 1994, 1996). As well as assessing how well schools were managing their delegated budgets, these reports provided comparative patterns of expenditure and resource use to enable heads and governors to compare the way their own school deployed resources with that of other schools.

Three general points emerge from this work: (i) the need for trends over time to be established in order for ‘value for money’ to be assessed; (ii) the need for rigorous statistical analysis to determine whether trends are significant; (iii) the need to ensure that comparisons are made on a ‘like with like’ basis and thus to exclude from the data any ‘external events’ which would make comparisons invalid.

Cost-effectiveness is an under-developed area of educational evaluation in the UK. Accordingly, the DfEE plans to fund a dedicated research centre for three years to develop a programme of interdisciplinary research that will address questions such as:

- ◆ how can educational interventions be linked to pupils’ skills?
- ◆ what effect does investment in education at different ages/phases have on long-term outcomes?
- ◆ how do policy options compare in education?

Part of the work of this centre will focus on developing the methodology for conducting cost-benefit analysis of educational interventions.

3.2.2 Evidence-based Educational Policy

As noted above, the agenda driving the ‘evidence-based’ education movement is closely connected to establishing value for money, in that the latter must be premised on having a systematic knowledge of ‘what works’, based on the best possible research evidence. Criticisms have recently been made about the lack of utility of much educational research, its unsystematic repetition (as distinct from replication) of other previous work or the simple failure to take account of other work, its sometimes dubious methodologies, unexplained rationales, unsupported findings, and above all its lack of relevance to either policy or classroom practice (for an analytical review of

research in education see Hillage *et al.*, 1998; for a more polemical view, see Hargreaves, 1998).

An international movement for evidence-based education, modelled to a large extent on principles taken from healthcare, is gaining rapid credence with researchers and policy-makers alike: for an overview of issues and approaches, see, for example, Davies (forthcoming), Harlen (1996), Sebba (1999), Slavin (1986). The main areas which need to be addressed are:

- the methodology for conducting comprehensive and synthesising research reviews;
- the undertaking of new research reviews in key areas;
- the quality assurance and control of empirical (including qualitative) research;
- the transparent analysis and reporting of research studies;
- the dissemination of research to policy-makers and practitioners.

It is early days to know exactly what the practical outcomes of the movement will be. The government's proposals include setting up 'an education and employment collaborative research resource' whose purposes would be to:

- ♦ 'increase the certainty with which evidence in a given area can be regarded, by generating relevant, reliable, up-to-date evidence which minimises bias;
- ♦ coordinate and support the collation of educational... research;
- ♦ provide evidence that is accessible to inform research, policy and practice, thereby contributing to the development of an overall dissemination strategy.'

The proposed resource would achieve these objectives by:

- ♦ 'developing a comprehensive database of ongoing and published research;
- ♦ setting up collaborative research groups to undertake reviews in specific areas and take responsibility for updating these. Membership should include researchers, policy-makers, practitioners and others who share an interest in accessing reliable evidence' [these are colloquially known as Cochrane Reviews, after the model established in healthcare].

If this proposal, or anything like it, becomes a reality, it will represent a major investment in research but would also bring a number of challenges to research institutions and individuals. These include agreeing the overarching methodological issues in reviewing educational research, framing 'Cochrane review' questions,

outlining feasible models and following through what may be far-reaching implications for educational research practice (including training and resources).

3.3 Other Issues to be Considered

3.3.1 Arising from the above discussion are three further issues which will need to be explored in the conduct of evaluation(s) by the NCECR in Slovenije. The first is the relationship between central government and independent research/monitoring institutions as it relates to the quality of research management by central government, especially the funding of strategic research programmes, the specification of research tenders and the procurement/purchasing of research contracts. This is too complex an area to discuss in this paper, but the NFER would be willing and able – as a separate exercise – to provide further detail for the NCECR, given the relative maturity of practice in the Department for Education and Employment (England) in this respect.

3.3.2 The second is the development of appropriate instruments, ranging from tests and other assessment of performance tools, through questionnaires, interview and observation schedules, to instruments for open-ended ethnographic studies. Again, this is far too diverse and technical an area to address here, and it is assumed that the NCECR will have access, through the Pedagoški Inštitut among others, to the requisite expertise. However, the NFER could offer technical guidance if required.

3.3.3 The third issue is qualitatively different from the previous two but of great importance for the development of a rational system of evaluation. It concerns the proliferation of national agencies concerned with evaluation, monitoring, testing, inspection and so forth; as Holt (1981) neatly puts it:

It is a characteristic of pluralist societies – and of educational planning in particular – that it is relatively easy to allow new organisations to come into existence, and almost impossible to get rid of them. As soon as they become institutionalised, their adherents form a political power base and what might at first have come about largely by chance suddenly appears to fulfil an irresistible purpose. (Holt, 1981)

In England, some rationalisation of the various quasi-autonomous non-government organisations (QUANGOs) in the field of education has taken place in recent years, so that the National Curriculum Council and the School Examination and Assessment Council, both created in 1988 by the Education Reform Act, merged in 1993 to become the School Curriculum and Assessment Authority; this body was itself merged in 1997 with the National Council for Vocational Qualifications, established

in 1986, to become the Qualifications and Curriculum Authority (QCA). The Department for Education and Employment (DfEE) started life after the second world war as the Board of Education, later the Department of Education and Science; it became the Department for Education in 1992, and then merged with the Employment Department (which had its own chequered history in the context of post-war 'manpower' needs and the provision of training) to become the DfEE in 1995.

Some further proliferation has also occurred, however. Within the DfEE is a semi-autonomous unit created in 1997, the Standards and Effectiveness Unit (SEU). The Unit is responsible for key areas brought into being by the new Labour administration through its White Paper, *Excellence in Schools*; these include school and local education authority effectiveness and the national literacy and numeracy programmes, as well as several targeted interventions such as Education Action Zones.

Her Majesty's Inspectorate (HMI) is a model of longevity by comparison, being created in 1839 to inspect educational provision in England, although the Chief Inspector was required to produce annual reports summarising the findings from its programmes of individual school and college inspections only as recently as 1988. During the 1980s, HMI became more closely associated with informing government policy, noticeably through unpopular or contested actions taken by government in the wake of critical HMI reports. On the other hand, consequent on the Education Reform Act of 1988, much of what HMI reported was in effect, though not necessarily in intention, critical of aspects of the reforms or at least their implementation and unintended consequences (Bolton, 1998). Its independence was perceived by both policy-makers and practitioners to be compromised, though for opposite reasons. HMI was replaced in 1992 by the Office for Standards in Education (OFSTED) which was created to give some 'teeth' to the government's accountability agenda by providing a framework for an enlarged and much-publicised programme of inspections – each school to be fully inspected once every four years – and training for the private teams of inspectors who needed to be contracted to carry out this programme. See Bolton (1998) for a view of these recent developments.

(The field of post-compulsory education and training, i.e. further and higher education and work-based training, have histories which are, if anything, even more convoluted, particularly with reference to funding and accountability mechanisms. It is not proposed to attempt an overview here! Useful references include Young and Spours, 1998, and Trow, 1998.)

Sometimes the effect of having more than one body with statutory interests in the same areas leads to ‘turf wars’, in which people seek to defend their organisational vested interests, and also to ‘mixed messages’ being given to schools and the public at large. A symptom of the extent of rationalisation that has still to be accomplished, even within a relatively contained area like the publication of performance data, is that it is still possible for QCA, the Office for Standards in Education (OFSTED) and the Department for Education and Employment (DfEE) to produce statistics on school and student performance which purport to help schools and local education authorities with monitoring and self-evaluation but which are calculated on different bases, take account of different variables and give different national norms.

It may be therefore be helpful to bear in mind one of the OECD recommendations, namely:

- ♦ The structure of institutions involved in the reform of secondary education seems broadly right for the current stage of reform, and in particular the central role of the National Curriculum Council is important. The Ministry of Education and Science should, however, keep the institutional structure under review, with the aim of simplifying it when appropriate.

4. SOME WORKED EXAMPLES OF ‘SUCCESS CRITERIA’

As was said in the introduction, the NFER considers it important that the key stakeholders should be involved in the definition and measuring of success criteria. At the same time we concur with one of the general recommendations made by OECD (Reid *et al.*, 1998), as follows:

Current developments should aim at a transparent and outcome based system in which the outputs of all parts of the education system are subjected to regular measurement. There should be a hierarchy of measures, at national, regional and local and institutional level, which would contribute to a general appreciation of the effectiveness of the system and its value for money.

The framework produced in the NFER’s first report was designed with this requirement in mind. To help with the development of success criteria, we have provided one or two worked examples – see overleaf.

[To follow as soon as possible]

REFERENCES

ASHBY, J., BURLEY, J., HARGREAVES, E. and McCULLOCH, K. with SAINSBURY, M. and SCHAGEN, I. (1995). National Curriculum Assessment 1995, Key Stage 1: Evaluation in the Core Subjects. Unpublished report. Slough: NFER.

AUDIT COMMISSION (1993a). *Adding up the Sums: Schools’ Management of their Finances*. London: HMSO.

AUDIT COMMISSION (1993b). *Adding up the Sums 2: Comparative Information for Schools*. London: HMSO.

AUDIT COMMISSION (1994). *Adding up the Sums 3: Comparative Information for Schools – 1994*. London: HMSO.

AUDIT COMMISSION (1996). *Adding up the Sums 4: Comparative Information for Schools – 1995/96*. London: HMSO.

BLACK, P. (1998). ‘Learning, league tables and national assessment: opportunity lost or hope deferred?’, *Oxford Review of Education*, **24**, 1, 57-68.

BOLTON, E. (1998). ‘HMI – the Thatcher years’, *Oxford Review of Education*, **24**, 1, 45-55.

BROOKS, G., HAGUES, N., HUTCHISON, D., MILNE, A., MINNIS, M., MORRIS, M., SAINSBURY, M., SCHAGEN, I., SCHAGEN, S., SIZMUR, S. and WHETTON, C. (1997). 'Measuring the effects of educational interventions'. Paper presented at the British Educational Research Association annual conference, University of York, 11-13 September.

BROWN, M. (1998). 'The tyranny of the international horse-race'. In: SLEE, R. and WEINER, G., with TOMLINSON, S. (Eds) *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements*. London: Falmer Press.

DAVIES, P. (forthcoming). 'Evidence-based healthcare/education', *British Journal of Educational Studies*.

DEARING, R. (1993). *The National Curriculum and its Assessment: Interim Report*. London: SCAA.

DEARING, R. (1994). *The Review of the National Curriculum: a Report on the 1994 Consultation..* London: SCAA.

DOWNES, P. (1998). 'The head's perspective', *Oxford Review of Education*, **24**, 1, 25-33.

FORD, K., CLARK, J., LEAT, D. and MILLER, J. (1998). An Analysis of Research into the Impact of the National Curriculum and the Implications for Teachers and Schools. Unpublished report. Newcastle: University of Newcastle upon Tyne.

FOXMAN, D. (1997). *Educational League Tables: For Promotion or Relegation?* London: Association of Teachers and Lecturers.

FOXMAN, D., HUTCHISON, D. and BLOOMFIELD, B. (1991). *The APU Experience 1977 – 1990*. London: School Examination and Assessment Council.

GOLDSTEIN, H. (1979). 'Consequences of using the Rasch model for educational assessment', *British Journal of Educational Research*, **5**, 211-20.

GREAT BRITAIN. DEPARTMENT OF EDUCATION AND SCIENCE. HER MAJESTY'S INSPECTORATE. (1989). *The Implementation of the National Curriculum in Primary Schools. A Report by HM Inspectors, Summer 1989 (HMI Report 332/89)*. London: DES.

GREAT BRITAIN. DEPARTMENT OF EDUCATION AND SCIENCE. HER MAJESTY'S INSPECTORATE. (1990). *Standards in Education 1988-89. The Annual Report of HM Senior Chief Inspector of Schools*. London: DES.

GREAT BRITAIN. DEPARTMENT OF EDUCATION AND SCIENCE. HER MAJESTY'S INSPECTORATE. (1991). *Standards in Education 1989-90. The Annual Report of HM Senior Chief Inspector of Schools*. London: DES.

GREAT BRITAIN. DEPARTMENT OF EDUCATION AND SCIENCE. HER MAJESTY'S INSPECTORATE. (1992). *Education in England 1990-91. The Annual Report of HM Senior Chief Inspector of Schools*. London: DES.

GREAT BRITAIN. DEPARTMENT OF EDUCATION AND SCIENCE AND WELSH OFFICE. (1987). *Task Group on Assessment and Testing: a Report*. London: DES and Welsh Office

GREAT BRITAIN. DEPARTMENT OF EDUCATION AND SCIENCE AND WELSH OFFICE. (1988). *Task Group on Assessment and Testing: Three Supplementary Reports*. London: DES and Welsh Office

HARGREAVES, D. (1998). *Creative Professionalism: the Role of Teachers in the Knowledge Society*. London: Demos.

HARLEN, W. (1996). 'Educational research and educational reform.' In: HEGARTY, S. (Ed) *The Role of Research in Mature Education Systems*. Slough: NFER.

HILLAGE, J., PEARSON, R., ANDERSON, A. and TAMKIN, P. (1998). *Excellence in Research on Schools*. Research Report RR74. London: DfEE.

HOLT, M. (1981). *Evaluating the Evaluators*. London: Hodder and Stoughton.

HOLTERMANN, S. (1998). *Weighing It Up: Applying Economic Evaluations to Social Welfare Programmes*. London: Joseph Rowntree Foundation.

KUSHNER, S. and MacDONALD, B. (1987). 'The limitations of programme evaluation'. In: MURPHY, R. and TORRANCE, H. (Eds) *Evaluating Education: Issues and Methods*. London: Harper and Row.

PARLETT, M. and HAMILTON, D. (1987). 'Evaluation as illumination: a new approach to the study of innovatory programmes'. In: MURPHY, R. and TORRANCE, H. (Eds) *Evaluating Education: Issues and Methods*. London: Harper and Row.

REID, G. *et al.* (1998). *Reviews of National Policies for Education: Slovenia. Examiners' Report*. Paris: OECD.

SAINSBURY, M. and SIZMUR, S. (1998). 'Level descriptions in the national curriculum: what kind of criterion referencing is this?', *Oxford Review of Education*, **24**, 2, 181-93.

SAINSBURY, M., WHETTON, C., ASHBY, J., SCHAGEN, I. and SIZMUR, S. (1992). *National Curriculum Assessment at Key Stage 1: 1992 Evaluation*. London: SEAC.

SAUNDERS, L. (1999). *Value Added Measurement of School Effectiveness: a Critical Review*. Slough: NFER.

SEBBA, J. (1999). 'Educational research: developing and implementing the Government's action plan', *Research Intelligence*, **67**, 19-20.

ŠETINC, M. (1999). 'International comparative studies of science achievement from the time perspective: the Third International Mathematics and Science Study versus the International Assessment of Educational Progress', *Assessment in Education*, **6**, 1, 43-56.

SLAVIN, R.E. (1986). 'Best-evidence synthesis: an alternative to meta-analytic and traditional reviews', *Educational Researcher*, **15**, 9, 5-11.

TROW, M. (1998). 'American perspectives on British higher education under Thatcher and Major', *Oxford Review of Education*, **24**, 1, 111-29.

WESTON, P. and BARRETT, E. with JAMISON, J. (1992). *The Quest for Coherence: Managing the Whole Curriculum 5 – 16*. Slough: NFER.

WHETTON, C. (1997). 'The psychometric enterprise or the assessment aspiration'. In: HEGARTY, S. (Ed) *The Role of Research in Mature Education Systems: Proceedings of the NFER International Jubilee Conference, Oakley Court, Windsor, 2-4 December 1996*. Slough: NFER.

YOUNG, M. and SPOURS, K. (1998). '14-19 education: legacy, opportunities and challenges', *Oxford Review of Education*, **24**, 1, 83-97.

**APPENDIX A. RECOMMENDATIONS MADE IN O.E. C.D.
EXAMINERS' REPORT ON SLOVENIA**

The following are the recommendations sections excerpted from the main text.

Chapter 2. Curriculum Development

- ◆ The curriculum content in primary and secondary education should be reduced. The curriculum should be based on the knowledge, skills and values that are relevant to Slovenian society and all the educational stakeholders, not least the pupils, not just those which are important from the traditional academic point of view.
- ◆ Curriculum development should build bridges between the different school subjects, particularly at primary level, with more inter-disciplinary and cross-curricular topics or themes in the Framework Curriculum in order to encourage schools to introduce such courses and activities.
- ◆ It is recommended that a national strategy should provide the municipalities and schools with ideas and approaches to combat the traditional teacher-centred methods of instruction and training, taking account of the understanding gained in the international community of school improvement.
- ◆ The National Curriculum Council has adopted the right approach by inviting all teachers in every school to discuss and comment on the Framework Curriculum documents and this should continue. It is also important to invite the Teachers' Union to participate to discuss the design of new curriculum with the authorities in charge.
- ◆ Educational research should be an important provider of innovation in the system of education and should examine the effects of the reform process in schools and in the system. The Ministry of Education and Sport (MoES) and the Ministry of Science and Technology should create a clearer strategy for educational research in Slovenia which supports and backs up the national reform efforts in education, involving the Educational Research Institute and the University of Ljubljana.
- ◆ The ongoing curriculum reform taking place in Slovenia could be enhanced by having a shared understanding of what is meant by curriculum. This requires, among other things, the involvement of educational researchers, in-service training for head teachers, and emphasising teachers' role as curriculum planners.
- ◆ In vocational education, development of curricula must be closely linked with labour market needs, and employers must be fully involved in the process of development with, where necessary, assistance to support employer networks.

**Chapter 3. Outcome Standards, Assessment, Certification and Quality:
I. Primary Education**

- ◆ Outcome-oriented standards are required in primary education, including for cross-curricular skills, and the Curriculum Commissions should work towards documentation on this basis. These outcome standards should focus on core skills and competences rather than on subject-specific knowledge.
- ◆ Assessment instruments should be developed centrally which enable teachers to check on a regular basis whether the results of their teaching meet required standards.
- ◆ Consideration should be given as to whether the mechanism for selecting students for more prestigious schools should give so much weight to primary education end-tests.
- ◆ Programmes for teacher training should be developed and implemented to help teachers supplement traditional testing with alternative forms, setting standards, and apply these standards reliably.
- ◆ The Ministry of Education and Sport should work towards instruments which provide an assessment of national educational performance.
- ◆ The process of education should be evaluated and the role of an inspectorate independent of the Ministry should be considered for the future.

**Chapter 4. Outcome Standards, Assessment, Certification and Quality:
II. Secondary and Post-Secondary Education**

- ◆ As with primary education, outcome-oriented standards are required for secondary education, and work towards this should be given high priority.
- ◆ Arrangements whereby pupils are able to move between streams in secondary education, currently being changed, should be made more flexible.
- ◆ While safeguarding the positive aspects of the *Matura*, the National Examination Centre should be asked to develop specialised assessment instruments, other than the existing examination structure, for use in the future.
- ◆ The structure of institutions involved in the reform of secondary education seems broadly right for the current stage of reform, and in particular the central role of the National Curriculum Council is important. The Ministry of Education and Sport should, however, keep the institutional structure under review, with the aim of simplifying it when appropriate.
- ◆ The development of vocational standards is still at an early stage, but the full development of a competence-based system of standards should be given high priority. It must have a clear relationship to labour market needs and should create a hierarchy or family of standards which industries and occupations can use.
- ◆ The effect of the *Matura* examinations on teaching and learning should be carefully researched and evaluated.
- ◆ Consideration should be given to how the accreditation of university programmes might be accelerated and deepened, in the interests of quality assurance.

Chapter 5. Regionalisation and Decentralisation

- ♦ Decentralisation within the education system is an important goal which should be pursued, so as to involve stakeholders at regional and local level more closely with decision-making.
- ♦ The Ministry of Education and Sport should consider whether the current method of funding schools provides the right signals within a decentralised system.
- ♦ Decentralisation is particularly important for vocational education. There appears to be no clear focus in this area at regional level and the Government should consider the proposal for 'workforce development boards' at regional level, with either co-ordinating responsibilities or resource allocation powers.
- ♦ The Ministry should consider carefully the educational case for a third university in Slovenia, the implications for public expenditure on education and the opportunity cost for other parts of the system, as against the expansion of post secondary vocational and professional institutions.

Chapter 6. Efficiency and Resources in the System

- ♦ The Government should consider whether cross-departmental machinery should be established to secure a consistent approach across all departments to labour force development.
- ♦ Existing work on labour market information, and vocational guidance, should be reinforced so that enterprises, individuals and Government can take decisions on the basis of soundly based and timely information.
- ♦ The Government should review the effect of demographic trends in Slovenia on the future shape and structure of the education system, in the long term and more immediately in particular, the effect on primary schools (especially in remote areas) and the policy and resource implications should be brought out clearly as a matter for Governmental consideration.
- ♦ The Ministry of Education and Sport should consider whether the recent legislation on higher education is adequate for the universities to adjust their governance to the new situation facing Slovenia.
- ♦ Current developments should aim at a transparent and outcome based system in which the outputs of all parts of the education system are subjected to regular measurement. There should be a hierarchy of measures, at national, regional and local and institutional level, which would contribute to a general appreciation of the effectiveness of the system and its value for money.

CONTENTS

	page
1. INTRODUCTION.....	i
2. ALTERNATIVE APPROACHES TO EVALUATION	iii
2.1 Formal Consultations and Consultative Reviews	iii
2.2 Measurement of the Impact of Reforms	iii
2.3 Reviews of Existing Research.....	iv
2.4 Ongoing Monitoring of System's Effectiveness	v
2.5 Specific, Local and/or Contextual Studies	v
3. EVALUATION APPROACHES FOUND TO BE EFFECTIVE.....	vii
3.1 Evaluation Methods in the UK	vii
3.2 New Generic Developments in Evaluation in the UK	xix
3.3 Other Issues to be Considered.....	xxii
4. SOME WORKED EXAMPLES OF 'SUCCESS CRITERIA'.....	xxv
REFERENCES	xxv
 APPENDICES:	
A. RECOMMENDATIONS MADE IN O.E.C.D. EXAMINERS' REPORT ON SLOVENIA.....	xxix