

Evaluating Educational Outcomes: What Are Our Priorities ?

Professor Patricia BROADFOOT
Centre for Assessment Studies
School of Education, University of Bristol, UK

Plenary Session

THE CHANGING CONTEXT OF EDUCATION

Changing work practices, technological advances and a demographic decline in the absolute numbers of young people in many advanced societies are highlighting an increasingly urgent need for students to be equipped with the skills and qualities that will enable them to adapt to a variety of occupational roles. Problem-solving ability, personal affectiveness, thinking skills and willingness to accept change are typical of the general competencies straddling cognitive and affective domains that are now being sought in young people. Although aspects of many such general competencies are embodied in more traditional subject syllabuses the tension between pressures for curriculum development to reflect the acknowledged need to provide students with learning opportunities in which they can develop these general competencies and the inability of the assessment industry, as yet, to give widespread support to such developments is becoming increasingly marked (Frederiksen and Collins, 1989). Yet as Nickerson (1989) points out, until such assessment procedures are available, not only will it be impossible to produce valid judgements about the success of the educational enterprise as a whole in terms of its goals, it will also help to guarantee that certain desired educational outcomes will be neglected in the classroom.

Historically, the practice of educational assessment has been largely driven by a perceived need to *measure* individual capacity. Associated with this has been the desire to measure individual differences as the basis for fair and appropriate selection to different opportunities and roles within society. Thus the psychometric tradition of assessment has predominated and with it, an overwhelming emphasis on the need for *reliability* in test application such that the assessment may be seen to operate fairly and consistently in the crucial business of determining life chances. The question of *validity* - whether the test does indeed measure what is intended to measure, has arguably been subordinated to the overwhelming need for comparability of results. This preoccupation with reliability has necessarily tended to lead to a concentration upon that which is more easily measurable such as knowledge and understanding and a relative, if not absolute, neglect of higher level intellectual skills such as "thinking" and of those, more affective qualities which are now recognized as a crucial part of both learning and future occupational success (IPM, 1984) but which are much more difficult, if not impossible, to measure using traditional psychometric techniques. Whilst higher-order skills and personal qualities have been the subject of a

considerable *informal* assessment industry of reports and references which testify to their perceived importance, much of this assessment has tended to be intuitive, impressionistic and explicitly subjective -and accepted as such. More systematic attempts to incorporate assessment of such learning outcomes have typically foundered because of problems both of conceptualisation and of the lack of acceptable assessment techniques (APU 1984, Shepard, 1979).

However, dissatisfaction with traditional testing approaches is now widespread, explicit and clearly articulated. It includes the emphasis on comparison between students rather than describing specific and changing levels of attainment; the frequent mismatch between curriculum and test content, the pressure to test in a relatively limited number of aspects of a programme of instructions; the assumption that students learn in a linear fashion and that therefore they must learn and be assessed on the 'basics' before going on to more complex intellectual processes; and that the criteria of differentiation and reliability are the most important, leading to an emphasis on quick, economic, multiple-choice achievement tests which can be shown to have acceptable psychometric properties and are easy to mark.

A second important fundamental reason for increasing dissatisfaction is that the tests can be shown not even necessarily to do what they aspire to do -to provide objective, reliable evidence of the attainments measured (Ingenkamp, 1977, Lee Smith, 1990, Raven, 1989).

A third related reason for dissatisfaction with traditional testing approach concerns what is perceived to be their undesirable 'wash-back' effect on instruction. It is also increasingly being recognised that the gulf between what is covered by formal assessment techniques and those desired learning outcomes that are not the subject of such techniques has major repercussions in the emphasis given in what is taught.

Thus a number of theorists are now arguing for a re-interpretation of construct validity which will take explicit account of this 'wash-back' effect whether it is intentional or unintentional. Starting from the premise that test validity must encompass the actual and potential consequences of test score interpretation and use, Messick, argues that a test should be regarded as 'systematically' valid when it induces in the education system curricular and instructional changes that foster the development of the (cognitive) skills that the test is designed to measure (Messick, 1989).

A number of other studies in the field of meta-cognition (see, for example Howie 1988, Crooks 1989) have revealed just how important

assessment is in defining the attitude students take towards their work, their sense of ownership and control, the strategies they employ in learning and their confidence and self-esteem, all of which impact profoundly on the quality of learning achieved. Thus the growing recognition of the limitation of traditional approaches to assessment, especially the need for a greater concern with validity; concern about the influential but potentially quixotic role of 'personal' assessment; a desire to harness the powerful impact of assessment to promote, rather than inhibit, learning, have fuelled the search for new assessment approaches. Above all, increasingly explicit demands by modern economies both for the encouragement of new kinds of learning outcome and for information about a much broader range of skills and qualities have combined to create a climate in which both the potential outcomes of learning, and whether or not these have been realised, are the subject of new definitions and new assessment techniques.

Thus, most current assessment techniques may be shown to have major shortcomings in a number of important respects. Firstly, their underlying theory of learning often does not reflect important new developments in cognitive psychology; secondly their content does not help to encourage the learning of all the competencies identified as goals for the education system. Thirdly the assessment procedures used may adversely affect students' approach to their learning and the classroom climate as a whole. Fourthly, the effects of such testing approaches encourage either as the basis for judging the true achievements of individuals or as a basis for accountability in judging the achievement of the education system as a whole.

There is thus a fundamental flaw in much contemporary assessment thinking which helps to explain the inability of most existing modes of assessment to affect performance. This, is the assumption the schools can be made better if they are made to try harder. Rather, as Baker et al (1990) argue we need assessment approaches which measure significant learning in a way that supports desired performance and which provide reliable information about incomes. In particular the assessment must address more generic cognitive learning tasks such as deep understanding and problem-solving as these are expressed in different learning tasks and subject demands. These are today's priorities for assessment; these are the basis for an international reform and certification designed to address these needs shows that the process has already begun as I and I shall illustrate in the next section.

SOME CURRENT EXAMPLES OF CHANGE

In reviewing current international development in assessment, some of the characteristics which international comparisons suggest may well prove to be the harbingers of a new assessment culture are as follows:

- 1 An increasing emphasis on formative, learning-integrated assessment throughout the process of education.
- 2 A commitment to raising the level of teacher understanding and of expertise in assessment procedures associated with the devolution of responsibility for quality assurance in the certification process.
- 3 An increasing emphasis on validity in the assessment process which allows the full range of curriculum objectives including cognitive, psychomotor and even affective domains of learning to be addressed by the use of a wider range of more 'authentic' technique for gathering evidence of learning outcomes.
- 4 An increasing emphasis on describing learning outcomes in terms of particular standards achieved -often associated with the pre-specification of such outcomes in a way that reflects the integration of curriculum and assessment planning.
- 5 An increasing emphasis on drawing together academic and vocational qualifications into a single framework that will provide for equality of esteem across a range of learning purposes and styles.
- 6 An increasing emphasis on using the assessment of individual pupil's learning outcomes as an indicator of the quality of educational provision, whether this be at the level of individual classroom, the institution, the state, the nation or for international comparisons.

Clearly none of these trends exists without the overlay of institutional traditions and national ideologies which powerfully affect how they are manifested in any one educational system. In what follows I shall refer briefly to some of the current assessment development in other countries in order to illustrate how general the existence of the trends I have referred to above are. At the same time, I shall argue the equally important point that it is the differences in the way that these common trends are being translated into policy in individual countries which, as well as the similarities, underline for us that there can be no one ideal solution to the qualifications

debate. Rather, there are a series of possible trade-offs or compromises as Noah and Eckstein (1991) demonstrate in their study of examination systems in eight different countries. The nature of the compromise made in any one national system is likely to be determined by the balance of power prevailing at any one time between different professional and political ideologies and the pressure groups that represent them. To the extent that this is so, it reinforces the point that decisions about evaluation techniques and qualification systems are not merely technical. They are powerfully affected by prevailing educational priorities as these in turn reflect the dominant values of a given society. They embody, implicitly or explicitly decisions about ends as well as about means. Some brief example of how countries vary in their approach to some of the key assessment issues of the day may serve to illustrate just how significant such differences can be in reinforcing a particular vision of society.

FORMATIVE ASSESSMENT

National assessment

One of the most striking developments of recent years in relation to evaluation is also the most general. It concerns the rise to prominence of an aspect of education that was for long regarded as a necessary evil which periodically punctuated the process of education but that had little to do with it. Teachers typically had little interest in it and even less training. Now the significance of evaluation in every aspect of education is being widely recognised.

In particular, many countries are now seeking to improve the quality of formative assessment being undertaken by teachers. Often this is being done by making available suitable nationally-created assessment instruments. Thus, in the Netherlands, for example, a wide variety of diagnostic tests have now been made available nationally which teachers can use regularly to gauge the progress of their pupils. The tests are related to 'Help books' which are available to guide teachers' remedial activities. Similar diagnostic tests are available in Sweden, Scotland and France but with important difference in the latter two countries that the periodic utilization of some nationally-provided tasks for formative purpose is compulsory at certain age levels.

In other countries including England, New Zealand, the United States and some Australian States, there also nationally-provided assessment

procedures which are intended to have an important formative purpose but this role is combined with a standards-monitoring function such that the information so generated is to be used in a summative way as an indicator of quality. Where this is so, it tends to lead to heated debates concerning the incompatibility of formative assessment procedures and those assessment procedures which are geared towards providing for public accountability. As a recent Report by the US Educational Testing Service -one of the leading test institutions in the world- argues, referring to the American context:

"Accountability testing does not fit easily with instruction: it is an imposed system, usually relying on multiple-choice, machine-scored tests to hold down costs, and too often used to make superficial judgements. New performance measures do not fit with accountability: they are closely tied to work done in individual classrooms and with individual students, are more expensive to operate, and do not readily yield simple scores which can be used normatively to assess trends in standards."

The same point is made forcefully by Codd et al in relation to the New Zealand 'Achievement Initiative' and by several other commentators in relation to the English National Assessment programme (Broadfoot et al 1991; Conner 1990).

In sum then we may conclude that many countries are concerned to improve the quality of the diagnostic and formative assessment which is undertaken in classrooms in the hope that by equipping teachers to target more precisely the support they give to individual students this will raise standards by helping to increase student motivation and by making teaching more effective. Most such countries are also implementing some mechanism for national monitoring of standards. In this latter aspect however, they differ in at least two important respects. Firstly there is the issue of compulsion. National assessment may be conceived of as a resource which teachers are encouraged, but not obliged to use, or it may be mandatory. Secondly national assessment may embrace only a small sample of students and schools or it may require comprehensive coverage. In those countries, where testing is both compulsory and comprehensive, it typically becomes 'high stakes' exerting a powerful and arguably undesirable wash-back effect on the education system as a whole. Where either of these features is missing, testing becomes 'low stakes', supportive, rather than dictatorial. The crucial difference between these two positions may best be illustrated by citing two contrasting examples -England and France.

In 1974 France set up a programme for evaluating, gathering, analysing and to put at the disposition of the Ministry necessary to evaluate in their various aspects the results of training activities and their longer term evolution'. Surveys are carried out at regular intervals (at least once every 5 years) and cover a wide range of knowledge, skills, behaviour and opinions across the curriculum.

Using mainly a mixture of multiple choice and short-answer questions, surveys are carried out on a sample of schools to take account of differences in size, type and rural/urban location. For each subject, the tests take about one hour and the total may amount to a few hours for a student. The tests are administered in carefully controlled conditions, marked by teachers according to strict guidelines to ensure reliability. Recently attempts have been made to include less objective techniques including oral assessment of verbal expression, role-play, language and observation, for example, in the evaluation of body language. Attempts are also being made to assess practical skills.

By re-using certain tests these surveys provide a means of comparing standards over time, although this is problematic when the curriculum has changed. The surveys also provide the data needed for the increasingly felt need for international comparisons. Thirdly the survey results are designed to identify particular areas of concern and relate particular factors in the education system concerning teaching approach and student background, to results. This is seen as providing an important aid to policy-making.

This permanent monitoring provision has been implemented since 1989 by the implementation of a comprehensive evaluation system annually in both state and private schools to the attainment of Maths of all students in the 3rd year of primary school and in secondary school and, in 1992, the first year of post-compulsory education. This important initiative is conducted in the second week of the school year in order to provide teachers as they begin work with a detailed diagnostic picture of the strengths and weaknesses of their children so that the teacher can respond differentially to their needs. The aim is to encourage a more individualised pedagogy and to improve the overall level of student learning given the

of students experiencing difficulties in basic skills on entry to secondary school which was identified by previous national monitoring.

This initiative is a reflection both of the increasing diversity of student need and the general lack of tradition or training among French teachers in coping with such diversity. An important part of the initiative involves the provision of training courses in assessment run by national and local inspectors for teachers. Although the assessment instruments have so far been devised nationally, the longer-term aim is for more local provision to become the norm, tailored to local conditions and needs. Crucially it is hoped that in time teachers can be provided -or indeed provide themselves- with assessment materials that they will, with training, be willing and able to use as an integral part of their teaching, in order better to provide for diagnostic and responding to individual student's needs.

Information about their child's performance is given to parents, and many schools are setting aside time for discussion of results with parents. The unprecedentedly high level of parent interest and response provoked by the evaluation is regarded as one of the major achievements of the initiative providing, as it does, both for the kind of constructive, close dialogue between schools and families which has not hitherto been a feature of French education and for a measure of direct accountability to parents concerning standards.

If the aim of improving student learning through nationally initiated formative assessment also embraces training teachers in assessment and helping them to devise ways of responding to student need in a 'regular and rigorous manner', an associated aim is to encourage a school-wide response to this need. Although the main role of these assessment is officially at least to provide formative information for teachers and to encourage and equip them to be better formative assessors themselves, the programme also has an important summative dimension in that aggregated results are published nationally so that parents, teachers and headteachers can compare their 'results' against national norms.

Although the availability of such information has provoked considerable media interest nationally, the Ministry has so far resisted pressure for results to be made public and thus to provide for inter-school or inter-region comparisons. The still strongly-held French belief in the notion of equality and commonality in educational provision militates against the adoption of the notion of competition and market-forces to stimulate efforts by individual

schools to improve standards. The assumption is rather of a common problem which schools must be enabled to tackle.

The initial reaction of teachers to the very rushed imposition of this initiative was predictably negative. The profession is accustomed to resisting change, and to seeing assessment of the kind as external and imposed, is not committed to the concept of an individualised pedagogy and has little or no tradition of formative assessment. Teachers were suspicious that the aim was really to assess them rather than their students.

However, research studies report an increasingly favourable response among both teachers and parents to the tests themselves and the information produced (Ministere de l'Education Nationale 1990, 1991, Gilles et al 1990).

- National assessment in England

Although, like France, England has also launched a national assessment programme based on anonymous testing of 'light' samples of pupils in the mid 1970s, changing economic pressures and political ideology have led to a much more 'high stakes' approach to national assessment under the provisions of the 1988 Education Reform Act. The national assessment programme requires that pupils be assessed against the attainment targets of the national curriculum by their teachers continuously and by external tests (called standard assessment tasks) at the ages of 7, 11, 14 and 16. At these ages the results of teacher assessment (TA) and the external test (SATs) are combined and must be reported towards the end of the school year. The results of individual pupils are confidential to themselves, their parents and teachers; results for a class as a whole and a school as a whole are to be available to the parents; results at school level are to be publicly reported at 11, 14 and 16; publication of results at 7 is not mandatory but is strongly encouraged by the Secretary of State. The publication of results is to be as part of a broader report by the school of its work as a whole. It is suggested that such reports should include a general report for the area 'to indicate the nature of socio-economic and other influences which are known to affect schools'. At 16 the external test is to be the General Certificate of Secondary Education (GCSE) which is currently taken by approximately 85% of the age group, and the grading system of the GCSE is to be merged with the ten level national curriculum scale.

The National Assessment programme was intended to fulfil diagnostic and formative as well as summative and evaluative purposes. It was inspired by an assessment philosophy with emphasises criteria-referencing, teacher-

based assessment, active, process-based assessment tasks and coursework assessment. The aim initially was to use practical tasks and observations as well as more conventional tests to limit the possibility of 'teaching to the test' and the negative effects normally associated with formal assessment. This aim has now been replaced by a move back to more conventional paper and pencil tests as the need for comparability and ease of administration associated with summative and evaluative purposes outwiders the original priority given to supporting learning through strengthening teachers' formative assessment practices.

It is interesting to note in this respect the important contrast with Scotland which has its own education arrangements even though it is part of the United Kingdom. Compulsory tests only apply at 8 and 12 years and are limited to reading, writing and maths. Teachers have considerable choice in which test they use, allowing them to integrate testing with teaching so that the tests fulfil an important formative function. As a result of widespread parental resistance to compulsory blanket testing, teachers are now allowed to use the tests at their discretion to support their own assessment at any point in the school year. Furthermore, rather than national assessment being used as an accountability device for schools, the Scottish have retained for this purpose their earlier light-sampling arrangements to comment on national standards of performance (BERA, 1991).

Records of achievement

The development of more formative, learning-integrated assessment is not principally associated with the provision of national tests however. The widespread experimentation with 'portfolios', profiles and records of achievement in, for example, the United States, the United Kingdom, Canada and Australia represents 'a conceptual approach to assessment which brings together instructional theory and assessment methodology' (Dwyer, cited in Nisbet 1991) and which represents an important departure from traditional psychometric thinking.

There are two main dimensions to this type of assessment initiative. The first is the attempt to use 'complex, integrated and challenging intellectual tasks... for assessment which mirror good instruction' (Sheppard, 1990). The second is the associated attempt to find ways to structure recording of achievements which will prioritise feedback to the student and the encouragement of motivation rather than their best work which not only provides 'authentic' evidence of achievement but also, by including students' reflections on their own work, their reasons for the choice of work

presented, the inclusion of early drafts and especially their self-assessment, acts as a key feature of the learning process itself.

In some instances, such as the United Kingdom development of 'records of achievement' what started out as largely an initiative to provide more meaningful and comprehensive summative assessment, has evolved into much more of a formative assessment initiative for which the rationale is essentially couched in theories of learning. Thus, for example, just as constructivist learning theory emphasis that meaning, understanding and the development of conceptual structures have to be created by the learners to understand and be partners in the assessment process (Sadler, 1989).

As Nisbet (1991) suggests:

"In recent years, cognitive psychology has increasingly come to recognise the importance of affective and social elements in learning, attitudes, motivation, relationships and self-image. Learners' perceptions are all-important: their perceptions of the tasks, of the process and of their own competence. Assessment plays a crucial part in influencing these perceptions...one of the criteria for judging an assessment system is its impact on the climate of learning within the educational provision."

In his comprehensive review of such evidence, Crooks (1989) demonstrated the many ways in which assessment methods impact on students' attitudes to learning; to motivation; affects their study skills and supports or inhibits the use of the more productive learning climate which cooperation between students brings about.

I have spent some time discussing what may be regarded as some of the most telling symptoms of an impending radical change in the dominant assessment culture. A comprehensive explanation of the source of these changes would be a complex and difficult task. However, both the rapid international growth in learning-integrated assessment and that of the more or less contradictory practice of national assessment for monitoring purposes, may be traced back to economic pressures. It is pressure from employers which has highlighted the need for more curriculum emphasis to be placed on a range of skills and qualities which cannot validly be assessed by traditional written examinations and multiples-choice tests. It is also the perceived need for economic competitiveness which has fuelled the move towards more constructive forms of assessment; towards assessment procedures which are designed to raise standards by encouraging students to keep on with education rather than, as for competitiveness which has fuelled

the rapid growth of interest in so many countries in monitoring national standards of performance

Where such monitoring is associated with 'high stakes' testing, it provides a clear demonstration of the current struggle between the two assessment cultures -the one emphasising competition and a carrot and stick approach to raising standards; which highlights results but offers little guidance about how to achieve them. The other, takes a national, scientific approach to improving learning by emphasising process and by seeking to incorporate into assessment procedures what is known about how students learn. As we consider how best to move forward in relation to evaluation issues, we need to be very aware of the powerful role played by these procedures in reinforcing the legitimacy of one set of assessment concepts over another and to choose accordingly.

EVALUATION IN THE FUTURE

Earlier I identified a number of features of the new assessment culture which is now emerging. These centred on the search for assessment procedures which would be more comprehensive, more meaningful, more constructive, more professional and less divisive. Despite the backdrop of contradictory pressures in some countries towards the narrowing of curriculum goals; towards a return of norm-referenced comparisons of individuals and of schools; towards declining teacher morale and pupil anxiety, all of which have been shown to be associated with 'high-stakes' testing (Cooper and Wilson) these trends continue to develop strongly. Nowhere are they more visible than in the changing structure of formal qualifications and certification. If we take the key features of:

- a) the integration of vocational and academic qualification;
- b) the inclusion of higher-order learning outcomes and core skills;
- c) increased institutional responsibility for summative assessment;
- d) explicit statements of learning outcomes;
- e) flexibility through the modularisation and rationalisation of courses;

we may note some or all of these aspects in a comparative review of current international developments in the structure and context of qualifications.

In Spain for example, hitherto one of the most examination-oriented education systems, it is planned to do away with external examinations during the compulsory years of schooling and instead to raise teachers' professional competence to undertake this task. Post sixteen, a structure of flexible modules is envisaged including both general and vocational provision with even the prestigious 'selectividad' university entrance examination including assessments of core study skills which will necessarily be assessed in non-traditional ways. (Ministerio de Educacion y Ciencia, 1990).

France provides us with a good example of a country which has made great strides in equating and integrating vocational and academic studies under the common rubric of the 'Baccalauréat'. The Netherlands has pioneered some of the most exciting new 'authentic' assessment techniques for assessing aspects such as social skills, planning skills and problem-solving. Germany demonstrates how far it is possible to go in devolving responsibility for assessment teachers and schools since certification procedures have long operated throughout the school system with only minimal external controls (Sutherland, 1991). Sweden has pioneered the creation of a system of qualifications which greatly increases access for mature students. New Zealand and the UK may be taken as leading examples of the attempt to rationalise qualifications into a coherent or flexible system that both encourages participation and facilitates communication with users. To the UK, along with Australia and the United States, we must give credit for highly significant developments in the way in which achievements are recorded and future targets set as these processes are embodied in Records of Achievement and Action Planning developments.

Each country has something to contribute to the new assessment culture; a contribution to make to creating practices that will help us achieve the educational goals of the 21st century. Equally, countries are manifesting in different ways the problem of shedding familiar, but fallacious assumptions about assessment procedures. Qualifications that emphasis passing and failing; fixed and narrow notions of intelligence; and a restricted range of assessment techniques can have no place in a society committed to maximizing the personal and professional growth of all its citizens. Instead we need qualifications which are based on actual performance; such tests need to focus on actual performance; on comprehensive coverage of learning goals; on the application of clearly-defined hierarchical criteria as the basis for both

reliability and for 'transparency' so that those being assessed understand the criteria being applied and thus can assess themselves and direct their learning appropriately. Qualifications need to be able to address both domain specific and general learning goals; both short and long-term desired outcomes; to combine cognitive and conative dimensions; to be usable in range of potential performance contexts -verbal, symbolic, physical and social; and to take account of idiosyncratic, unanticipated learning as well as instructional goals. They also need to be designed in a form that allows for aggregation across individuals and to be relevant for making policy decisions as well as those concerning instruction and selection.

The current international interest in the generation of effective indicators of educational quality (Walberg, 1990) means there is a very real danger of invalid, shallow data and inappropriate correlations being generated in an attempt to short-cut the necessary development work that still needs to take place in designing new assessment approaches. After over a hundred years in which a particular subject-based approach to both the definition of learning goals and their assessment has predominated, and in which the canons of psychometric assessment have emphasised the need for objectivity as an overriding goal, the establishment of a very different set of assessment principles among both educational professionals and the wider population is likely to take a good deal of time and effort. As with most other problems of assessment, the issues are both technical and political involving both the generation of suitable techniques and rendering these acceptable for the various social purposes that assessment fulfils.

This paper has described some of the strides that have already been made towards this goal. "To the extent that assessment experts prove themselves able to meet the challenge of the development of a new theory of test design and validation -one that emphasis the communication of individual learning rather than of individual differences (Baker et al, 1990, p. 32), they will have done far more than provide some elements of a more valid system of qualifications, important as this is. The 'wash-back effect', will ensure that education systems around the world begin to direct their efforts towards generating those vital intellectual, social and practical competencies which for so long have been largely empty rhetoric on the agenda of both educational goals and public concern. More fundamentally still it will help to challenge individual societies to consider the kind of educational goals to which they should aspire as the start of a new milenium.

If the last century has bequeathed us any lessons for the next one, it is that the central assessment questions are not technical ones. That is

because we have allowed ourselves to become overly preoccupied with them, that our education systems are not currently meeting the needs of the societies they serve as effectively as they should. As we come to recognise the key role that evaluation procedures play in defining the shape and context of education we come closer to understanding what needs to be done to effect real improvements in learning.

REFERENCES

APU (1984) What it is and how it works, London.

BAKER E.L., FREEMAN M., CLAPTON S. (1990) Cognitive assessment of subject matter: understanding the marriage of psychological theory and educational policy in achievement testing, Boston, AERA.

BROADFOOT P.M. (1991) "Implementing national assessment issues for primary teachers", Cambridge Journal of Education, Vol. 21, N° 2, pp. 153-168.

CONNER C. (1990) "National curriculum assessment and the primary school: reactions and illustrations of emerging practice", The Curriculum Journal, Vol. 1, N° 2, pp. 139-154, Scottish council for research in education, Edinburgh.

CROOKS T. (1989) "The impact of classroom evaluation practices on students", Review of educational research, Vol. 58, N° 4 winter, pp. 468-481.

ECKSTEIN M.A., NOAH H.J. (1991) "Controlling education through examinations: a comparative study of eight nations", unpublished ms.

FREDERIKSEN J.R., COLLINS A. (1989) "A systems approach to educational testing", Educational researcher, Vol. 18, N° 9 Dec., pp. 27-32.

GILLES Marc and Associates (1990) "Bilan diagnostic des effets, chez les enseignants de l'opération évaluation-formation CE2/6ème".

HARLEN W. et al (1991) "A comparison of the proposals for national assessment in England and Wales, Scotland and Ireland", British Educational Research Association (BERA) Assessment Policy Task Group.

- HOWIE (1988) "Assessment learning outcomes", OECD Paper, Paris.
- INGENKAMP K. (1977) Educational assessment, Windsor, NFER/Nelson.
- LEE SMITH M. (1990) Meaning of test preparation, Boston, USA, AERA.
- MESSICK S. (1989) in LINN B. (ed) Educational measurement, 3rd Edition.
- MINISTERE DE L'EDUCATION NATIONALE DE LA JEUNESSE ET DES SPORTS (1990) "Evaluation CE2/6ème: résultats nationaux septembre 1990", Ministere de l'Education reports.
- MINISTERIO DE EDUCACION Y GENCIA (1990) "White paper for the reform of the educational system", Madrid.
- NICKERSON R.S. (1989) "New directions in educational assessment", Educational researcher, Vol. 18, N° 9 Dec., pp. 3-7.
- NISBET J. (1991) "Assessment in education in the U.S.A.", University of aberdeen mineo.
- RAVEN J. (1989) "Questionable assumptions in test construction", Bulletin of the international test commission, N° 28/29 april/october, London.
- SADLER R. (1989) "Formative assesement and the design of instructional systems", Instructional science, Vol. 18, pp. 119-144.
- SHEPARD L. (1979) "Purposes of assessment" in ROGERS W.T. (ed) op cit.
- SHEPARD L.A. (1990) "Psychometricians beliefs about learning", AERA Annual Meeting, Boston.
- SUTHERLAND M. (1991) "Assessment in Germany", Paper prepared for OECD, Paris.
- WALBERG H.J. (1990) "OECD Indicators of educational productivity", Educational researcher, Vol. 19, N° 4 june/july, pp. 30-33.