CSE
MONOGRAPH
SERIES
IN
EVALUATION

12

LEARNING FROM PUPIL ASSESSMENT: INTERNATIONAL COMPARISONS

Edited by
Paul Black
and
Alain Michel

A report on a conference arranged in collaboration among

The Organisation for Economic Cooperation and Development
The UK Office for Standards in Education
The UK Department for Education and Employment
The United States Department of Education

and held at the Garden House Hotel, Cambridge, UK
20-25 February 1996

CSE MONOGRAPH SERIES
IN EVALUATION

SERIES EDITOR

Eva L. Baker

Center for the Study of Evaluation
UCLA Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, California 90095

1998

LEARNING FROM PUPIL ASSESSMENT: INTERNATIONAL COMPARISONS

Edited by
Paul Black
and
Alain Michel

A report on a conference arranged in collaboration among

The Organisation for Economic Cooperation and Development
The UK Office for Standards in Education
The UK Department for Education and Employment
The United States Department of Education

and held at the Garden House Hotel, Cambridge, UK
20–23 February 1996

CSE MONOGRAPH SERIES IN EVALUATION

NUMBER

1. Domain-Referenced Curriculum Evaluation: A Technical Handbook and a Case Study From the MINNEMAST Project
   Wells Hively, Graham Maxwell, George Rabehl, Donald Sensio, and Stephen Lundin

2. National Priorities for Elementary Education
   Ralph Hoepfner, Paul A. Bradley, and William J. Doherty

3. Problems in Criterion-Referenced Measurement
   Chester W. Harris, Marvin C. Alkin, and W. James Popham (Editors)

4. Evaluation and Decision Making: The Title VII Experience
   Marvin C. Alkin, Jacqueline Kosecoff, Carol FItz-Gibbon, and Richard Seligman

5. Evaluation Study of the California State Preschool Program
   Ralph Hoepfner and Arlene Fink

6. Achievement Test Items—Methods of Study
   Chester W. Harris, Andrea Pastorok Pearlman, and Rand R. Wilcox

7. The Logic of Evaluative Argument
   Earnest R. House

8. Toward a Methodology of Naturalistic Inquiry in Educational Evaluation
   Egon G. Guba

9. Values, Inquiry, and Education
   Hendrik D. Gideonse, Robert Koff, and Joseph J. Schwab (Editors)

10. Evaluation in School Districts: Organizational Perspectives
    Adrianne Bank and Richard C. Williams with James Burry (Editors)

11. Assessing Student Achievement: A Profile of Classroom Practices
    Donald W. Dorr-Bremme and Joan L. Herman

12. Learning From Pupil Assessment: International Comparisons
    Paul Black and Alain Michel (Editors)

The work reported herein was partially supported under the Educational Research and Development Centers program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this publication do not reflect the positions or policies of the National Institute on Student Achievement, Curriculum, and Assessment, the Office of Educational Research and Improvement, or the U.S. Department of Education.

TABLE OF CONTENTS

PREFACE

Widespread interest is evident within national systems of education in defining educational objectives and measuring learning outcomes. Both tasks, defining objectives and measuring outcomes, present challenges to our professional expertise and to the ability of education systems to implement effective change.

A case can always be made for the sharing of experience across national boundaries. But, that case having been made, the circumstances must be right for the ideas to flow freely, reshaping our thinking about what has already been done and what might be possible. The Organisation of Economic Co-operation and Development (OECD) Conference held at Cambridge, England, in February 1996 was such an occasion; something much more than a routine exchange of information and ideas. This book has been prepared to bring together the contributions to the conference made by key speakers and conference delegates from twenty-one countries.

The paper presented by the opening speaker, Dr. Barry McGaw, set the scene by offering an overview of the concepts and practice associated with "Assessing student performance." Further papers on the three themes of "Formative assessment for learning," "Using summative assessment," and "Systematic reform: national policies" then provided ample stimulus for the interesting exchanges which took place in working groups throughout the conference. The set of papers collected here also includes brief summaries, prepared by delegates, outlining the current situation in respect of curriculum and assessment reform in each of the countries represented at the conference.

The conference was organised by the External Relations Team of the Office of Her Majesty's Chief Inspector England (OFSTED).

A successful conference depends on many people playing their part in organising activities and stimulating ideas. Thanks are due to Santosh Sadanand of OFSTED's External Relations Team for his often inspired support in administering this conference,

for the secretarial help of the Cambridge office of OFSTED, and to all those who chaired groups, presented papers and acted as rapporteur. OFSTED is indebted to other UK organisations, namely the Department for Education and Employment and School Curriculum and Assessment Authority for financial support, and to OECD and the United States Department of Education for their guidance and encouragement.

DIANE SIMMONDS
Her Majesty's Inspector of Schools

RICHARD DAUGHERTY
Professor of Education
University of Wales

CHAPTER 1

INTRODUCTION

PAUL BLACK, King's College, London
ALAIN MICHEL, Inspecteur Général de l'Education Nationale, Paris

The work presented in this report was not an isolated exercise—it was part of a process. A group of representatives from the Organisation for Economic Cooperation and Development (OECD) member countries had already been working together for more than a year. This group was contributing to a coordinated study under the general title of "The Curriculum Redefined." In exploring this broad theme, the countries involved had decided to concentrate on three main themes, curriculum, assessment, and teachers' professionalism. As one of the countries working in the assessment group, Britain acted as host to a meeting which was to contribute to the development of thinking about the assessment issues.

Prior to this conference, the thinking of the assessment group had been formulated in a draft document that formed one of the background papers for the conference and that will be refined as further inputs, notably from this conference, serve to develop the ideas. The eventual aim will be to tease out the main policy issues about the curriculum of the future that national and local governments ought to address. Thus, the focus on assessment is meant to identify more precisely its role in improving school learning and in the development of the curriculum, not to study assessment in artificial isolation from such issues as curriculum, pedagogy, and teacher development.

It was made clear in the background paper that any study must distinguish clearly the different purposes of assessment; the tensions among them; the interaction of assessment with the formulation of curriculum and standards; the roles of teachers and of pupils in assessment; the manageability, cost, and dependability of various forms of assessment; and issues of accountability and reporting. Within this complex of issues there were three that stood out: the nature and development of

formative assessment, the strengths and weaknesses of external summative assessments, and the need to explore comprehensive and coordinated—and therefore systemic—approaches to policy formulation for assessment. It was therefore decided that the conference should have these three topics as its main foci.

Each of these three themes was addressed by two speakers. To establish a common framework for the discussions, the speakers were preceded by an overview paper by Barry McGaw. His paper makes clear the distinction between the two broad areas of purpose, the formative for the development of learners and the improvement of programmes, the summative for accountability. It then goes on to explore two issues which are central to all three themes—the establishment and use of standards and the adequacy of the data required.

The papers by Margaret Brown and Eva Baker are both grounded in the authors' experience of working with teachers on formative assessments. Both papers stress, with different relative emphases, the benefits that might accrue from development of formative practice, and the substantial and radical changes in professional practice that such development requires from most teachers. The ensuing discussions were well focused, partly because at the level of classroom practice that is central here all countries have a great deal in common. It was also clear that this is an area for future growth in most countries, so that it was appropriate that we had two researchers/developers to discuss and illustrate the issues.

In contrast, Claudine Peretti and Hans van Aalst discuss summative assessments from different perspectives. The first presents an account of national and regional summative assessments in France, showing how a particular national policy uses such assessments as a lever for reform. In the second paper, van Aalst adopts a broader theoretical approach to explore meanings currently attached to summative assessments and to speculate about directions for future change. The discussions of this theme were more diverse, partly because of these two different perspectives. A more significant obstacle to communication was that, for this theme, the differences among the national and regional systems of different countries meant that the meanings of the terms and the assumptions underlying accepted practice were not easily shared.

The authors for the third pair of papers have both been closely involved in development of national policies in their respective

countries—David Stevenson in the United States Department of Education and Alejandro Tiana Ferrer from the National Institute for Curriculum and Assessment in Spain. Their papers on the systemic dimension describe the development of comprehensive reform policies in which the authors were personally involved. Both accounts show how, in quite different ways, the interactions of the formative and summative aspects and the particular pressures for accountability give rise to tensions, to needs for support, and to links among curriculum development, professional development, and assessment reform. The diversity here was further evident as these same issues were reflected across the many countries represented in the three discussion groups.

In seeing the work of this conference as a whole it is helpful to bear in mind not only the six different countries of the main authors of the papers, but also the sixteen further countries who were represented at the meeting. The twenty-two brief outlines of the country assessment systems given at the end of this report will serve to emphasise this point, and are also included as a valuable resource of information in their own right.

It is not possible to represent fairly in this report the very useful work done among members in the discussion groups to which half of the time of the conferences was devoted. We have chosen to use the many ideas contained in the records of these discussions as sources for our closing summary in the final chapter. There we reflect on some of the most significant points arising from the conference, including the invited papers, the discussion group reports, and our own thoughts on the issues raised. The overall purpose has been to focus on issues that merit close attention by any who may be engaged in influencing or formulating assessment policies in their own countries.

CHAPTER 2

ASSESSING STUDENT PERFORMANCE[1]

BARRY MCGAW, Australian Council for Educational Research

The purpose of this paper is to provide a conceptual framework for the discussion of educational assessment, central to the OECD conference. The paper develops the framework under three broad headings: the purposes of assessment, the point of reference in assessment, and the scope of data required. Major developments and issues are illustrated with examples of current practice.

Purposes of Assessment

The purposes of educational assessment can be divided into two broad categories, one being development and improvement and the other being accountability. The terminology introduced by Michael Scriven to distinguish similar broad categories of purpose in program evaluation could just as well be used here to name the two broad purposes of assessment. *Formative* assessment is directed to developing and improving students' learning. *Summative* assessment is directed to judging the levels of achievement that students have reached for certification or for other, perhaps aggregated, uses of the status measures in ensuring accountability for performance.

Development and Improvement

It is a relatively easy matter to assert that assessment is engaged in only for noble purposes of facilitating the development of individual learners or the improvement of programs. If the claim is made, it ought to be supported by the

[1]Based on keynote opening address delivered at the conference, *Assessment in Schools: International Comparisons*, organised jointly by the Organisation for Economic Co-operation and Development (OECD) and the UK Office of Standards in Education (OFSTED), Cambridge, England, 20-23 February 1996.

discipline of establishing how links will be built between the outcomes of assessment and the development of policies and programs.

This imposes the obligation to consider, in advance, the potential relevance of the kinds of data gathered and the links required between those who will obtain the data and those who might be the agents of change. In technical terms, it imposes a strong validity requirement on assessment practices. It is not enough that the assessment instruments or procedures provide valid measures of performance in particular domains of intended learning; they must also inform subsequent decisions. This can be thought of as a variant of the notion of "consequential validity" introduced by Messick.

Accountability

Where the purpose of assessment is linked with accountability rather than development, there are important questions of focus. The most visible, recent forms of accountability involve system-level assessments of student performance as a basis for judgment of the system. Another system-level approach is seen in the formal examination systems through which all students in a particular cohort are assessed in order to certify their levels of achievement. In that case, it is more the student than the school or the system that is being held accountable.

System-level assessments can also yield regional or institutional data and thus extend the accountability requirements downward from the system to its parts and its individual schools.

For both system-level and school-level accountability it is important to be clear and realistic about the criterion against which accountability is to be judged. The criterion could be set in terms of some desirable level of achievement to be reached, though that requires definition in some form that can make the criterion explicit. Attempts to define "minimum competency" levels of achievement in particular domains of learning, such as literacy and numeracy, are early examples of this approach.

An alternative is to set the criterion in terms of growth in levels of achievement; that is, to expect not just that systems or schools will reach and maintain some level of achievement, but that they will continue to improve on it. Where the measures used refer to some criterion scale of achievement, the notion of growth

involved is an absolute one, however well or poorly it might be measured. Where the measures used are norm-referenced, then the notion of growth is a relative one which can accommodate growth of some only at the expense of the relative decline of others.

In the case of individual schools, it needs to be recognised that there are multiple causes of both the levels achieved at any time and the subsequent changes in the levels achieved. This point is so obvious—the potential influences beyond the school itself including changes in the mix of students enrolled and a whole range of factors relating to family and other social supports—that it remarkable that anyone should even consider using straightforward measures of achievement as a direct reflection on schools.

Educational institutions are not the only ones subjected to such a simple-minded notion of accountability for outcomes. It was recently proposed that death rates among patients be published among a set of outcome measures for hospitals in one state in Australia. The implication was that a low death rate would indicate a hospital was performing well. Quite the opposite could obviously be true. Death rates would be highest in hospitals treating the most serious cases and they might well be the very best hospitals. How might such an outcome measure be used to judge the quality of care in hospitals for the terminally ill?

The inadequacy of approaches to accountability that assume simple, direct relationships between institutional practices and outcomes is obvious in this example with hospitals and should be no less obvious with schools. An alternative being explored in a number of places is the development of estimates of school effects that might be thought of as the "value added" by the school. The technical problems in obtaining good estimates are not trivial but pursuit of this alternative ought not to be abandoned. It also imposes substantial requirements on the data sets to be gathered. It is best, for example, if the "value added" is estimated using current and prior achievement in the relevant domains of learning and not current achievement and proxies for prior achievement such as socio-economic status indicators. The task of relating achievement measures on different occasions is rendered much more difficult if there is significant movement of students between schools in the intervening period.

Point of Reference in Assessment

The point of reference against which an assessment is interpreted could be the performance of some relevant cohort of students or it could be an expected standard of performance.

Performance of Cohort

Historic retreat to the norm. Early measures of human performance involved investigations in psychophysics in which human judgments of physical phenomena were calibrated against external measures of the phenomena using independently calibrated scales. Examples include human judgments of weight of object, volume of sound, pitch of note, brightness of light. The strategy typically used did not require judgment of the absolute magnitude of the property but rather judgment of which of two objects had more of the property; that is, which was heavier, louder, higher, brighter, and so on. The primary interest in this research lay, not in the measures yielded by the human judgments, but in the process of human judgment.

When psychologists turned their attention to the measurement of psychological phenomena, such as intellectual performance, attitudes, and values, they were working in domains for which there were no external scales against which to calibrate human judgment or performance. Other ways of interpreting the data were required. Thurstone's approach separated the process of scale calibration from the process of measurement in much the way these processes are separated in the construction and use of scales with physical phenomena (for example, calibrating a rule and using it to measure length). This approach introduces the possibility of the meaning of the scale providing the basis for interpretation of the performance but it was not taken up in any substantial way until the last decade or so, as new statistical models for calibration and measurement procedures have been developed and applied.

In the intervening period, the view developed that the only sensible way to interpret a person's performance was in comparison with the performances of some relevant group of other persons. The strategy was to locate the individual's performance within the distribution of other individuals' performances using scores that indicated, for example, how much above or below the average (or norm) the individual's

performance was. This historic retreat to the norm impoverished educational and psychological measurement, ceding the territory to researchers interested in individual differences rather than individual performance *per se*.

Choice of cohort. If the point of reference in interpreting the performance of an individual or a group of individuals, such as the students in a particular school or system, is to be the performance of others, then the selection of those others is central to the interpretation. It may be better to refer to the measurement strategy as "cohort-referenced" rather than "norm-referenced" to make clear that there is a choice in the norm involved as well as the use of a norm.

The most obvious approach when the focus is on the individual student is to make the cohort of reference the cohort of students of which the student is a member. If the cohort is small, for example, just a single classroom of students, the comparative data will tell very little about the student. If it is to extend beyond the classroom or the school, then some common measurement must be imposed.

Where the focus is on a group of students, such as all those in a school, a region or a system, two broad approaches to cohort definition are possible. One is to make the cohort of reference similar students at other times. The other is make the cohort of reference other groups of students with which comparison might reasonably be made.

The problem with comparisons over time is that there is usually no way to link the achievement scales, since the results are usually normed internally. The most obvious example of this problem occurs with high-stakes external examinations at key stages in education systems, such as at the end of secondary education when the examination results are used for selection of students for higher education. Here the preoccupation is with ranking students within the current group to facilitate the comparisons required in selection decisions. Comparison of results for students in one year with those in other years cannot readily be made. If the student cohort is large enough, it is usually assumed that the overall distributions of performances will be essentially similar from year to year. Nevertheless, making such an assumption and building it into the measurement strategy removes the possibility of investigating whether a change in performance levels over time has occurred.

A stronger version of the assumption is that the distribution of performance levels in the full population is so fixed that any changes in the participation rate of the population in a cohort will automatically change the standards of performance of the participating cohort. In Australia, this has led to the assumption that increased participation of the population in the final years of secondary schooling has lowered the standards of upper secondary education. In England, recently, it led to the presumption that an increase in the pass rates for various A-level examinations must have resulted from a lowering of standards and not an enhancement of performance levels.

Where the cohort of reference is students in other places, rather than similar students at other times, these could be students in other schools, other systems, or other countries. An alternative is to make comparisons with a generalised group of comparable students using "norms" derived from the performances of students in some representative sample from the population of students with which comparisons are sought.

In any of these cases, there can be a temptation to search for generous comparisons that make the local performance look good. Cannel's revelation that virtually all of the states in the United States were being shown to be above average by normative comparisons of student performances invites the conclusion that the definition of "average" had been rendered meaningless, either by manipulation of the norms to which comparisons were made or by manipulation of the local group or its performances.

Cross-national comparisons may seem to be immune from that kind of manipulation but there can be similar difficulties because of differences in definition of the participating populations from which samples are drawn, particularly for levels of schooling at which participation rates are less than 100 percent, and differences in the validity of the test for local curricula. Nevertheless, the comparisons can provide a useful provocation to review local performance levels and policies and practices.

Standards-referenced Assessment

A scale used to display the results of a survey of students' test results illustrates the transition from cohort-referenced assessment to standards-referenced assessment. One of the questions asked by the Minister of Education who

commissioned the survey was what proportion of the students were failing to reach some minimum acceptable level of achievement in numeracy and literacy.

The scale displayed specific items in the tests along the axis in order of difficulty, for Grade 5 and Grade 9 tests. The items give meaning to the scale, their ordering by difficulty up the scale showing what increasing levels of performance measured by the scale mean in terms of numeracy tasks able to be performed. The broad descriptions used, ranging from "simple arithmetic calculations" to "more complex problem solving" give a general sense of the development. Numerical values on the scale, ranging from around 20 to 60, provide a metric which is helpful but arbitrary in the same sense that the numerical Celsius and Fahrenheit scales provide helpful, but numerically arbitrary, scales for talking about temperature.

With tasks located on the scale it is possible to determine a location that represents "minimum acceptable performance for adult life" in the manner sought by the Minister. The location defined in this way is at 35 on the scale and about 95 percent of Grade 9 students were performing above that level.

The definition of the point on the scale represented by 35 as the minimum acceptable level for adults, and thus for students completing the compulsory years of schooling, could be contested, of course, but the argument would need to be engaged in terms of the scale itself. It would be possible to argue that all adults should be able to perform satisfactorily on more difficult tasks than those located up to the point of 35 on the scale and so to set the cut off higher than 35. This debate need pay no attention to the proportion of students above the level. The argument can proceed in terms of desired standards of achievement, which can be expressed in terms of location on the scale of achievement.

Levels of generality of standards. Much more interesting than an argument about the location of particular points, such as "minimum competency," on a scale of achievement is development and interpretation of the whole scale. A fully developed scale can represent a developmental sequence which can, in turn, define successive standards of achievement to be reached as students progress through their schooling. This invites definition of standards as a way of defining scales and driving test development and scale construction. It can give

primacy to curriculum considerations and not assessment considerations.

Beginning with the specification of standards raises the question of how detailed the specification should be. At the most general, it would be possible to set broad goals for schooling but it would be difficult to translate these into sensible scales. Their purpose can be to set the framework for more detailed specifications. An example of a set of broad goals is provided by those adopted by Australian Ministers for Education:

1. To provide an excellent education for all young people, being one which develops their talents and capacities to full potential, and is relevant to the social, cultural and economic needs of the nation.
2. To enable all students to achieve high standards of learning and to develop self-confidence, optimism, high self-esteem, respect for others, and achievement of personal excellence.
3. To promote equality of educational opportunities, and to provide for groups with special learning requirements.
4. To respond to the current and emerging economic and social needs of the nation, and to provide those skills which will allow students maximum flexibility and adaptability in their future employment and aspects of life.
5. To provide a foundation for further education and training, in terms of knowledge and skills, respect for learning and positive attitudes for life-long education.
6. To develop in students:
   - skills of English literacy, including skills in listening, speaking, reading and writing;
   - skills of numeracy, and other mathematical skills;
   - skills of analysis and problem solving;
   - skills of information processing and computing;
   - an understanding of the role of science and technology in society, together with scientific and technological skills;
   - a knowledge of and appreciation of Australia's historical and geographic context;
   - a knowledge of languages other than English;
   - an appreciation and understanding of, and confidence to participate in the creative arts;

- an understanding of, and concern for, balanced development and the global environment; and
- a capacity to exercise judgment in matters of morality, ethics and social justice.

7. To develop knowledge, skills, attitudes and values which will enable students to participate as active and informed citizens in our democratic Australian society within an international context.

8. To provide students with an understanding and respect for our cultural heritage, including the particular cultural background of Aboriginal and ethnic groups.

9. To provide for the physical development and personal health and fitness of students, and for the creative use of leisure time.

10. To provide appropriate career education and knowledge of the world of work, including an understanding of the nature and place of work in our society.

In moving to more detailed statements, important questions about the level of generality remain. In some cases, expected outcomes of schooling are declared in generic terms such as communication skills and problem solving skills, inviting the conclusion that these might become the developmental focus of curriculum and pedagogy and then the focus of assessment. Such generic specifications of desired outcomes are often the product of committees given general briefs to define purposes of schooling and to set a framework for thinking about quality outcomes. This is particularly so when employers' perspectives are introduced to offer outcome specifications in terms of "employment-related competencies."

Psychological research on the acquisition of expertise suggests that high level expertise is quite specific to domains of knowledge, however, and that it does not transfer readily between domains. Generic competencies are, therefore, weak competencies that might serve a person well when moving into a new domain as a novice. They will not provide the basis for the distinctive problem representations and problem solving strategies of experts in a particular domain. If expertise is being sought, then the outcome specifications need to developed in some detail in specific domains and not kept at a level of generality that suggests substantive relevance across domains.

Sources of standards. The first step then is to choose the domains into which the curriculum is to be broken. The choice is somewhat arbitrary but not trivial. In the case of Australia, eight broad learning areas have been chosen: the Arts, English, Health and Physical Education, Languages other than English, Mathematics, Science, Studies of Society and Environment, and Technology.

The choice is quite similar to that made in England, the key difference being that Australia has brought history and geography together in social studies but added environmental issues as well to create a domain called "Studies of Society and Environment."

The difficult task is not the creation of the domain categories but the specification of standards within them. For that, the issues of sources for the standards is crucial. One approach is to draw on past curriculum statements, transforming specifications of what is to be taught into specifications of what is expected to be learned. The translation is not straightforward, however, for two reasons. The first is that the conceptualisations of what is taught and what is learned are often different. The second is that the two differ in scope. Curriculum statements tend to set out "all that might be taught," which would translate into "the most that might be learned" and that would provide a definition of expected learning that would be unrealistically high.

Appeal can be made to normative information as well as to prior curriculum statements in seeking to specify standards. Knowledge of what students typically can do at various stages of schooling can provide a useful guide. It need not inhibit attempts to raise standards by setting higher expectations but it can bring a useful touch of reality to the initial specifications. In some domains, such as mathematics and science, there is a sufficient body of research on learning to give some guide to the development of a sequence of expected outcomes over successive years of schooling.

As a number of nations move to the specification of standards of achievement expected in schools, cross-national comparisons can also guide in the adjustment of local specifications. The fact that others expect much more or much less might lead to reconsideration of local expectations and priorities.

The specification of expected learning outcomes in each key learning area in Australia has involved the development of a *statement* that provides a framework for what will be taught and

defines the learning area in terms of *strands* that specify content and process. In English, for example, the strands are speaking and listening; reading and viewing; and writing. In Studies of Society and Environment they are investigation and communication; time, continuity, and change; place and space; culture; resources; and natural and social systems. Among these strands, the first involves key processes and the other five identify key concepts to be learned. Strands are broken into *strand organisers*. In the case of English, the strand organisers are texts; contextual understanding; linguistic structures and features; and strategies.

In addition to the statements, there are *profiles* that set out what students are expected to learn in terms of outcomes to be achieved by students during the first ten years of schooling. The progression is set out in eight levels. For example, the outcome for the first level for the strategies strand organiser for the speaking and listening strand of English is "monitors communication of self and others"; for the fifth level, "assists and monitors the communication patterns of self and others"; and for the eighth level, "uses listening strategies which enable detailed critical evaluation of texts with complex levels of meaning."

Speaking and Listening is a strand of English that is not usually specified in curricula in the same kind of detail as reading and writing so specification of learning outcomes for this strand proved to be quite novel for teachers in Australia. However, the specification of the outcomes proved to be insufficiently precise to guide test development. Test developers at the Australian Council for Educational Research using the profile had to establish a more precise set of outcome statements for the speaking and listening strand.

Debate about framework

Adopting a framework for the curriculum and specifying learning outcomes will not end debate about how to think about standards. It will sharpen it.

There will be normative questions about the suitability of the classification of learning areas. Questions to be faced will include whether some areas are privileged (such as mathematics and science which usually occupy discrete cells in any classification) and whether others are too broad and incoherent

(such as the Arts and Studies of Society and Environment, in the Australian case.)

There will also be questions about what is missing as well as about the organisation of what is included. In the Australian case, there has been some dissatisfaction with the treatment of values. There are value components in the outcome statements in many parts of the profiles but it has been suggested there is no coherent value framework. As a response to this critique, two groups of Christian schools (Anglican and Catholic), Islamic schools, and Jewish schools in Western Australia have worked together and with academic consultants to develop a minimum values framework as an additional component of the curriculum structure. In doing so, they have adopted the language of the current reform and developed a specification of learning outcomes with which to represent the value dimension.

There will also be empirical questions about how the specified outcome sequences fit with the actual developmental sequences of students moving through the school years. Subsidiary questions to be faced will be how students whose development does not fit the standard pattern will be treated and, more generally, whether some particular developmental sequence will gain a dominance gives an advantage to the cultural group whose development it best represents.

Scope of Data for Assessment

In addition to questions about purposes and point of reference, there are important questions to be considered about the scope of data for any assessment program. One set of questions relates to the scope of the population of concern, another to the sources of data to be gathered.

Scope of population of concern

There are some circumstances in which the concern is with a full cohort. One obvious case is where something needs to be said from the assessment program about the performance of all individuals and not just about the system as a whole or in some part.

It may also be necessary to have data for a full cohort if there is to be any substantial investigation of the performance levels and circumstances of subpopulations with a relatively low

membership. They would be too few in any representative sample of the population as a whole and they may be distributed in a way that does not include enough of them in a general sample that is otherwise representative of the overall population.

Finally, full data sets are necessary if there is to be any effective attempt to investigate school effects in order to establish the value added by individual schools. This kind of estimation is best done with performance measures obtained on more than one occasion. If there is considerable movement of students between schools, then only a universal data set will be likely to yield a sufficiently large set of data for students remaining in a school to enable the estimations to be made.

There will be circumstances, however, where data from a sample will be sufficient to address concerns about a full cohort. If the general level of achievement of students in a system is required, this can be adequately estimated using only a sample. In fact, working with only a sample rather than a full cohort, limits the size of the data sets to be gathered and analysed and can permit more adventurous data gathering to give a richer representation of student learning.

Sources of data

Data sources can be school-based or external. In system monitoring they are typically external but the dominant mode of assessment in the routine of school life is school-based. It is not a question of one or the other but rather one of balance.

With cohort-referenced assessment, it was often argued that teachers need some assessments that set their students' performances in a larger context to give them a better idea of how their students' performances match up to the rest of the cohort. This need is usually seen to be greater for teachers who have been in the one school for a considerable period since their recent professional experience will not of itself provide the wider perspective.

When the interpretive focus is shifted from the performance of others in a cohort to a set of standards, it can be argued that individual teachers and schools do not require evidence about what others can do to interpret the performance levels of their own students. However, some recent evidence suggests this may not be the case.

Data were obtained from teachers at Grade 3 and Grade 5 to assess students' performance levels based on the outcome statements in literacy The calibration process used locates the relative difficulties of the outcomes separately for Grade 3 and Grade 5 teachers. The data showed that Grade 3 teachers differentiate between outcome Levels 2 and 3 much more than do Grade 5 teachers. In fact, Grade 3 teachers perceive the increase in difficulty from Level 2 to Level 3 to be so much greater than do the Grade 5 teachers that the Grade 3 teachers equated outcome Level 3 to the same range of difficulty as Grade 5 teachers judged Level 4. Similarly, the Grade 3 teachers locate Level 4 outcomes at a difficulty level used by Grade 5 teachers for Level 5 outcomes. A similar set of differences is evident in other data comparing the judgments of Grade 5 and Grade 10 teachers.

The point of this analysis is that it shows that teachers at different grade levels do not use the outcome statements that specify standards in the same way. They make finer discriminations in the range of outcomes in which their own students predominantly operate. That finding is consistent with much psychological research on perception which shows finer discriminations are made among things closer to a person's own experience or position (for example, in differentiation of values or attitudes, in judgments of distances between geographic locations).

For the use of specifications of standards of achievement expected of students, the data suggest that teacher may well need the support of external judgments across grade levels as a check on their interpretations of the standards if consistent use is to be made of the standards.

Summary

The purposes of assessment can be for improvement or accountability. In either case, it is important that the links are clear between the assessment program and its data on the one hand and fulfillment of the purpose on the other.

The interpretation of much educational assessment has traditionally been cohort-referenced because it was accepted that the only way to provide a sensible interpretation of what one person could do was by comparison with what others could do. Modern psychometric methods have separated scale calibration

and measurement in a way that permits scales to be developed that have meaning in terms of the relative difficulties of tasks calibrated along the scale. It is possible to begin with the scale construction and let the meaning arise from the product but it is also possible to start with scale specifications (standards definitions) and to construct scales which reflect them. With this latter approach, the curriculum specification can drive the assessment program and not the other way round.

There is, however, clearly more to be learned about how to set and use standards, and there is evidence that, to ensure consistent use by teachers working at different grade levels, supporting data from outside schools needs to be added to what is gathered by in-school assessment programs.

CHAPTER 3

FORMATIVE ASSESSMENT FOR LEARNING: GENERAL ISSUES ILLUSTRATED BY EXAMPLES FROM ENGLAND

MARGARET BROWN, School of Education, King's College, London

The Nature of Formative Assessment

It is clear that there is a growing interest in many countries in using assessment to provide feedback to students and teachers for the purpose of improving standards of learning.

For example in mathematics alone, both the National Council of Teachers of Mathematics (NCTM) in the Assessment Standards (1993) and the Mathematical Sciences Education Board of the National Research Council (1993) have highlighted the importance of formative assessment. Lauren Resnick has followed up her earlier work (e.g., Resnick & Resnick, 1992) by stimulating the production of materials for teachers (New Standards, 1995). Similar developments leading to classroom materials are taking place in Ontario ("benchmark tests"), Australia (e.g., Masters, 1994), and elsewhere. Within Europe, France, Holland, and other countries are introducing national assessment systems which have formative purposes. However, there has as yet been little widescale evaluation of the effects of new forms of assessment on learning.

It is not my intention to give in this paper a survey of developments worldwide, such as that by Broadfoot (1995). Nor will I stop to survey the existing literature on and theory of formative assessment, since others have already undertaken this more than adequately (e.g., Black, 1993; Gipps, 1994). Instead I have chosen to consider some specific issues in the field of formative assessment using examples from both development and from evaluation in which I have been personally involved in the UK.

First, it is necessary to clarify some meanings. Teacher assessment (referring to assessment by teachers), authentic assessment, or performance assessment involve classroom-based modes of assessment that reflect broad aspects of performance in a normal classroom context. However, these assessment modes are not necessarily formative in the sense that "assessment information is used, by both teacher and students, to modify their work in order to make it more effective" (Black, 1993, p. 49). Similarly, national written testing, as in France, may produce diagnostic information on students' attainment that teachers and students may use formatively. Thus formative assessment describes a function and not a method of testing.

Nevertheless, clearly some forms of assessment are more appropriate for formative assessment than are others, because of features such as the nature of the design framework, the mode, the method of assessment, the style of reporting, and the validity.

In concentrating on formative assessment I will not take into account more global uses of information even if these have formative aspects. For example, a teacher who evaluates his or her teaching by examining overall class results on a topic test and who may decide to re-teach the topic is using information formatively but not at the individual student level. Alternatively, a teacher may give a student a low grade and reassign the student to lower book or a lower-level class. This would involve action at the student level but the modification of work would be only in a global sense.

True formative assessment requires the result of assessment to convey meaning to the teacher or student in terms of providing evidence to suggest whether the student possesses a competence, understanding, or appreciation which is the aim of further learning activities.

A belief in the efficacy of formative assessment does not require a constructivist theory of learning, but a constructivist theory renders invalid the traditional behaviourist "teach a skill—then—test it" model, and thus enshrines formative assessment as the only reasonable course of action. (I have argued elsewhere (Brown, 1993) that the continuing battles over assessment in the UK and potentially elsewhere reflect uncommunicated differences in epistemology between the behaviourism of politicians, parents, and many teachers, and the constructivism of educationists and other teachers.)

Constructivism is a tough faith for teachers to follow. It requires a sound subject knowledge in the sense of a grasp of an overview of the significant understandings and skills and the interconnections between them that form the basic structure of the subject. Constructivist beliefs then demand that each student's progress in learning that subject be assessed in relation to these understandings and skills, and finally that that knowledge be used to set further work activities. These may well be the same activities as are deemed appropriate for all or some other students in a class, but they may sometimes be different.

Although these requirements do not make for an easy life for a teacher, there are teachers who successfully employ formative assessment as part of constructivist practice, as can be demonstrated in the next two sections. Whether or not this is an effective way of raising standards remains to be seen.

Introducing Formative Assessment in Secondary Mathematics: The Graded Assessment in Mathematics (GAIM) Project

Background

The need for formative assessment became clear to me while working on a series of projects on the development of mathematical understanding in children.

The first and largest project enabled us to identify key steps in mathematical attainment in a number of domains, and to find what proportion of secondary students of different ages had attained those steps (Hart, 1981). This revealed the fact that there was a huge range of attainment in each year-group, that the range differed hardly at all from year to year, and that the mean attainment advanced very slowly. This and other data suggested that in five years a child was only likely to progress across about a third of the range in the sense that, for example, students in the lowest third of the population at age 11 would only have reached at age 16 the positions occupied by the middle third at age 11. They would not, by 16, have caught up with the performance of the top third at age 11. Wiliam has demonstrated that this pattern of attainment and progression is typical in a number of subjects including English (Wiliam, 1992).

This pattern was in contrast to the assumptions of most curriculum materials in mathematics, which march doggedly forward from year to year. Certainly there seemed to be a poor match between the implemented and the attained curriculum for many students, as has been demonstrated at primary level in mathematics and other subjects (Bennett et al., 1984).

Our results suggested that teachers often badly overestimated what students knew; however, in some cases they underestimated it. For example, in one school we asked to assess the understanding of place value and decimals of twelve children, one in each of the three attainment bands for each of four year groups. We were surprised to find that the second highest attaining child of the twelve was in the middle band of the youngest age group, and out-performed a student in the top group two years older. Thus research generally suggests that there is a need for formative assessment since many teachers do not always have a firm knowledge of students' areas of competence, and are not able to provide teaching at an appropriate level.

This and other experiences suggested that the structure we had identified and the assessment questions we had used to analyse the development of mathematical understanding (Hart *et al.*, 1985) could be used by teachers as a formative assessment tool. The materials, known as the *Chelsea Diagnostic Tests*, have been used in over 20 countries but did not sell well at home, presumably because there was no incentive for teachers to use them. The tests provided information on students' attainment that teachers found difficult to act on, since most were committed to a textbook scheme from which it was difficult to depart.

We found this problem also when evaluating another scheme of authentic assessment arranged in levels which related to a popular set of textbooks. Some students were failing to get beyond Level 1 having repeated the assessments several times, yet they were working through books pitched at Level 3 which assumed earlier skills. The teachers admitted that the material was inappropriate but felt they had no other option given that no other texts were available and given the difficulty of having students in the same class on different texts (Close and Brown, 1989). This suggests that an important principle of formative assessment is that there is little point in it unless there is both the willingness and the resources to act on information obtained.

However, in 1983 I was asked by the Inner London Education Authority and what is now the University of London Examinations and Assessment Council to lead a team of teachers and researchers in developing the Graded Assessment in Mathematics (GAIM) project. This ran alongside parallel schemes for Science (for which Professor Black was a consultant) and for Technology, English, and Modern and Community Languages. The outcome was to be an assessment framework built around levels of attainment, together with related authentic assessment materials. These were to be used at any time by teachers both for formative assessment and summatively to monitor students' progress across the age range of 11 to 16 years. It was an important principle that the results of the scheme should be reliable and valid enough to enable them to be used as a basis for the award of a grade in the General Certificate in Secondary Education at age 16. This was expected to provide an incentive for teachers to use the scheme, since it provided a continuously assessed alternative to what was for most students an externally set and marked examination.

Structure

The levels in the assessment framework were designed using the results of research so that students would expect on average to progress at an average rate of one level a year. In order to cater for the extremes of attainment it was concluded that 15 levels would be needed; students would start on whatever level was appropriate at age 11. The median starting point was expected to be Level 6, although children with learning difficulties (i.e., students in the third percentile) might start as low as Level 1 and very high attainers (i.e., 98th percentile) as high as Level 11.

Using research results from earlier projects, our own and those of other groups, we were able to define a set of "topic criteria" in each of the areas of numbers, algebra, space, probability and statistics, and logic, for each of 15 levels. Each criterion was separately trialled and was illustrated by at least one example to convey to teachers the level and range of performance expected. We found that isolating criteria that represented significant steps in understanding, expressing them clearly at the most appropriate level of generality and with appropriate examples to convey meaning, and checking that they

these steps could be identified reasonably reliably by teachers in their students, was an exacting and lengthy process. Nevertheless we found, as others have done previously (e.g., Hoge & Coladarci, 1989) , that teachers were able to understand and internalise the criteria as a result of being willing to invest time in working with them, and were able to develop confidence in using them to assess students. Moreover many teachers volunteered that the effort of understanding and internalising criteria, in discussion with other teachers, had greatly improved their own professional knowledge and skills and had made them more competent in professional judgements about, for example, the quality of curriculum materials.

Assessment Materials and Modes

Following the recommendations of the Cockcroft Report (Department of Education and Science, 1982), realistic problem-solving tasks and mathematical investigations were selected to be the major assessment instruments. These were open in the sense that they could be tackled at different levels to provide differentiation by outcome, in the same way as for a piece of writing. From the teacher's point of view this was helpful in that they could be used by a whole mixed ability class, working either in small groups or individually. Eighty such tasks were devised and trialled by groups of teachers, and for each one detailed assessment notes containing examples of students' work and cross-referenced to the topic criteria were provided. These are similar to the teachers's notes in the draft performance standards from the United States (New Standards, 1995), but more detailed and indicating a range of performances characterising different levels.

Students could receive certificates when they had been assessed as achieving a core of the topic criteria at any given level once they had in their portfolio more than eight open tasks that had been assessed at that level or above. Alternatively, at any time students could achieve a "record of achievement" indicating progress in each topic area by level.

Because teachers found it difficult to assess all topic criteria by open assignments, a more closed, but not routine, "topic task" was provided for each criterion (providing "differentiation by task"). Criteria were also cross-referenced to tests and assignments in the most popular curriculum schemes. Teachers

were told that they had to wait at least a fortnight and preferably longer between teaching and testing, in order to make sure short-term learning was not credited.

Teachers were also encouraged to credit students with any criteria for which they observed the evidence, or elicited it by discussion, provided such ephemeral evidence was not used for more than half the criteria. We found that when teachers first adopted the scheme they tended to use the more instrumental forms of assessment, such as topic tasks and curriculum tests, but as they internalised the criteria they tended to move towards more informal methods of observation and discussion.

Self-assessment

Record sheets were provided by the GAIM team for students' own use, so criteria were translated into forms readily understood. In many schools, students successfully maintained their own records, checked by teachers, and took on much of the workload that is a barrier to teachers undertaking formative assessment.

In many cases, this led to students claiming that they could fulfill additional criteria, in which case teachers were expected to demand evidence before crediting them. Although self-assessment records were produced for each individual level, teachers were encouraged to also give students record sheets for a higher level than that they were working on, so that they could see where they were going and be encouraged by the few higher level criteria that they were likely to have already achieved. Many students, especially in more deprived schools, were highly motivated to understand what was required in higher levels and to demonstrate they could achieve these. Teachers felt that this feature of sharing criteria was important in raising students' mathematical standards.

Implementation

The project started off with ten London pilot schools and finished its development phase with 77 schools spread across England. Participation was on a voluntary basis, although sometimes strongly encouraged by a local mathematics adviser. Often one school's participation would bring in other schools from the same locality. Each school joined a cluster served by a

local assessor responsible for induction of new schools and teachers. Central training courses were provided also, at which annual attendance was required for participating schools. The emphasis was on sharing practice and agreeing on common standards. Local assessors were responsible for visiting schools to check ten percent of portfolios, so as to moderate and correct standards, especially for GCSE certification. As a result of agreement ("social moderation") meetings and assessors' moderation visits, it was judged that teachers were able to make judgements which were reliable; the GCSE statistics suggested that these were at least as reliable as those of external written examinations.

National Curriculum Links

In 1987, it was announced that there would be a national curriculum for England and Wales. The selected framework of levels and criteria, teacher assessment, and open tasks (Department of Education and Science, 1987) was modelled on those piloted by GAIM and the other graded assessment schemes, although considerable differences in the detail necessitated revision of the GAIM materials to match. Nevertheless, the graded assessment teams were each awarded one of the first set of contracts to design and pilot national assessment tasks for age 14.

Changes in the Climate

Unfortunately, over a period of about five years after the start of the national curriculum implementation in 1989, there was a change in the political climate towards "back to basics" in education, accompanied by four changes of Secretary of State, two further revisions of the national curriculum in mathematics and science, changes in the national testing, and general turmoil in the schools. Teacher assessment was initially implemented with the enthusiastic support of local advisory staff at Key Stage 1 (ages 5–7) but only half-heartedly at Key Stage 3 (ages 11–14). It was not supported centrally by either ideological commitment or by funds, and the frustrated and over-worked teachers used the heavy additional unpaid workload of formative assessment as the legal justification to boycott all national assessment procedures over 1993 and 1994. Part of the price the

teacher unions exacted from Sir Ron Dearing for agreeing to participate in the national tests was the removal of any requirement to undertake teacher assessment, except as a one-off intuitive summative judgement against a set of level descriptions. This move reinforces one of the earlier GAIM findings that formative assessment will not be readily implemented when it is imposed, but only when teachers have previously adopted or come to adopt beliefs and practices that can be loosely characterised as constructivist and that fit in with working in this way.

Nationally, attention was given to the development of national tests, which, from the more progressive pilots in 1990 and 1991, became increasingly traditional and paper-and-pencil-based in style and content. This was particularly the case with GCSE at age 16, which, in 1989 and 1990, could be 100 percent by teacher assessment, but from 1994 at least 80 percent of the marks had to be awarded from externally set written examinations.

Given this hostile climate in which teacher assessment and open tasks were no longer valued and could not contribute more than 20 percent to students' grades at the end of compulsory schooling, and a situation in which teachers were fully occupied in adapting—without suitable teaching materials—to a curriculum and examinations that were regularly changed, it is surprising that GAIM survived at all. There was no incentive for teachers to devote time to their own formative assessment which was in addition to all the other requirements. Nevertheless, there are still 58 schools formally taking part in the scheme and many more using the assessment materials. Some of the 58 are new schools, although there are a large core of schools where the teachers have remained committed in spite of hostile circumstances. This suggests that when teachers have voluntarily operationalised and integrated into their practice a scheme of formative assessment, they do not abandon it lightly.

The School Curriculum and Assessment Authority has recently been forced to provide a slim additional document containing exemplars of assessed tasks to assist teachers in achieving consistency in their summative assessment at the end of each key stage, something in the style of the Performance Standards books referred to earlier (New Standards, 1995). In the Key Stage 3 version of this document (SCAA, 1995), more examples

are taken from GAIM tasks, both practical problems and investigations, than from any other scheme.

The development and implementation of GAIM and the other graded assessment schemes at secondary level have highlighted some important issues concerning formative assessment. However, it should be noted that this was a development project which was not rigorously evaluated. The next section, which describes results of a research project to evaluate teacher assessment at the primary level as part of implementing assessment of the national curriculum, reinforces some of these conclusions.

Evaluating the Introduction of Teacher Assessment in the Primary School: The National Assessment in Primary Schools Project

Formative Assessment in the National Curriculum

The second initiative concerning formative assessment in which I have participated, as Co-Director with Professor Caroline Gipps at London University Institute of Education, is a six-year evaluation project, National Assessment in Primary Schools. This project monitored the introduction of both teacher assessment and national tests in Key Stage 1 (ages 5–7) and Key Stage 2 (ages 7–11). We studied 32 schools—eight from each of four Local Education Authorities—in each phase, selected as a stratified random sample. The education authorities were situated in different parts of the country and had different socio-economic characteristics.

The report of the Task Group on Assessment and Testing (TGAT) (Department of Education and Science, 1987), which set the framework for national assessment, gave the highest priority to formative assessment by teachers, with national standard assessment tasks as part of the moderation arrangements. The rationale was clearly expressed, highlighting the role of formative assessment in raising standards. Nevertheless, as already noted, the government and the various curriculum and assessment agencies and their directors were distrustful of teacher assessment and chose to concentrate finance and effort on national tests as the main instrument of national assessment. Meanwhile, teachers were left with the legal requirement to carry out and report the results of their

teacher assessment but with very little central guidance as to how to go about it.

The first phase of schooling in which national assessment, both national tests and teacher assessment, was implemented was at the end of Key Stage 1 (age 7) in 1991. Although Local Education Authorities worked quickly to get systems and staff in place, and to run training sessions for teachers and headteachers, teachers had to report results for all the core subjects in 1991 with very little training or support. Although we interviewed teachers shortly before the deadline, many were unable to describe clearly the processes they were going to use. Many schools had experienced a frenzy of activity designing a variety of complex record sheets, but the actual methods of assessment remained at a vague hypothetical level.

Formative Assessment at Key Stage 1

Because of the problem of obtaining clear information, the following year, six of the 32 schools were selected for more detailed case studies of teacher assessment methods. The information from these allowed us to characterise different styles of teacher assessment, which we were later able to use in generalising to all other teachers of Year 2 children according to how they performed their assessment for the 1992 round. There were basically three "ideal" types: intuitives, evidence gatherers, and systematic planners.

Teachers' beliefs and behaviours could be described in terms of these types, although most teachers demonstrated aspects of more than one ideal type. The first two of the three types will be described very briefly here as it is the third type, the systematic planner, which was closest to embodying true formative assessment. More detail is given in Gipps et al. (1995a).

About one in five teachers we characterised as intuitives, who found it difficult to relate the national curriculum framework and systematic assessment into their practice. In some cases, this was for ideological reasons because they were loyal to a child-centred model of education commonly espoused by early years teachers and they objected to a reductionist subject-centred model imposed by the central government. In other cases, very experienced teachers were used to operating within their own framework of teaching and testing and, although not necessarily antipathetic, found it difficult to adapt to new ways of thinking

and working. Many intuitive teachers claimed with some justification to be carrying out formative assessment, but they were working unsystematically and within their own constructs which they found difficult to share. When faced with the need to provide teacher assessment levels, they tended to make intuitive judgements based on what they believed to be their knowledge of children, without referring to any evidence.

Teachers who were closest to the evidence gatherer type (about 50 percent of our sample) were those who conscientiously tried to adapt their practice to the national curriculum, but who prioritised teaching over assessment of individual children. They related their teaching plans, either for the whole class or for separate subgroups, to the national curriculum, but did not divert from those plans to any great extent. Thus, assessments mainly took the form of pupils' work that was collected and stored, often in individual portfolios. When these teachers came to a point when they needed to record their assessment, they assembled together all their evidence and tried to use it to make a judgement against the criteria.

In contrast, about a third of teachers, who were characterised as systematic planners, were operating a formative assessment system. Although they had longer term aims, they planned and recorded on shorter timescales, often weekly. They tried to assess regularly against the national curriculum criteria, often by setting specially designed tasks to groups and observing the responses of individual children. These responses were then fed into the planning of work for small groups, and sometimes individual children, in future weeks. These teachers often took notes, either jotting them down while they were working or at the end of the day.

Some of the systematic planners had internalised many of the criteria statements in the national assessment framework, and tended to rely more on informal observation, discussion, and note-taking, integrating teaching and assessment; others organised more distinct assessment events and collected more systematic feedback.

It is not difficult to see why teachers, especially conscientious systematic planners, have become first angry and then cynical as a result of continual changes in the criteria statements within a six-year period, with mathematics and science statements now in their third version. Each change has required teachers to

redesign record sheets and put in the effort of relearning statements in order to make formative judgements.

Because of these curriculum changes and the boycott of national assessment which resulted in considerable loss of momentum in the implementation of teacher assessment over the years 1993 and 1994, it is difficult to discern a longterm view of teachers' development. Over time some movement from intuitive to systematic planners was observed, where teachers became less hostile to the national curriculum framework. From our observation it appeared that the style of teacher assessment adopted was that most in tune with a teachers' beliefs; constructivist teachers tended to be either intuitive or systematic planners, while teachers who were more inclined to transmission views emphasised their planning of teaching and tended to be evidence gatherers.

It should, however, constantly be reiterated that individual teachers could rarely be matched precisely to one "ideal" type; most expressed complex views about teaching and learning which demonstrated aspects of both transmission and construction beliefs.

Formative Assessment at Key Stage 2

The research on methods of formative assessment was extended to teachers in the final year of Key Stage 2 (Year 6, age 11), either in the same school or in other randomly chosen schools in the same area. Although there was considerable overlap with practices used by teachers of 7-year-olds, there were some clear differences. Perhaps not surprisingly, there was more use of class testing and fewer teachers who were strongly loyal to a child-centred methodology.

The four types of teacher assessment which emerged from this study were markers, testers, frequent checkers and diagnostic trackers (Gipps et al., 1995b).

Markers, about 30 percent of our sample of 28, worked in a way that is more typical of secondary teachers, focussing on teaching so as to cover content, setting class tasks and marking them according to their own intuitive judgements with no reference to the national curriculum criteria. They thus shared some features with the less ideological Key Stage 1 intuitive teachers, and had made few changes to accommodate their practice to national developments.

Testers, about 40 percent of the sample, also emphasised teaching but tended to plan and assess with relation to national curriculum criteria, often in collaboration with colleagues. Assessments were planned well in advance, usually on a class basis and towards the end of planning cycles. Thus testers were similar to evidence gatherers at Key Stage 1, keeping children's assessed work in order to perform summative teacher assessment at the end of the year.

Other teachers adopted more formative styles. Frequent checkers, who made up about 20 percent of the sample, had some features in common with testers but tended to use more frequent, more informal and more differentiated assessments. Although planning was long term, assessment information was used more readily to modify plans. However, the focus was still on teaching, albeit often at the small group level rather than individual learning. Hence there were formative aspects to practice.

Finally the only truly formative practice was among diagnostic trackers, about 15 percent of teachers, who behaved in a similar way to the subgroup of systematic planners at Key Stage 1 who managed to integrate assessment and teaching. They used day-to-day tracking of individual children by questioning, observing and recording incidents as they happened, using the framework of the national curriculum criteria. They saw no need for formal testing, but kept specific pieces of work that they felt showed development.

Thus teachers' assessment practice at Key Stage 2 varied on a number of dimensions: systematicity, integration with teaching, focus (individual, class, or curriculum), approach (informal qualitative or formal quantitative) and rationale (formative or summative). As at Key Stage 1 there seemed to be a link between teachers' beliefs and their practice.

Conclusions: Issues and Problems in Formative Assessment

Out of the two different experiences of developing a formative assessment scheme at secondary level and of evaluating the national implementation of teacher assessment at primary level, there are some common patterns:

- There is a significant minority of teachers, whatever the age of the students, who are strongly committed to a formative assessment model of operating, in which they are continually collecting information about individuals to assist them in day-to day planning of work. This seems to be founded on a constructivist ideology, although that may not be explicitly held. These teachers appear to select this way of working whatever the educational climate.
- The availability of an assessment framework, provided it is appropriate, encourages these and other teachers to work systematically rather than intuitively. It can help to raise teacher professionalism, since teachers are forced to master it in order to make accountable and comparable judgements.
- Teachers may need to start operating a formative assessment system in a formal way, but as they internalise the criteria, they may move towards a more integrated pattern of teaching and assessing.
- More than half of all teachers put their emphasis on teaching rather than learning, and hence prefer to assess in a summative manner. They are likely to consider formative teacher assessment as an unnecessary workload. Even if they have the information about individual attainment, they find it difficult to act on it. A minority of these teachers find it hard to adapt their accustomed practices in any significant way.
- Involving pupils is not only a way of reducing teacher workload, but also increases student motivation and can help to raise standards by combatting low teacher expectations.
- Formative assessment is attractive to politicians as long as it is not used as a basis for summative judgements. It is perceived to be unreliable in a summative role, although where teachers are properly trained and moderation occurs, evidence suggests that teacher assessment is at least as reliable as, in addition to being more valid than, external tests.

- Nevertheless, formative assessment will become more attractive to teachers only when it can be used as the major basis for summative reporting.

Experience of formative assessment in England is thus both optimistic and pessimistic; it will be important to judge whether similar findings emerge from other cultures.

REFERENCES

Bennett, N., Desforges, C., Cockburn, A., & Wilkinson, B. (1984). *The quality of pupil learning experiences.* London: Lawrence Erlbaum.

Black, P. (1993). Formative and summative assessment by teachers. *Studies in Science Education, 21,* 49–97.

Broadfoot, P. (1995). Performance assessment in perspective: International trends and current English experience. In H. Torrance (Ed.), *Evaluating authentic assessment.* Buckingham: Open University Press.

Brown, M. (1993). Clashing epistemologies: The battle for control of the national curriculum and its assessment. *Teaching Mathematics and Its Applications, 12*(3), 97–112.

Close, G., & Brown, M. (1989). *Graduated assessment in mathematics: Report of the SSCC study. Part 2.* London: Department of Education and Science.

Department of Education and Science, Committee of Inquiry into the Teaching of Mathematics in Schools (1982). *Mathematics counts* (The Cockcroft Report). London: HMSO.

Department of Education and Science, Task Group on Assessment and Testing (1987). *A report.* London: DES/Welsh Office.

Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment.* London: Falmer Press.

Gipps, C., Brown, M., & McCallum, B. (1995a). *Intuition or evidence? Teachers and national assessment of seven-year-olds.* Buckingham: Open University Press.

Gipps, C., McCallum, B., & Brown, M. (1995b). *Models of teacher assessment among primary school teachers in England.* Paper presented at the BERA/ECER Conference, Bath.

Hart, K. (Ed.). (1981). *Children's understanding of mathematics 11–16.* London: John Murray.

Hart, K., Brown, M., Kerslake, D., Kuchemann, D.E., & Ruddock, G. (1985). *Chelsea Diagnostic Tests.* Windsor: NFER-Nelson.

Hoge, R.D., & Caladarci, T. (1989). Teacher-based judgements of academic achievement: A review of literature, *Review of Educational Research, 59*(3), 297–313.

Masters, G.N. (1994). Profile reporting and assessment. *Curricular Perspectives, 14*(1), 48–52.

Mathematical Sciences Education Board, National Research Council (1993). *Measuring what counts: A conceptual guide for mathematics assessment.* Washington, DC: National Academy Press.

New Standards (1995). *Performance standards, Vols. 1, 2, 3.* Pittsburgh: National Center on Education and the Economy.

National Council of Teachers of Mathematics (1993). *Assessment standards for school mathematics.* Reston, VA: NCTM.

Resnick, L.B., & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford & M.C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37–75). Boston, MA: Kluwer Academic Publishers.

School Curriculum and Assessment Authority (1995). Consistency in teacher assessment: Exemplification of standards—Key Stage 3, Levels 4 to 8. London: SCAA Publications.

Wiliam, D. (1992). Special needs and the distribution of attainment in the national curriculum. *British Journal of Educational Psychology, 62*, 397–403.

CHAPTER 4

FORMATIVE ASSESSMENT FOR LEARNING: NEW GOALS, NEW MODELS

EVA BAKER, CRESST, University of California, Los Angeles

Teacher Use of Assessment in the Classroom

Teachers have always made judgments about the accomplishments of their students as part of their regular work. These judgments have been made based on methods as diverse as student recitation, review of homework, evaluation of classroom discussion, observation of behavior, test performance, and analyses of student projects. Growing over the last ten or so years has been concern about how teachers make these judgments about their students' competencies. The issues underlying the renewed emphasis on teacher use of assessment in the classroom relate both to broad policy concerns and to the technical and professional development needed to strengthen teachers' efforts in this area.

In the policy arena, a key impetus for concern about teachers' use of test information is the prevailing belief about the competitiveness of U.S. students in comparative, international studies of achievement. U.S. students are not performing well in these comparisons. On the national level, studies of student performance in the National Assessment of Educational Progress report that most students are failing to achieve proficiency. Similarly, when the Scholastic Achievement Test (SAT) used for college admissions by a large number of postsecondary institutions needed to be "recentered" or adjusted to reflect lower averages, another question was raised about student performance. In state assessments created to measure curricular goals, most states also report that many students are achieving below expectation. Colleges and universities enroll ever-increasing numbers of underprepared students who have met admission standards. Business and industry leaders complain that they must mount expensive training programs to prepare

new hires for entry-level positions. There are numerous counter-examples that argue that today's students are performing as well as ever and that expectations have risen just as the students we teach have grown more diverse. Nonetheless, the consistency of information and of beliefs about inadequate student achievement from a wide variety of sources has given rise to educational reforms on the local, state, and national level.

One obvious question is whether teachers have similarly detected poor performance in their own classrooms. Teachers' summative judgments, corresponding to external tests such as state assessments, are encapsulated in the grades they give students. Judging from general reports of grade inflation, the answer seems to be that inadequate student performance is not reflected in teachers' grading practices. Without getting sidetracked by a discussion of the utility of grades and their motivational effect, or of alternatives to conventional grading practices, let us turn to the practical consequences of external testing information that conflicts with summative teacher judgment. One direct response of policy makers has been to seek to assist teachers to be more effective in the judgments they make about students. One strategy is to improve their use of tests and of test results.

There are three additional reasons that teacher testing practices are of interest—fairness, effectiveness, and efficiency. In the case of fairness, there are questions about the meaning and consistency of grades from classroom to classroom and from school to school. When the population of students was similar from school to school, no one was much concerned about variations among stricter or more lenient grading teachers; any differences were expected to average out. Now that many urban schools deal almost exclusively with disadvantaged students, it is important to document that grading standards are fair, with no particular group of students getting special advantage. This assumption of fairness is needed to assure the meaningfulness of high school diplomas and when comparative judgments are made based on students' grades, for example, in college admissions procedures.

The effectiveness and efficiency arguments relating to teacher testing practices have far less to do with grading and what is called summative evaluation and much more to do with helping all students attain their maximum level of performance. Formative evaluation is the technical term that describes the use

of test results to improve teaching practice and student learning. The purpose of such tests—for instance, those given monthly—is to identify areas needing additional attention or effort as well as to provide evidence of accomplishments. The use of test results in this case is intended to improve the learning of students who were tested by involving (a) students in reflecting on their work, (b) parents for appropriate assistance, and (c) teachers who would need to undertake additional approaches to help students reach academic goals. Considering instructional improvement in longer cycles, it is reasonable for an instructional team to look at the performance of 4th-grade students in 1996-97 and take into account their areas of weaknesses when planning for the 1997-98 school year. This type of formative evaluation provides information to the teaching staff. At the core of these uses is suspicion that teachers may not see student performance as malleable, and may attribute all poor performance to factors outside the instructional setting. Consequently, teachers might be more likely to accept results as the best that could have been achieved—instead of analyzing findings and making changes in instructional practices that might contribute to some improvement.

These analyses are not presented as an assertion of their truth but rather to offer an explanation for the attention given to test data in general and teachers' use of test results in particular. Were U.S. students performing well in international comparisons and other external test results, were all teachers highly respected as professionals, such discussions would be markedly less likely.

Improving the Use of Classroom Tests

To improve classroom use of tests, five basic properties of the assessment should be considered: (a) The assessment must be valid; (b) it must be fair; (c) it must be credible; (d) it must be practical; (e) and it must generate useful results. Let's briefly consider each of these properties.

Validity is a concept that means the test measures the aspect (or construct) of student performance of interest. For example, if a good writer can get a high score on a history essay knowing very little history, the test would not be a valid measure of the student's understanding of history. If a student who can do complex mathematics fails a math test because the word

problems are written in an unfamiliar language, the test would not be a valid measure of mathematics for that student. Validity also has to do with test content, particularly if the test includes or excludes content that it would be expected to measure. For example, a test involving literary devices that excluded similes would have its validity challenged by English teachers who believe that similes represent an essential component of that content area. Validity also has to do with the purpose of the test. Certain tests may be very good for identifying the best or worst students in a class but not very good for providing diagnostic information. When tests are discussed or proposed, it is always a good idea to raise a question about how the validity of the test has been documented, for what students, and for what particular purposes or uses.

Fairness, our second major element, undergirds much of American values and is an extremely important part of testing practice. Most simply, the precept of fairness means that students should receive examination scores that reflect their particular level of expertise and are free of influences based on group membership, such as gender, language group, or cultural background. Of course, the concept of fairness also implies that the "rules" for preparing, administering, and scoring a test were followed and that scores were not influenced by cheating, inappropriate practice, or unacceptable hints, for instance. Scores should not be influenced by students' background, their differential familiarity with the method of testing, or the use of scoring practices that favor one type of student over another. Because fairness is so important, it is one motivating force for the development of tests that can be scored in an objective way.

A third important attribute of tests is their credibility to relevant parties. Credibility means the extent to which a test is perceived to be worthwhile and that its results are trusted. If tests are not credible to teachers, they will not be administered, or if mandated, their results will not be taken seriously. If tests have low credibility with students, many may not try hard and their results will not be good measures of the real level of their accomplishments. If tests are not credible to parents and the public, their results will be dismissed as meaningless, and efforts will be made to change tests to be more in line with public expectations.

Practicality of a test describes the ease of use in reasonable classroom settings. If a test is discretionary and not practical, it

will not be used at all or not be used for long. If a test is required and not practical, in that it takes too much time, requires special materials not easily available or managed, and so on, unforeseen adjustments in administration procedures might be made in classrooms that could well invalidate the validity and fairness of the results.

Finally, classroom tests need to generate results that are useful. Tests that give teachers an overall estimate—that students do well or poorly—are less desirable than tests that give guidance about specific topics or skills that need improvement. But more detail is not always better. The level of detail must match the ability of the teacher to make use of the information. For instance, teachers are unlikely to be able to make use of highly refined test results in situations where limited instructional materials are available. For teachers to make very detailed test results useful in planning for each student in a class, there must be deep resources in teachers' personal repertoires or school curriculum and teaching assets.

Two developments that are part of the current educational reform movement may help teachers to use student results in more effective ways. The first is the development in almost every state of "standards" to guide instruction and testing. One kind of standards focuses on the identification of important goals in content to be achieved at various grade levels or age ranges. Some states and districts have identified standards to be met for every grade, whereas others have chosen important points in student development (for example, at the end of primary education, at the end of elementary school, at the beginning of high school, and at graduation). In most cases, these standards bear remarkable resemblance to what used to be called curriculum goals, curriculum guidelines, or frameworks and are intended to describe what students should be expected to accomplish. They are often phrased in a general way—"students should apply linear and geometric measurement principles to real life problems"—and then may be augmented by more explicit content specifications, for instance, types of included polygons, to provide further guidance. For the most part, states and districts have reviewed model content standards prepared by national groups focused in particular subject matters, such as those of the National Council of Teachers of Mathematics or the National Science Foundation. In a second part of this standards reform, "performance standards," content standards are made

more explicit either by describing a type of task students would be expected to do (estimate distances on a baseball field through geometric methods) or by describing the expectations for performance (answers should be expressed in x units and procedures used should be explained in enough detail that another student could complete them). These standards are intended to provide guidance for determining both the kinds of learning students should be encouraged to experience and the types of examinations that should be given to students. In the purest form, standards-based assessments would provide a coherent framework for teaching and learning. What is particularly interesting about this cycle of educational reform is its emphasis on high standards for all students. Not since the Sputnik era, where once before international competition was a major impetus for reform, have experts in subject matter provided what are called "challenging" goals for students and educational systems.

A second important component of this round of reform is the emphasis on assessments that map to standards. Many of the standards require students to complete tasks that involve multiple days, multiple steps, and collaboration. The type of assessment most appropriate to measure many of these more complex standards is performance-based assessment. Performance-based assessment is an important initiative because in addition to assessing statements of standards, it is based on the idea that assessment must be like student learning. This simple idea turns on its head the more familiar idea of "alignment," that student learning needs to match the methods used in testing. Because of its deep dependence on learning as psychologists and researchers have come to understand it, performance assessments have certain characteristics. Students are expected to *construct* their answers, because research in learning suggests such constructions are the way students acquire meaning. Students may be asked to perform tasks that have multiple steps, acquiring knowledge or determining the next operation based on the results of a prior activity. Students may occasionally work in groups because collaboration has been shown to be an important approach for many learners. The consequences of these attributes are performance assessments that take considerable time (no more two-minute test items), that require judgments or raters to determine level of performance (no more answer key), and that use language as the basis for explaining how and why

problems were solved rather than simply inspecting solutions. The time these performance assessments take inevitably limits the number than can be administered in any one subject area, in interdisciplinary topics, or to any one student. The cost of ratings by judges also limits the number of performance assessments that can be administered. Consequently, it is not feasible for the educational system to assess the attainment of all of the standards it seeks to achieve. So it will be expected that teachers will use their own classroom examinations or other methods to assess many of the standards.

How Can Teachers Use Standards-Based Performance Assessments in Their Own Classrooms?

Ideally, professional development will help teachers organize the way they go about their assessment job, and research has shown that it takes considerable time, energy, and knowledge of subject matter to do a good job with performance assessments. A simple set of guidelines may help people get underway. First, determine which standards are being measured by external means, through the use of district or state assessments, commercially available tests, and so on. Seek to acquire whatever you can that specifies the content to be assessed (expanded content standards) or the type of measures and student expectations in the assessments to be administered (performance standards). One source of classroom examinations should be based on components or comparable measures that will be used to evaluate externally the students and the school. For example, if students are expected to use particular procedures in solving mathematics problems, such as a number line, you should develop a classroom assessment that uses the same general procedures. Second, determine which standards have been articulated that are impractical for the education system to measure formally. Progress toward these standards can be examined by using your own personal examination procedures. There are numerous models available for the design of performance assessments. One we have used in a variety of districts and states at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) creates specifications for each type of learning to be measured and then uses the same general approach in different subject matters. The schematic below shows the five key types of learning CRESST

has identified (Figure 1). Most goals (and relevant assessments) in schools are made up of a combination of these learning types, in the same way as an individuals DNA is made up of different combinations of genes.

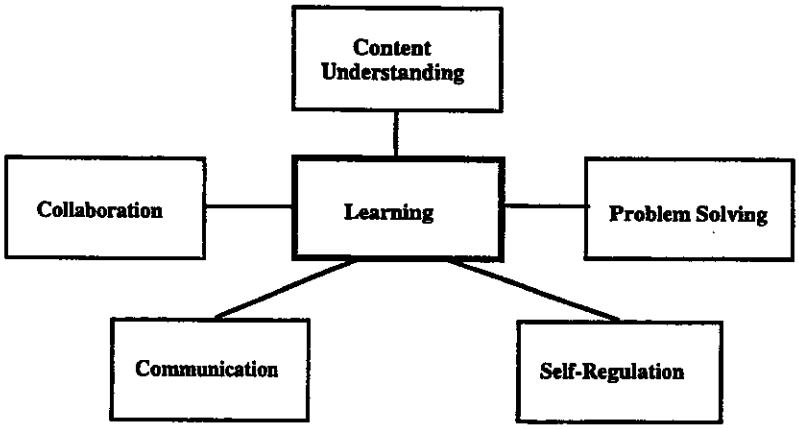

Figure 1. CRESST Model of Learning

For each of these areas, at least one set of specifications for guiding performance assessments has been developed. These specifications include the types of materials provided to students, the administration procedures, and the scoring criteria or rubric. For example, in the area of problem solving, the specifications provide a number of choices for the assessment team. They might choose to focus on problems that have multiple right answers or single right answers. The goals of interest might emphasize problems that are clearly formulated. As a comparison, the focus of the assessment could be problems that are complicated and need further clarification. Whether students are provided with all the information necessary to solve the problem, are expected to have already acquired the information, or are provided opportunity to seek relevant information (for instance, in their books, the library, or on the Internet) are other choices. Performance in problem solving can use criteria related to problem identification, strategy selection, use of prior knowledge, and fluency. It should be clear that such

components in scoring criteria not only encompass important attributes of problem solving (documented in the research literature) but provide areas that teachers can address in both initial and any continuing instruction to assist students in acquiring competence. What is most important about the problem-solving model as it has been formulated is that it can be applied to different subject matter areas, for instance, social studies, mathematics, and science. So a teacher or assessment team need only to learn well the specifications (or create their own) and apply them to multiple subjects taught in the classroom or school. Specifications have been widely used in the area of content understanding as well, where students are presented with source materials that include relevant content information. Primary source materials are used in history and geography—speeches, letters, films, and maps; in language arts, literature or literary analyses or audio tapes can be used; in science, descriptions, write-ups, graphs, or videos of experiments may be provided. In mathematics, problems are provided and alternative strategies or solutions may be given. The students then prepare analyses of the source materials. So far, two major approaches have been used. One involves the preparation of a written explanation where the student explains to a friend or relative the key principles involved and explains why the material is important. These explanations are scored using a well-researched approach involving an overall judgment of content understanding, and the student's use of principles, prior knowledge, and argument appropriate to the subject matter. These essays are often scored for English writing conventions as well. Another set of specifications related to content understanding uses the same sort of stimulus materials, but permits students either individually or in teams to create graphical depictions of the relationship among key principles and ideas in the form of a concept map. These maps are scored in terms of their fidelity to maps made by an expert or teacher. One recent breakthrough is our ability to score these maps automatically using computer support. Specifications for performance assessments also exist in the areas of collaboration, communication, and self-regulation. The benefit of using this type of approach is its general applicability to multiple subject matters and the relief it provides from reinventing assessment approaches time after time. There is also some evidence that teaching towards one assessment (for instance in a different

topic in history) helps students apply the appropriate analytical process to an assessment on a different topic.

Without doubt, creating one's own performance assessments is a task involving great commitment. But to do so where standards-based reform is underway provides a unique opportunity to integrate two aspects of assessment that have been largely separated from each other: external or accountability-focused assessment, and teacher-designed classroom assessment intended to provide feedback and improve learning. If both sorts of assessment can be generated from conceptions of learning that also guide instruction, there is real opportunity to improve the performance of our students in a coherent, sensible way. Integrity of the educational system will ultimately depend upon the coincidence of goals, learning, teaching and assessment. Approaches that attract these elements rather than set them in opposition to one another are essential to the success of our schools.

CHAPTER 5

ASSESSMENT: A LEVER FOR CHANGE IN THE EDUCATION SYSTEM

CLAUDINE PERETTI, Department for Assessment and Forecasting, Ministry of Education, Higher Education and Research, France

In the middle of the eighties, in order to meet the expectations of the nation, the French education system was given the target of bringing 80 percent of each age group up to the Baccalaureate level by the year 2000.

Attaining this aim required an increase in educational performance and, as a result, the creation of a steering structure within the French Ministry of Education. To meet this need, a department for assessment and forecasting, the DEP (*la direction de l'évaluation et de la prospective*), was created in 1987 and commissioned to develop external assessments of the education system to give an account of its effectiveness.

The Education Act published in July 1989, which laid as a principle permanent global evaluation within the education system, has extended the DEP mission to include the dissemination of assessment tools for those involved in the education system.

Thus it is now time to ask how and where the development of external assessments and the dissemination of a new assessment culture were instrumental in changing the system and improving its results.

How Has Assessment in the Education System Developed?

The DEP brief is twofold: First, to organise external assessments in order to measure the results of the education system and to take account of its effectiveness; second, to develop a new assessment culture within the system by giving diagnostic and formative assessment tools to teachers and education staff.

For the first of these, the measurement of the effectiveness of the education system, four approaches of increasing complexity have been selected.

The first approach is the assessment of pupil's competences and knowledge. In France, the knowledge and skills expected of primary and secondary school pupils are defined at the national level in the aims and curricula. As a result, the DEP conducts assessments of pupils' attainment, which take place within three different frameworks:

- Cohort studies of pupils set up periodically to track and understand school careers;
- Periodic sampling at particular levels of education, for example at the end of lower secondary school, to establish whether training objectives have been fulfilled;
- Mass testing organised at entry to particular classes to help teachers measure the difficulties and gaps in learning of their pupils.

In order to enrich these measures, the DEP often builds into its tests either a longitudinal dimension which seeks to measure the development of pupils' skills and knowledge over time or an international comparison with one or more countries so that the results of French pupils can be placed in a wider context than the purely national.

The second approach is the evaluation of education policies used to promote a more favourable global learning environment and to help specific groups of pupils to overcome particular difficulties. The aim of these assessments is to measure not only the agents' satisfaction but also the effects of policies on pupils' achievements.

The third approach is the evaluation of the agents in the system, namely those responsible for implementing education and training policies; the main action here is a research study of the effects of the agents' characteristics and practices on pupils' attainments.

The fourth approach is the evaluation of teaching units, with the specific aim of enhancing our understanding of the impact of organizational constraints in schools and of the interaction of the people in them on pupils' learning and achievements. The

concept of added value of schools underlies the whole work in this area.

For the second part of the brief, the requirement to develop a new assessment culture within the education system, the DEP has undertaken to distribute diagnostic and formative assessment tools.

The tests the teachers are accustomed to using are summative assessments. Their aim is to check whether pupils have the knowledge and skills they are supposed to have acquired. Tests of this type are not pedagogical tools because they are not involved in a learning process. For this reason, the DEP introduced a national compulsory assessment of pupils' skills in 1989 at the beginning of the CE2 (age 8) and at sixth, the start of secondary schooling (age 11); and, in 1992, at second the beginning of the first year of upper secondary school (age 15).

The aim of these annual assessments is, for each key moment of schooling, to help teachers measure their pupils' difficulties and the gaps in their learning and to plan remedial action. The fact that they are systematic has helped encourage teachers to make use of diagnostic assessment during the course of the school year.

In the same way, DEP produces and disseminates assessment tools—diagnostic or formative—for use by teachers at will. Currently these tools exist for primary grades for every branch of learning except arts and physical education. For lower secondary school, they exist for mathematics and French language. For upper secondary school in the general and technical branch, these tools exist for mathematics, French, foreign languages (English and German), history and geography; and in the vocational branch, for mathematics, French, economics and management or industrial subjects.

In another area, DEP disseminated an indicators system to all secondary schools in 1995. The objective is to provide schools with tools for working out a school project and managing it. This system proposes some twenty standard indicators common to all secondary schools, enabling them to calculate their own specific values of these indicators and thereby to assess their own performance rating; to take stock of the way they are run, their available resources and the various constraints which they have to take into account.

What is the Influence of Evaluation within the Education System?

There are clear signs that evaluation is beginning to contribute to the regulation of the education system at both micro and macro levels and, consequently, to the improvement of its results.

At Classroom Level

A 1993 study of the use of national assessments at the beginning of CE2, sixth, and second and of the tools provided to primary school teachers showed that three-quarters of teachers think that these assessments are useful even if half the teachers emphasise the extra work that they require. At each level, teachers recognize that diagnostic assessment provides a better means for identifying pupils' difficulties than traditional summative tests. Half the primary school teachers say that they use the diagnostic and formative tools.

At the same time, the dissemination of a new assessment culture contributes to teachers' awareness of the aims of the teaching they have to deliver and of the achievements expected of their pupils. For example, when the 1989 tests at entry to CE2 showed areas of weakness in geometry, primary teachers automatically took the lesson to heart: the following year, in a similar test with comparable items, pupils' attainment in geometry had risen.

At School Level

To an increasing extent, primary and secondary schools are using assessment results to ensure that their projects or pedagogical actions are appropriate to pupils' difficulties, such as during extra-curricular activities.

Although it is too early to measure the use of the management indicators disseminated in secondary schools, it is already perceptible that these new tools arouse a great interest. The dissemination of these tools is helping schools give meaning to their autonomy.

At Regional and District Level

The systematic and standardized assessments at the beginning of CE2, sixth, and second also help the district managers to take account of pupils' and schools' heterogeneity and, as a result, to plan a local educational policy.

For instance, the results of these assessments are used to allocate more financial and human resources to the schools and pupils with the poorest results. This is a contribution to equity within the education system. In addition, the assessment results and the management indicators are increasingly used by the inspectors for diversifying and improving their evaluations of schools.

At National Level

The results of the external assessments help policymakers measure the effectiveness of implemented educational policies in improving pupils' success. In this way, they influence the decision-making process. At the same time, the results of these assessments help regulate and rationalize public debate about school effectiveness.

What is the Way Forward?

Even if there are real benefits for the education system in the development of external and internal evaluation, the generally positive verdict should not mask the limitations of present research and studies.

The first limitation is that evaluation is not exhaustive; it does not cover every area and, within the areas that it does cover, it leaves some fields unexplored. The processes within the classroom, the teacher–pupil relationship and the learning process particularly, are insufficiently explored. Furthermore, too little use is being made of the results of research to improve teachers' practices.

The second limitation is that evaluation cannot be the only regulator within the education system. Everyone knows that the decision-making process, both at the Ministry level and the classroom level, is the outcome of a large number of parameters in which rationality and neutral knowledge are not always outstanding—nor is it necessarily the case that they should be.

The third limitation is that evaluation will be truly effective when its procedures and its results are used at every level of the system. For the moment, this aim is far from being attained. Many teachers, and many inspectors, are not yet convinced of the value of these new approaches and they prefer to rely on the traditional summative tests. It is true that not every teacher has been trained to use these new tools and some of them find it difficult to adopt an informed and critical attitude towards their use so that they have no interest in using them.

To conclude, if we want every component of the education system to gain real benefit from assessment results and tools, it is indispensable, on the one hand, to develop initial and in-service teacher training and, on the other hand, to encourage the inspectorate to promote increased awareness of, and trust in, evaluation among teachers.

CHAPTER 6

ASSESSMENT IN SCHOOLS: A MATTER OF CHANGE

HANS F. VAN AALST, Organisation for Economic Cooperation and Development (OECD)

I started work on this paper[2] by describing current practices in assessment in Dutch schools, its use, and recent changes. Writing it down, I wondered why the drawbacks of that practice—which have been well known for a long time—have triggered so little change.

Obviously, there are several explanations. One of these is that some suggested changes in assessment do not match certain expectations of education held by the general public and many teachers. So the question arises how certain practices of assessment relate to value-positions about education. I decided to sum up the description of Dutch practice and to explore the meanings of education, the beliefs about what education should and can achieve, and the relations of those meanings to the roles of assessment. This may not be the best approach to the issue. Many innovations start from small practical changes and not from broad concepts. But I believe that many practical attempts at innovative assessment have created confusion or have failed because of a lack of communication with stakeholders about the underlying change of meaning of what education is about.

Repeatedly, international studies have shown that assessment serves more than one function at the same time. Optimisation of practice towards one function may disturb the others. Solutions to that problem have not been found yet. Some advocate different practices for different functions, others try to find practices that may serve several functions at the same time. Although I do not have a solution at hand, I feel that a solution can't be reached by functional reasoning alone but requires

[2]The views expressed in this paper are those of the author. They do not in any way commit the Organisation for Economic Co-operation and Development or its member states.

clearer communication about the meaning of education. That might help to decide which functions are crucial in a certain context and to what extent.

The exercise was really a first try and arrived only at the bare outline of an idea. Nevertheless, it leads to some suggestions about priorities for assessment policies.

The Practice of Assessment: The Dutch Case

Assessment[3] in Dutch schools is a normal part of educational life. In secondary education, there are frequent short tests[4] and, about twice a term, large and more formal ones. Teachers usually start lessons with oral questioning of some of the students. Assessment helps in organising learning time and learning effort. It serves the allocation of students to stages along the different educational pathways.

Allocation of students from primary education to secondary schools is based on the experienced professional judgement made by the primary school, often based on informal methods of

---

[3]For a definition of the term "assessment" I refer to Harlen et al. (1992):

> Assessment in education is the process of gathering, interpreting, recording and using information about pupils' responses to an educational task. At one end of a dimension of formality, the task may be normal classroom work and the process of gathering information would be the teacher reading a pupil's work or listening to what he or she has to say. At the other end of the dimension of formality, the task may be a written, timed examination which is read and marked according to certain rules and regulations...All types of assessment...involve interpretation of a pupil's response against some standard of expectation. This standard may be set by the average performance of a particular section of the population or age group, as in norm-referenced tests. Alternatively,...the assessment may be criterion-referenced. Here the interpretation is in terms of progression in skills, concepts or aspects of personal development which are the objectives of learning...

[4]I use the term "test" for a formal educational task, under constraints of time, meant to judge responses. I do not use the term here in the sense of a measuring instrument, detached from education, as is often the case in the USA. I use the term "examination" to indicate a final test at the end of a period of schooling.

assessment. Sixty percent of schools add the results of a centrally set achievement test to this professional advice.

In secondary schools, marks play a central role. Promotions from one school year to the next are based on marks from formal tests. Final examinations are important in secondary schools. They set guidelines for educational practice and the marks are used for certification. Half of the marks are based on school-based tests. Universities accept the secondary school certificate as adequate for entry.

Changes

Centrally Provided Tests for 200 Attainment Targets

The intention of recent policies is to raise the level of basic education for all. In 1993, the government set more than 200 attainment targets which have to be met by all students after three years of lower secondary education. All schools should use centrally provided tests, which are supposed to cover all the attainment targets. The first experiences seem rather chaotic. Main points of concern are whether the data used for certification of individual students should also serve—by some means of aggregation—to evaluate schools and monitor quality at the national level, and whether the tests are appropriate for weaker students. Other concerns include the workload for teachers and the costs and administrative burden for schools. For all these reasons, revision of the initiative is currently envisaged.[5]

Examinations in Upper Secondary

Changes in final examinations are envisaged for both lower and upper secondary schools. For lower secondary education a major revision of examination syllabuses is taking place, together with a new definition of educational streams.

In general upper secondary, the curriculum is changing dramatically. An integral approach has been adopted instead of the current subject-based approach. Students will also have about 20 percent of their time available for a wide variety of activities, including the possibility of work experience. Innovation is geared to more independent learning.

---

[5]An advisory report is expected in April 1996.

The accompanying changes in the final examination have a pragmatic character. They are intended to provide more opportunities for flexibility in learning, and to decrease the amount of teaching time that the examination takes. A school-monitored portfolio will replace the school examination. It is recognised that, in the long run, a form of formative assessment is needed in upper secondary, from which summative assessment may be drawn.

Teacher Networks between Secondary and Tertiary Education

It is beginning to be understood that examination marks have a rather low predictive validity for success in higher education. This is why teachers from secondary schools and from tertiary education have started networks to discuss together and improve their instructional practice. Additional activities for students of both educational sectors are provided as well, in order to facilitate the transition from secondary schools to higher education.

Functions of Assessment

Both teachers and students see assessment in the first place as a means to monitor commitment to learning and judge general ability. Teachers often claim that they need testing as a means to stimulate students to put effort in their studies. Many students work for the test in relation to their perception (and even calculation) of chances to pass at the end of the year. Both teachers and students do accept low marks on tests, as long as they do not endanger a promotion to a further stage. Marks on tests are central. Marks are used for decisions about passing from one school year to another, about the selection of subjects for examinations, and for certification. At the system level assessment serves a specific selection function: differentiation between students on a general scale that reflects a combination of ability and adjustment to certain socially valued behaviours. This "screening-function" of education is currently discussed in relation to a more allocative model of selection: the "qualification-function."

The government and the general public value examinations as a means to guarantee quality and standardisation. Examinations in the Netherlands serve also as a means to consolidate certain

innovations in the curriculum. There is a tendency to compensate for the negative backwash effects of central written examinations by setting guidelines for the school examination.

Assessment operates also as a ritual. Questions asked by the teacher often mark the beginning of instruction, the formal test at the end of a period of work marks the moment to conclude a subject, to leave it behind. And assessment is the expression of power of the teacher.

If one considers the basic instrumental functions of assessment as identified by Harlen et al. (1993)—

- the formative role: as a means for providing feedback to teachers and pupils about ongoing progress in learning;
- the summative role: as a means to summarise the nature and level of students' achievements at various points in their schooling and when they leave;
- the certification role: as a means of summarising for the purposes of selection and qualification;
- the quality control role: as a provision of information used in judging the effectiveness of schools and of the system as a whole—

then one can conclude that the summative and certification roles steal the show in the Netherlands. The summative role mainly serves as a ritual, the certification role a specific interpretation of selection: screening.[6] The formative role is not systematically practised. The quality control role is exercised only at the national level and is focused on standardisation. The emphasis in quality control is currently shifting to the school level and to more explicit means.

Reflections About Change

Drawbacks of Current Practice

The current practice of assessment not only serves the positive, desired roles: summing up of achievements,

[6]As I shall explain later, this is related to the meritocratic function of education in the Netherlands.

certification, and quality control. It has also serious drawbacks. In summary, the present practise of assessment has the following effects:

- It creates failure: Currently used tests and the way they are marked produce cumulative failure; within each group a 30 percent failure rate occurs again and again. Students tend to get indications of what was wrong instead of praise for what was good, which does not help to develop self-esteem. All this reinforces bias between social groups and between boys and girls in terms of educational achievement.
- It is inefficient in terms of allocation[7]: It is well documented that marks have low predictive validity for higher education and employment. The most that an examination might do is to provide a reliable average for a school group. Pass/fail decisions on the basis of marks are often wrong; many students who are not allowed to pass would have been able to do so, while many others who are allowed to pass, do fail; this is—ironically enough—especially true for students who have medial scores. Students do not learn much from current assessment about their potential for a wide range of possible future tasks.
- It discourages higher order learning processes and innovative and entrepreneurial talent: Assessment often creates a culture that values the right answer and discourages risk-taking, expression of doubts, and social interactions.
- It creates a race between students for ever higher educational qualifications, while the added value of such investments decreases where a long time is being spend only in initial formal education.

The first category of these drawbacks is currently explicitly attacked, at the level of lower secondary. As we have seen, the first experiences are chaotic. This is not only the case in the

[7]The term "allocation" is used to refer to a process of selection by which the emphasis is on finding a place inside or outside education where a person may develop further as well as possible.

Netherlands (OECD, 1995, p. 46). The other points are not yet on the agenda! One may wonder why this is so, given the priority of many governments for more effective education.

Apparently, current assessment practice is quite resistant to change.

The Stability of the Current Practice

There are several reasons for this.

First of all, current practice is embedded in history and culture and any suggested change finds itself in a defending position against it. Parents have survived it, or even profited from it. It serves teachers as a semi-objective tool to deal with complex and difficult decisions in education and the accountability of those decisions to parents. This has been called the "rational authority" of marks (Hoskin, 1982).

Secondly, the drawbacks are felt to be compensated by other means. The system compensates for the extreme effects of the drawbacks by allowing variation on other determinants of the system such as alternative pathways and repeating of school years. Schools and teachers compensate for the sharp edges by allowing special solutions for special cases.

Thirdly, teachers who are inclined to change assessment are discouraged by a lack of practical support and instrumental expertise. National testing programmes look towards standardisation and application across schools, and do not provide information about what to do if results do not meet standards. At the same time, a claim is being made for a formative role for assessment, aiming to influence the individual learner directly. These "double" messages confuse teachers, and not only them.[8] Teachers find the information gained from tests difficult to use for feedback and remedial treatment. Test results are also felt to be rapidly outdated. Teachers also mention serious organisational difficulties if they want to respond to children's results in a differentiated way.

Last, but certainly not least, such pragmatic and practical reasons are embedded in a culture where a specific meaning is

[8]Several authors think that both roles cannot be served by the same set of instruments (OECD, 1993; Harlen et al., 1992; Resnick & Resnick, 1992), others hope for special conditions under which that is nevertheless possible (Black, 1993b).

attached to education: People believe that educational achievement depends on the individual's merit, that it reflects individual commitment and ability. One expects education to reveal differences in individual merit. One is suspicious if education claims to achieve equal results for all, and is reluctant to invest in it, if it encompasses more than very basic skills.

Any change suggested has to deal with each of these forces and, indeed, with their interactions. This sets a very complex agenda for action. In this paper I will concentrate on the question of how different value positions may relate to different options of assessment.

Meanings of Education

A culture carries sets of socially established meanings. What are the meanings constructed around education, what are the underlying values and how are they reflected in assessment practices? Three sets of meanings may be relevant here: "education as supply," "education for results," and "education for qualification."[9]

Education as Supply

Description. "Education as supply" is a vision where the primary task of educational service is seen as the supply of materials, teachers, classrooms, and other conditions. In this view, a rich and varied educational environment serves the personal development of students. Debate concentrates on the question what "good" supply is. Supply is either equal for all students, or it is different for different groups of students, according to certain entry characteristics of students. Traditionally, results on tests play a role here.

The achievement of results is certainly seen as related to the quality of supply. But, whether students actually achieve results is considered to be dependent on their individual capabilities, effort, and perseverance ("merit"); it is their individual responsibility and/or that of their parents.

---

[9]In an earlier publication I introduced a fourth meaning: "education for effect." I do not deal with it in this paper, because the consequences for assessment are not yet apparent.

Educational policies. Educational policies aim to ensure "good" supply, guarantee accessibility, and control attendance. Whether students acquire results is not regulated, but summative assessment is.

Assessment. Assessment could be basically summative and could very well lead to a profiled summary of achievements.

In practice, supply-based systems tend to assess achievement in general terms of "school success" or "level." This is, on the one hand, an artefact of specific psychological theories (Gifford & O'Connor, 1992). On the other hand, it suits the generally held view about individual merit. Together, these factors match the value that success in society should not be derived from religious or societal positions, but should only depend on personal merit. It is the educational system that is expected to show and measure such merit (meritocracy). This position reflects a stake of educational institutions.

Related to this value is a specific meaning of selection: the "screening" or "credentials" theory. The underlying idea is that the primary function of education is not to enhance individual abilities, but to reveal potential—which individuals carry basic abilities already—and achievement of a certain level of education is an indicator of perseverance, adaptability, and learning competency. This explains why assessment for certification is dominant in supply-based systems and why it is based on unidimensional scales, referred to as "levels."

Status. I believe that "education as supply" has been the dominant position of educational policies for many decades, and still is the dominant position of the general public and many teachers.

Education for Results

Description. "Education for results" reflects a political stake and a mission of some educators. The basic service provided by education is no longer "good" supply but results for all students. In the beginning of this movement, the term "results" referred to a set of "minimum results for all"; later "the best possible results for all" is added to its meaning. The debate is about what such minimum results or such best possible results for all should be.

The position is related to concerns of educators about the systematic bias in educational results, both in the case of common supply and in the case of differentiated supply, leading

to systematic disadvantage of low socio-economic groups and girls. The findings of research that schools differ in their effectiveness to accomplish results for their students have contributed to the new mission as well.

On the political side, "education for results" is connected to a general need for governments to look for cost-effectiveness and the belief that decentralisation and demand for accountability would serve that purpose.

Educational policies. Educational policies focus on defining "attainment targets," make the targets obligatory for schools by regulation, and allow more autonomy of schools in terms of educational provisions (supply): curriculum, teachers, buildings, etc. Tests are provided to control whether the desired results are achieved.

The school is accountable to parents and the government. Teachers lose autonomy and are bound to work as a team towards the school's objectives. School management is an object of primary concern, more than teachers' professionalism, which is left to the school to foster.

Assessment. Assessment in schools is far more complex than in the case of education as supply. It is expected to fulfill more roles: First, the formative role of getting continuing evidence about students' progress and involvement, in order to improve those; second, the role of summing up whether and to what extent the required results have been achieved; third, profiled certification based on differences in specific domains, not on the basis of a position on one scale; and last, assessment for accountability of schools seems dominant, whether by inspection or by indicators aggregated from summative assessment.

Status. "Education for results" is the dominant position taken by most national educational policies since the early eighties (Husen & Tuijnman, 1994).

Difficulties and limitations of "education for results" (I). In spite of the political rhetoric and pressures on schools to raise standards, education for results is far from being a reality.

Reasons for this include the following:

A lack of instructional know-how: Concrete know-how about learning by students of specific subject matter is not readily available. To the extent that it is available, it shows that learning is much more complex than a model based upon simple progression through certain levels would suggest. Tacit

knowledge of practising teachers is not systematically used for improvement[10].

Organisation, management, and facilities of schools do not allow for a highly differentiated instructional approach. A change from a supply practice to a results vision needs a real "re-engineering" of processes in schools and this is often not addressed[11].

The emphasis on the school as an evaluation unit does not match teachers' sense of professionalism. Teachers feel under-valued. This, together with the lack of concrete support and expertise, makes them suspicious, defensive, and even cynical in relation to the proposed change, while for successful change, teachers should understand, be willing to work towards an innovation, and feel ownership.

External pressures for accountability distract attention from primary instructional processes. This is indeed a paradox: The attention given to summative assessment and accountability in order to improve results tends to push formative assessment and differentiated approaches to learning to the margin and leave it under-resourced, whereas better results inevitably require enhanced investment in primary processes.

Difficulties and limitations of "education for results" (II). Of a different nature is the following set of reasons. Here, doubts are expressed about the basic values of "education for results" and its rigidity of use:

The general public still adheres to a supply vision, and is suspicious about equal results. Fishkin once wrote about the "trilemma" among the values of equal chances, individual merit, and the autonomy of the family. His analysis revealed that it is indeed impossible to do justice to each of these values at the same time (Fishkin, 1983). Recruiting processes do reinforce the "screening" function of education by using achieved levels of

---

[10]There are of course exceptions, such as the Dutch work on mathematics, some English work on science, and the Australian PEEL project.

[11]Innovation in schools, which involve major and lasting transformations, demands special arrangements, which are different from managed change as applied in some transformations in business or industry, because of little means available for changing personnel. Karen Seashore Louis (1994) gives some clues about the special challenges.

education as a first order selection criterion and, as we have seen, such levels do not refer to specific domains of content.

Others are suspicious about the targets set within the educational community. They fear that such targets are too abstract, too far away from human commitments in daily life, or inadequately selected. These suspicions refer to inherent problems of the "education for results" approach: The question whether it makes sense at all to base education on defined desired performance within the context of schools and only at a given moment, and whether it is possible to define such results in a way which carries meaning for the stakeholders on the one side, and are relevant for educational improvement on the other.

The conclusion is that the political (and educational) desire for "education for results" is being imposed on a system which in practice is matched to "education as supply." This causes confusion, all sorts of unclear compromises, and lack of mutual trust between policymakers and the educational community. The efforts to change that practice have not been well defined, communicated, and resourced.

At the same time, research about the complex processes of transfer from schools to the outside world reveals doubts about the idea that the primary mission of education should be the achievement of well defined results. The quest is for more interactive processes.

Education for Qualification

Description. "Education for qualification"[12] refers to a position where educational elements, such as "good" supply or attainment targets, are directly connected with contexts in the world outside education. It is a response to the lack of validity of current school certificates for purposes of recruitment. Historically, it referred to methods to improve the definition of desired learning outcomes by careful analysis of current

[12]The term "qualification" stands for a definite meaning here, the "quality" a person may possess to perform a (future) task. A person is better qualified if the task is likely to be more efficiently performed. In principle, "qualification" includes personal characteristics such as height, but for education the concept condenses to learnable things: skills, competencies, knowledge. Crucial is the relation to given contexts/tasks. Thus, "qualification" does not refer to certification here.

practices in higher education, in the world of work, and in leisure, and by involving people from outside education in the definition and selection of these desired outcomes. It was found that such methods alone—in fact, a variant of the "education for results" approach—have limited value and need direct interactions on the level of learning processes and assessment to make them effective.[13]

Education for qualification finds its ground in theories of learning on the one hand, and is inspired by the wealth of experience and tacit knowledge about learning in organisational settings (referred to as "the learning organisation," "empowerment," "human resources management," etc.) on the other. It also relates to emerging modes of knowledge production, where the traditional notion of "learn knowledge first and then apply it" has changed to a notion where theory and practice are much more intertwined (Gibbons et al., 1994).

Initial education in formal schools is seen as valuable in some respects—for example, as a place where you are allowed to make mistakes, a place for reflection, and a place where you can think for a certain time without time pressures; or as a place to practice and gain some specific skills; basically "supply" factors—but only within the context of some agreement between the school, the student, and some external agency.

Learning is seen and understood as an active process of construction and as contextual. The assumption that learning implies mastery of basics (skills or concepts) which can later be applied to a varied set of contexts is believed to be valid only in a limited sense. Fixed results determined by central governments are thus not seen as a major indicator for meaningful learning. Even if certain results are established, it is not at all certain whether such results are of much value for later life. Thus, results are seen as important, but only in relation to some sort of external context. This means, in operational terms, that schools share ownership with other parties.

Learning is also seen as dynamic; results change over time. If certain results are not achieved at a certain time, they may be achieved later, and possibly more efficiently. Learning for qualification fits into initiatives for life-long learning.

---

[13]The Dutch networks of teachers of secondary and tertiary education to which I referred earlier illustrate the point.

Educational policies. Educational policies in the context of "education for qualification" focus on communication and co-makership between education and other institutions in society. Diversity in institutional settings and flexible arrangements and conditions for learning are allowed. Teachers' professionalism is expected to cover both learning in schools and other learning processes. Mobility of teachers between schools and other places in the working society is encouraged.

The objects of control and evaluation are the pathways of students through education and work settings, the transparency of learning arrangements for students and local quality control through accreditation by a diversity of interest groups ("multiple accreditation," van Vught & Westerheijden, 1994).

Assessment. Assessment includes local, formative, and dynamic arrangements, as in modern styles of human resources management, including assessment centres. Certification is based on portfolio methods, and may relate to agreed qualification levels. Certificates can be gained through work experience as well as through educational arrangements.

Status. Education for qualification is an emerging vision.

Conclusions and Issues for Discussion

My conclusions and suggestions for discussion are the following:

- Recent educational policies focus upon "results for all" and accountability of schools. This sets an enormous agenda for work, which does not exist in most countries.
- The current emphasis on education for results addresses only one aspect of the drawbacks of the present practise of assessment. Other aspects can only be addressed if educational policies shift from "education for results" towards "education for qualification."
- The current emphasis in educational policies is a gradual shift from "education as supply" towards "education for results." In many respects, it may be recommended to skip the results phase and jump directly from "education as supply" to "education for qualification." This would place initiatives for

"education for results" in a more balanced position and avoid some of the inherent shortcomings and operational problems of the movement for results.

- However, this does not mean that education will be modelled according to the properties given to the qualification vision in this paper. Instead, education in the future will have elements of all three visions. Elements of supply will focus on matters such as higher order learning processes, group learning, creative learning, reflection, etc. The "results approach" will focus on very basic processes only and education for such results will only take a limited amount of time of learning time. Education for qualification will cover the other domains of knowledge and skills. The process of change in such a direction will be slow.
- Schools may differ in what they offer, and may take responsibility for different combinations of the different modes. Some schools will develop towards a sort of educational broker, connecting students needs to learning places and guiding progression.
- If such perspectives make any sense, it helps to set priorities for assessment policies, such as the following:

  formative—formative assessment is always a priority, but it takes different forms in different settings;

  summative—development of innovative tests in fields of higher order styles of learning, or domains of new curricular contents;

  certification—development of profiled summative certificates, geared to allocation and recruitment purposes, in a context of communication between schools and recruiters and arrangements for life-long learning;

  quality-control—development of methods, which fit the distinct missions of schools; a mix of methods is indicated rather than one specific method.

REFERENCES

Black, P.J. (1993a). Formative and summative assessment by teachers. *Studies in Science Education, 21*, 49–97.

Black, P.J. (1993b). Assessment policy and public confidence: Comments on the BERA Policy Task Group's article "Assessment and the improvement of education." *The Curriculum Journal, 4*(3), 421–427.

Fishkin, J. (1983). *Justice, equal opportunities and the family.* New York:.

Gibbons, M. et al. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies.* London: Sage Publications.

Gifford, B.R., & O'Connor, M.C. (Eds.). (1992). *Changing assessments, alternative views of aptitude, achievement and instruction.* Boston: Kluwer Academic Publishers.

Harlen, W., Gipps, C., Broadfoot, P., & Nuttall, D. (1992). Assessment and the improvement of education. *The Curriculum Journal, 3*(3), 215–230.

Hoskin, K. (1982). Examinations and the schooling of science. In R. Macleod (Ed.), *Days of judgement.* Driffield: Nafferton Books.

Husen, T., & Tuijnman, T. (1994). Monitoring standards in education: Why and how it came about. In A.C. Tuijnman & T.N. Postlewaite (Eds.), *Monitoring the standards of education.* Paris: Pergamon.

OECD (1992). *Curriculum reform/assessment in question.* Paris: OECD Documents.

OECD (1995). *Schools under scrutiny.* Paris: OECD.

Seashore Louis, K. (1994). Beyond "managed change": Rethinking how schools improve. *School Effectiveness and School Improvement, 5*(1), 2–24.

van Vught, F.A., & Westerheijden, D.F. (1994). Naar een stelsel van meervoudige accreditering? ARO, Utrecht; Sdu-DOP, Leiden.

CHAPTER 7

SYSTEMIC REFORM: NATIONAL POLICIES ON CURRICULUM AND ASSESSMENT—A US PERSPECTIVE

DAVID STEVENSON, Department of Education, Washington, D.C.

I have focused my remarks on three topics: Why we have turned to standards and assessments as important policy tools in improving the performance of the education system; the promise of high standards and assessments; what technology has to offer?

Why Have We Turned to Standards and Assessments?

One of the reasons for the OECD project on curriculum redefined is the interest across OECD countries in improving the productivity of our education systems. There is widespread dissatisfaction with the performance of the education systems, even if the systems are seen as improving or high performing.

In the United States, there is a general dissatisfaction with the education system. The quality of education, for the first time ever, was American voters' top concern in the last Presidential election. Sixty-seven percent identified it as the top issue affecting their voting, ahead of crime and the economy.

This is the case in spite of the fact that performance has been improving:

- Participation in tougher courses increased—there was an increase in the percentage of students taking academic courses from 13 to 47 percent between 1982 and 1994.
- The average performance in mathematics improved between 1978 and 1992.
- Science achievement was higher in 1992 than 1982.
- The gap in performance between white and minority students has narrowed.

- The dropout rate declined from 14.6 percent in 1972 to 11 percent in 1993.
- Postsecondary education has continued to expand—the percentage of high school graduates immediately enrolling in college has increased from 47 percent in 1973 to 65 percent in 1993.

Given these encouraging trends, why is there such dissatisfaction? There appear to be two main sources of concern.

The first is that demands for "human capital" are greater than ever. There are more positions requiring high levels of cognitive skills. There is a widespread belief that increased efficiencies require more skilled individuals, together with a general dissatisfaction on the part of employers with the skills of newly hired employees. The advantages of increasing human capital are presumed to be readily apparent. Unemployed workers are those with low skill levels. There are high wages and real growth in the incomes of highly trained workers, and stagnant real wages for those who are less well trained workers. In general, a nation's human capital is tied to its competitiveness in the world economy.

Some of the arguments that policymakeres have made for reform, which are widely shared, include:

- The Council on Competitiveness rates improving primary and secondary education as the number one priority for boosting United States competitiveness; 91 percent of its members say this should be the top priority for United States policy for the next five years.
- 84 percent of voters agree that the nation's ability to compete in the global economy is directly affected by the quality of the educational system.
- 95 percent believe that it is urgent that the United States adopt school reforms to make the labor force more competitive; 81 percent think that this will require changing the present education system in major ways.

The second source of concern has focused on the quality of public life. In the language of the OECD this would be similar to the issue of social cohesion. Academics, including James

Coleman, Robert Putnam, and others, have talked about an aspect of public life that relates social capital to features of social organization (social networks and norms)that enable citizens to coordinate and cooperate for mutual benefit.

One easily observable measure of social capital is the degree of civic participation in voluntary associations. DeTocqueville commented on the strength of voluntary associations in America and saw them as a source of strength for the democracy. There has been recent evidence of a decline in the levels of civic participation.

- Participation in many conventional voluntary associations (parent-teacher associations (PTAs), community chests, Red Cross, recreational athletics, etc.) declined by 25 to 50 percent over the last two or three decades.
- People spend 25 percent less time on informal socializing and visiting.
- Collective participation is down:
  attending a rally or speech—down 36 percent;
  attending a town or school affair—down 39 percent;
  working for a political party—down 56 percent.
- Over the last 25 years, there has been a decline of roughly a third in the measures of social trust among all education levels.

The many reasons suggested for a decline in the vitality of civic life are well known: the rise in the number of households in which both wives and husbands work, the increase in single parent households, significant geographical mobility, increased television watching, and a general decline in "free" time.

This decline in civic life reflects an ironic shift—some of services that were provided by various voluntary civic organizations in the past are now purchased in the market. Even if the market-based services are individually quite good, this can lead to an erosion of social capital. Alternatively, there is increased reliance on state organizations, such as schools, at the same time as there is increased suspicion of their effectiveness.

This decline in social capital has at least two consequences for schooling. First, it made civic participation an issue that policymakers and others hope will be addressed through

schooling. This could be described as the socialization character of schooling, helping students to become citizens and to define their role in civic life. Here, schools are seen as playing a restoration role.

The second consequence is that the decline in social capital in general has consequences for how schools can operate. Concrete examples of social capital would be when one friend asks another about schoolwork, or when parents talk with each other about their children's schooling or participate in a volunteer day at the school. When communities are strong and play an important role in schooling, they reinforce norms and teach the importance of the work ethic and other societal values.

How Will Standards and Assessments Address These Two Issues?

Standards-based education systems can define challenging expectations for both students and teachers. Standards that set challenging goals for students and teachers can stimulate changes in the curriculum and instruction and enable students to reach high academic standards.

Such standards become the keystone for developing coherence in the educational policies. In developing standards, we must decide whether they will define basic or advanced skills, be general, or discipline-based. Standards need to be specific enough to provide guidance for teachers and students but not overly restrict school innovations nor prescribe detailed instructional strategies.

Assessment is seen as a powerful tool to help realize the goals of standards-based reform. It is a prime source of motivation and feedback, for both teachers and students. The formulation and dissemination of valid assessment instruments can help to translate standards into concrete and operational terms. In the past, very few assessments were aligned to what was taught. This strange tradition was a product and a symptom of the fragmentation of policy.

Countries differ in their beliefs about the efficacy of large scale assessment exercises. For the United States, facing a crisis of legitimacy for the school system, the demand is to strengthen accountability through assessment. The aim is for greater transparency—schools cannot expect to be funded on the basis of a blind trust that learning is taking place. Thus, the challenge

is to secure accountability through assessment methods which are valid, i.e., which are aligned to the learning aims expressed through agreed standards. Assessment should reflect and reinforce the outcomes of teaching and learning that reforms aim to secure, which implies that assessment methods must provide evidence that is a valid reflection of the achievement of these aims. This is a demanding requirement because the direction of many reforms is towards enhanced practical action by pupils in tackling realistic and complex problems. The capability they may thereby develop cannot be assessed by traditional methods that employ external tests designed to combine reliability with low cost.

One purpose of assessment, the immediate improvement of learning in the classroom, will only be well served by a significant improvement in teachers' competence in formative assessment. Such competence is particularly necessary where teachers are implementing new programs to realign their efforts in light of the standards which societies expect their schools to achieve. Good feedback is essential to the process of turning a new idea or aim into effective practice.

Thus, there is a need for close coordination of efforts to redesign curricula, improve both formative and summative assessments, and renew support for the professional development of teachers.

Challenging standards and assessments can also address the issues of human and social capital in the following ways:

- If standards and assessments raise the level of performance of the education system, the graduates will have increased knowledge and skills—specifically those skills that are of value to the economy.
- Standards provide a common set of intellectual knowledge, which should be an important resource of background knowledge for media, for the press, and for national conversations, as well as for more local or community dialogues.
- Standards create a set of expectations that can be shared among students, teachers, and parents. These expectations can become the basis of a set of mutual obligations. Such shared norms and

obligations are the basis for the development of social capital.

- Standards and assessments can increase the public's trust in educational institutions. By making the work of the institutions more transparent they will make it easier for these institutions to build trust, to command resources, and to build legitimacy. The days of "blind faith" in the efficacy of public institutions is past.

The Role of Technology

One topic that has been under-investigated, and has not been an adequate part in the OECD conversations about standards and assessments, is the role of technology. While the push for the adoption and adaptation of technology for education is not new, there has never been as great a push to incorporate technology into schools as we are experiencing at present. While schools have always been trying to catch up to technological advances, never has the access to such advances been so close at hand and seldom have the advances seemed to have such dramatic implications for education.

A key is how the technology is used. Investments in technology should complement the standards and skills being taught in schools. If used appropriately and effectively, technology can assist in the efforts to challenge students to achieve higher standards and higher levels of understanding.

Computers have penetrated all levels and features of American society. Almost 50 percent of our teachers have computers at home, access to the Internet is becoming very common, the home page of the Department of Education has several thousand users everyday, and major resources such as materials from the National Archives and Library of Congress are being digitized for access over the Internet.

We have examples of high technology schools where student achievement has improved, student attendance has increased dramatically, and there is an increased excitement about learning. Technology can clearly be a powerful lever to improve student learning. It can also be a resource in the creation of social capital. The Internet makes it possible for the wide dissemination of information about standards as well as assessment results. Computer networks can be established by schools to better

communicate with parents. There have been some efforts to link homes and schools through computers to make available student assignments or school information.

The use of computers in schools on a widespread scale creates new opportunities to think about assessments and how assessments can be redesigned to take advantages of computer technology. To what extent can assessments designed to monitor the performance of the education system be administered through computers? The technology promises to give both lower costs and increased speed, but we need to know if this can be done without sacrificing quality. Other questions arise, notably:

- Can computer administration of assessments make them more reliable and valid?
- Can the use of computers aid in the creation and maintenance of student work?
- Can students store their work for a year on a single CD-ROM?
- How can the use of computer technology increase the reliability of scoring of student work?

Last year, President Clinton announced a technology challenge to the country: within five years, every teacher in the nation will receive basic training in the use of computers to help students achieve high academic standards; schools will have available in every classroom one computer for every five students, the more than 84,000 public schools in America will be connected to the Internet; and every state and local school district will be focusing on the use of technology to help students learn both the basics and challenging content to achieve high academic standards.

Pessimists will ask whether this is another silver bullet that will not lead to change and will wonder whether computers will wind up being unused—dusty and locked in a closet. The evidence is that such pessimism is not justified. There is overwhelming support and interest in using computers in schools, especially among parents. There is also widespread industry interest—companies are volunteering to do the cable work to hook up schools, refurbishing and updating computers for school use, loaning employees to schools to help them network computers. Although few local bond issues to raise public finance are receiving support, those focusing on

technology are having success. There is now an average of 70 computers per school.

There has not been a recent school reform effort that has elicited such high levels of corporate support and interest. The challenge is to harness the capacity of computer technology in the service of higher levels of academic performance, and not to have technology just to for the sake of having technology. To be sure, computer technology can be "addictive," but in part it is addictive because it is empowering.

Standards and assessments are traditional policy tools. The push for more productive education systems has encouraged us to become more rational about using them as policy levers. But in many ways they are an effort to "shape up the system," to provide the system with greater coherence around a set of high academic standards, to refocus the system. Such an attitude can lead us to distrust encouraging innovations that are less controllable. Certainly, computer technology is such an innovation. Once linked to the Internet—there are almost endless places one can go, people to talk with, and resources to access—the innovative enterprise can go well beyond the boundaries of the classroom, the school, the community, the nation. Such intellectual explorations are to be welcomed and can work in the service of student achievement. We must be certain that the efforts to improve the education system do not leave the schools beginning the 21st century looking like they did at the end of the 19th century.

CHAPTER 8

SYSTEMIC REFORM—NATIONAL POLICIES ON CURRICULUM AND ASSESSMENT: THE SPANISH EXPERIENCE[14]

ALEJANDRO TIANA FERRER, Instituto Nacional de Calidad y Evaluación (INCE), Spain

An Education System in a Process of Reform

The Spanish education system has been undergoing a process of reform for several years now. The basic characteristics, structure, and organisation of the system currently being changed date from 1970, when the General Law on Education (*Ley General de Educación* or LGE) was passed within the context of a more wide-ranging process of modernisation in Spanish politics and society (Casanova, J.V., unpublished Ph.D. thesis, 1982; Tiana, 1992). Subsequently, the end of the Franco regime, the transition to democracy, and the passing of the Constitution of 1978 changed the basic points of reference according to which the educational system was governed, forcing it to step into line with the new political and social circumstances.

Changes in education had already gotten underway during the first democratic governments, with the twofold aim of developing the Constitution and adapting the education system to new social and economic needs. But it was after the 1982 socialist victory when transformation of the whole system was embarked upon. The general framework for the Spanish education system began to change, and the principles of participation, university autonomy, the right to education, and

[14]This paper is a shorter, revised version of a longer one, called "Quality, curriculum, standards, assessment. A Spanish perspective," prepared by the author as a contribution to a US/OECD project with the same title.

the distribution of educational responsibilities began to be applied.

Decentralizing the Education System

This final aspect is particularly important to understand the current organisation of the Spanish education system. Indeed, whereas our historical tradition had been based on centralism, after the passing of the Constitution in 1978, a definite process of transfer of responsibility to the so-called *Comunidades Autónomas* (autonomous regions) started. Although Spain was not proclaimed to be a federal state, the State comprising Autonomous regions is quite similar to that model. To a certain extent, the State forgoes management of those matters which can be taken on by the regions themselves. This is the feature which best characterises the current process of decentralisation in Spanish education when compared to developments in other countries (De Puelles, 1992).

The educational responsibilities established by the Constitution as exclusive to the State are very limited, confined to the general regulation of the system, the determination of minimum requirements for a school, the conditions for obtaining and issuing qualifications, the regulation of the core curriculum, general planning on education investment, policy on study assistance, ownership and management of Spanish schools abroad, and international cooperation on educational matters. The autonomous regions, for their part, assume full responsibility for the regulation and administration of all aspects, levels, and forms of education, except for those responsibilities attributed exclusively to the State. This means that the Spanish education system has a common structure and core curriculum for the whole State, but that it is managed in a decentralised way by the autonomous regions. I will consider the implications of this for curricular development and assessment at a later stage.

The decentralisation process initiated in 1978 has still not reached its completion. At the current time, only seven of the 17 Autonomous regions (Catalonia, the Basque Country, Galicia, Valencia, Andalusia, Navarre, and the Canaries) have assumed full responsibility for education. The other ten will do so gradually over the next few years. The Ministry of Education and Science currently has a dual role to play: On the one hand it has to exercise the functions of the State in education

nationwide, on the other hand, it also has to assume some of the responsibilities which belong to the Autonomous regions in ten of these regions. It is important to bear this duality in mind to understand some of the issues which will be dealt with below.

Reforming the Education System

Together with the transfer of educational responsibility to the Autonomous Regions, the early eighties saw the beginning of the reform of primary and secondary education in a limited and experimental way. Between 1985 and 1986, education policy in Spain was reviewed by the OECD (Ministerio de Educación y Ciencia, 1986), which slightly deflected the experimentation process. From then on, the appropriateness of starting an overall reform of the education system began to be considered, bringing the former experimental period gradually to a close. Thus, in 1987 the Ministry of Education and Science presented a project for educational reform (Ministerio de Educación y Ciencia, 1987), which was widely distributed and debated. The public debate was synthesised into five volumes (Ministerio de Educación y Ciencia, 1988), and the so-called *White Paper for the Reform of the Education System* was published (Ministerio de Educación y Ciencia, 1989). The *White Paper* served as the basis for the subsequent drafting of the new Law on the General Ruling of the Education System (*Ley Orgánica de Ordenación General del Sistema Educativo*, or LOGSE), passed in 1990. This was the start of a far-reaching reform of the education system, with a view to responding to social, cultural, and economic needs.

The process of educational reform encouraged by the LOGSE is moving in three complementary directions: structural transformation, improved quality of education, and curricular renovation. In the first place, the law transforms the structure of the Spanish education system. With a view to bringing it closer to the models which have found most widespread acceptance internationally, it establishes new educational stages or levels: primary education, lasting six years, followed by compulsory secondary education for four years, and a two-year Baccalaureate. In this division, the primary level corresponds to the period where knowledge and instrumental skills are acquired, whereas the secondary involves tackling the different areas and subjects in greater depth. On the other hand, both

levels are conceived on a cyclical basis, divided into two-year cycles. Each cycle is conceived to be a single unit, with the final decision as to whether the pupils move on to the next cycle being taken at the end of each. Nonetheless, the second cycle of secondary education is organized in a different way from the rest, given the specific characteristics of this level and of the pupils.

As well as restructuring the system, the LOGSE extends compulsory schooling up to the age of 16, thus bringing Spain into line with most of its neighbouring countries. The main novelty is not so much in this extension, but rather in the fact that the model followed is a comprehensive one, and in accordance with this, all pupils follow the same route through school until the age of 16. Although pupil diversity is adequately dealt with, this does not mean that they are divided into different streams.

Improving the Quality of Education

The second direction that the reform process is taking is aimed at improving the quality of education, one of the central purposes of the reform carried out.

In fact, the LOGSE devotes a full chapter to the factors that contribute to improving the quality of education. Specific attention is paid to in-service teacher training, which has constituted a political priority during this period of reform; autonomy of schools, allowing them to develop their own curricular project; the promotion of research and innovation in curricular, methodological, technological, and organisational matters; provision of academic, psycho-pedagogical, and vocational guidance; the redefinition of the role of the inspectorate; establishing a new mechanism for the evaluation of the educational system by means of creating the National Institute for Quality and Assessment (INCE).

In the belief that improvement in the quality of education should be an ongoing concern of educational administration, the Spanish Parliament recently passed a new Act for improving the participation, evaluation, and administration of schools (*Ley Orgánica de la Participación, la Evaluación y el Gobierno de los Centros Docentes*, or LOPEG, 1995).

A New Model of Curriculum Design and Development

As well as restructuring the education system and adopting a series of measures aimed at improving quality, the LOGSE has introduced a new model of curricular design and development. There are two complementary factors that have motivated the introduction of this model.

Firstly, the need to reformulate the curriculum established in 1970, to make it more coherent and adapted to new requirements, became evident at the end of the eighties. The highly prescriptive curriculum established had to give way to a more flexible one, in the development of which the teaching staff were to take a more active role. Schools were to enjoy greater freedom in curricular development than under the previous system. The basis for this was to be found in a constructivist view of learning (Coll, 1987).

Secondly, the new distribution of responsibilities and the start of the decentralisation process towards the autonomous regions demanded the design of a new curricular model in which the core curriculum established by the State would be completed by the regions themselves.

The LOGSE has tried to respond to the new situation and the needs expressed, developing a curricular model with the following basic features:

- It is an open and flexible model, in the development of which Autonomous Regions, schools and teachers had to take an active part. The purpose of this conception is to allow a wider range of specific aspects to be added to curricula.
- It aims to respond to four fundamental questions that allow the planning and development of educational practice: what to teach, when to teach, how to teach, how and when to assess.
- It is a broad model and not simply limited to the acquisition of concepts and knowledge, but also includes practical skills, attitudes and values.
- It has two different functions: to make the intentions of the educational system explicit and to serve as a guide to teaching.

The curriculum takes on more specific form through three different levels. The first one is the so-called basic curricular design or official curriculum, comprising the compulsory nationwide core curriculum and the curriculum established by the autonomous regions. At the second level there are curricular projects corresponding to specific stages and schools, drawn up by the teachers from each individual school. At the third level are the class programmes, in which a school's curricular project for a particular set of pupils is specified.

First Step: Setting the Basic Curricular Design or Official Curriculum

The first level of specification of the curriculum is the so-called basic curricular design or official curriculum. This is defined by each of the autonomous regions with full educational responsibilities (at the moment, seven), and must incorporate the core curriculum established across the State. According to the LOGSE, the basic contents of the latter need not occupy more than 55 percent of the school timetable for those regions with an official language other than Castilian Spanish and 65 percent for the other regions.

This core curriculum is intended to be a guide on the one hand and to be prescriptive on the other. This means it serves as a basis for the subsequent curricula developed by the autonomous regions, schools, and teachers, and it is also compulsory. However, the core curriculum represents minimum goals for a pupil to reach, and education should not be limited to these goals.

The core curriculum includes various components. Firstly, it states the general goals of the stage or level (primary or secondary). They are defined as a series of different abilities (locomotive, cognitive, affective or emotional balance, interpersonal relationships, and social action and integration) which pupils have to develop or learn in the course of their schooling as a consequence of their education.

Secondly, the core curriculum includes broad curricular areas for primary and compulsory secondary education:

Primary education

- Knowledge of the natural, social, and cultural environment

- Artistic education
- Physical education
- Language, official regional language and literature
- Foreign languages
- Mathematics

Compulsory secondary education

- Natural sciences
- Social sciences, geography, and history
- Physical education
- Visual and plastic education
- Language, official regional language, and literature
- Foreign languages
- Mathematics
- Music
- Technology

Cross-curricular subjects

- civic and moral education
- education for peace
- education for health
- education for gender equality
- environmental education
- sexual education
- education of the consumer
- road safety

In each of these areas, the curriculum includes various components:

- An overarching statement or explanation of the sense, approach, and the general principles of the area, accompanied by guidelines for its teaching.
- General objectives, expressed in terms of abilities that the pupil is to have attained by the end of the stage or level. Unlike the stage or level goals, these add an explicit reference to contents.
- The most suitable contents to develop the abilities included in the stage or level goals and area objectives. This is not a list of subjects to be dealt with, but is rather a catalogue of blocks to be worked on at different cycles of the level. The area

contents do not refer solely to conceptually based contents, but also incorporate contents relating to procedures, values, and attitudes.

- Attainment targets or criteria to design activities that enable assessment and are coherent with the general goals of the stage or level and the area objectives.

With this core curriculum as their basis, the autonomous regions establish their respective curricula. Analysing the official curricula of the autonomous regions reveals a certain homogeneity between the various area objectives and the attainment targets. The slight variations noted consist of extending some of the goals, objectives, or targets, almost always with the intention of including specific features of the region itself. More differences are visible in terms of contents, given the considerable flexibility that the various regions have in completing their curriculum.

As the core curriculum, whether national or regional, has to be formally approved and passed in the form of a decree, it is a matter of discussion in the State (or regional) school council, a body of participation in which teachers, parents, students, and administrators are represented. Subjecting the proposals made by the different administrations to public scrutiny in these bodies allows the different views about educational objectives and contents to balance.

Second Step: Drawing Up the School Curricular Project

Once the official curriculum has been established by each autonomous region, schools have to adapt it to their particular requirements. This adaptation takes the form of a document called the curricular project (in the event that a school offers more than a single stage or level, it must draw up a project for each of them). This document represents a second step in the process of curriculum development, now at school level. In fact, when writing it, teachers have to discuss and agree about area objectives, contents, attainment targets, and assessment criteria and procedures. Although it is true that certain teams of teachers were already doing this, the novelty of the current reform is in making this general practice.

The curricular project is a central feature of the Spanish model of curricular development. It constitutes an intermediate link between the official curriculum and the teaching activities that take place in schools and classrooms. Its aim is to ensure the continuity and coherence between the two poles, promoting a specific identity alongside a "school spirit." At the same time, it constitutes an element of reflection on school practice and a stimulus for team work on the part of the teachers (Coll & Martín, 1994).

When a school tackles the task of drawing up its curricular project, it has to bear a number of factors in mind. First, it has to relate the curricular project to the school's educational project, a document that defines its marks of identification and major aims, in line with the feelings expressed by the members of the school community. Second, it has to analyse the context of the school itself, keeping in mind the psychological and social characteristics of the pupils of each level. Third, the school must link the project to previous experiences. Fourth, it must take the official curriculum of the corresponding autonomous region as a reference point.

The curricular project of every school should reflect the relevant decisions adopted on the following elements of the curriculum (Del Carmen & Zabala, 1991).

General goals of the level. The school must adapt the goals set in the core curriculum to its specific characteristics. Furthermore, the cross-curricular subjects described above must be incorporated into the curriculum.

Sequence of objectives and contents of each cycle. The official curriculum does not distribute the general objectives for each area, their contents, and attainment targets in cycles. This represents one of the key decisions that teachers have to make when drawing up the curricular project. It should set out the abilities and contents that have to be worked on in each cycle.

Methodological strategies. The curricular project must also include references to elements such as methodological principles and options for each curricular area, the criteria adopted to put pupils into groups, organisation of space and time at school, or criteria for the use of the materials and teaching resources.

Assessment strategies and procedures. As regards decisions on assessment, the curricular project must first tackle the question of what to assess. On the basis of the attainment targets included in the official curriculum, the school should review

them according to their own characteristics, adapting these where necessary, and draw up attainment targets for each cycle according to the sequence of objectives and contents carried out in the project itself. The question of how to assess also should be tackled, with reference to situations, strategies, and procedures. The curricular project must include a decision about when to assess, considering at least three basic times: initial, formative, summative. Also, the document must determine the type of report that is to be used at the school and to whom the information regarding the results of the assessment is to be given. The curricular project must establish criteria for: deciding whether or not to move pupils up to the next cycle or grade; supporting the pupils who receive negative assessments; and deciding if a pupil will attain the final qualification.

Measures to ensure diversity. The curricular project must, finally, establish the contents and structure of the guidance programmes which are to be developed in the school, the optional subjects to be offered (in the case of secondary schools), the way in which curricular diversification will be organised, as well as the organisation of resources for pupils with special educational needs.

Responsibility for drawing up the curricular project lies with the team of teachers for the particular level at each school. Coordination of the process is assured by the Commission for Teaching Coordination (*Comisión de Coordinación Pedagógica*) of the school. Once it has been drawn up, the project is discussed and approved by the Teachers' Senate, which takes it to the School Council (comprising representatives of parents, teachers, and pupils from the school) to issue their report. Finally, the project is supervised by the Inspectorate, who can suggest further elaboration of specific points.

The drawing up of the curricular projects has taken place in parallel with the implementation of the new educational structure. As schools have incorporated the new rulings, they have had to draw up their own projects. The specific process followed has varied from one autonomous region to another, but without any extreme differences. Usually, the first version of the project is developed over a period of one or two years. Nonetheless, the Ministry of Education and Science has tried to avoid making the process too bureaucratic, insisting on the need to consider it as something unfinished and to review it from time to time. In accordance with this concept, schools have to draw

up an initial proposal at the beginning of the school year, applying it during the year, assessing its repercussions, and, where necessary, revising it. In reality, the drawing up of the curricular project should be conceived of as a lengthy, almost permanent activity, and as a basic ingredient in exercising the profession of teacher. It implies giving as much emphasis to the process as to the product (Coll & Martín, 1994).

Third Step: Drawing Up the Classroom Programme

The third level of specification of the curriculum is composed of what are known as classroom programmes. These consist of specifying the decisions taken in the level or school curricular project for each specific group of pupils. The programmes are drawn up by each teacher for his or her particular set of pupils and for each year, including the sequence of contents, the teaching units to be developed, and the assessment procedures to be applied. In fact, programmes set standards for each group of students and year. In a sense, every single teacher has some room for setting concrete standards for their students.

A Flexible, School-Based System of Assessment

In line with the Spanish model for designing and developing the curriculum, the mechanism established for the assessment of students has three basic characteristics:

- It is a flexible system, since the specific attainment targets and criteria of assessment, accreditation, and promotion to the next level are established by each school through its curricular project. This means that the school can adapt the general criteria to its particular circumstances, depending on the particular features of its milieu and its pupils.
- It is a school-based system, since the assessment takes place entirely within the school itself. In contrast to other countries, Spain has no national or external examinations, either at the end of the primary or secondary education.[15] Pupils are

[15]The only external examination takes place after the Baccalaureate for those pupils who wish to enter a university. Whereas this is a prerequisite

assessed by their own teachers throughout their period of compulsory schooling.

- It is a system which emphasises the need to assess not only the results the pupil obtains, but also the process of education in the school, the classroom, and the teaching carried out by teachers.

According to that model, students are assessed within the school itself. The first consequence of this decision is that schools should fix attainment targets which will be applied to each level, cycle, and area in their curricular project. These targets need to be based on those set out in the official curriculum but should be adapted to the school's circumstances and those of its pupils, establishing its sequence in different cycles. In certain cases, the target is assigned exclusively to a single cycle, while in others it is distributed over successive grades and cycles.

On the basis of this sequence of attainment targets, which is associated with a further contents sequence, teachers determine their own assessment mechanisms in their respective classroom programmes.

This approach to assessment is based on a constructivist conception of learning, which has inspired the Spanish model of curricular development. Requirements—such as the continuous, global, and integrated nature of assessment, and insistence on the formative, regulatory, guiding, and self-correcting nature of the assessment, which must be geared to improving processes and results—are derived from the need for coherence between the curriculum and the assessment of pupils. In daily classroom activities, teaching, learning, and assessment are very closely linked, forming part of a teaching and learning continuum.

In line with these general principles, the ministerial regulations establish that assessment should be carried out collectively by the teachers of a group of students under the coordination of their corresponding tutor. This group of teachers meets at least three times a year to analyse the learning situation of each pupil, giving him or her a grade. In the event that this assessment is negative, the teachers have to take the necessary reinforcement measures and, when necessary, adapt the curriculum. The

---

of university study, it is not required to obtain the diploma of secondary education graduate or the Baccalaureate.

results of these assessments are given to pupils and their families, in accordance with the criteria set out in the curricular project.

Approaches to assessment must constitute the basis for moving up a grade, for the accreditations and qualifications attained by pupils. Therefore, at the end of each cycle, the teachers decide whether pupils will go up to the next cycle or level. In the case of pupils with learning difficulties, it is possible to choose between continuing, together with the necessary proposals for remedial activities, or repetition. Repetition can only be chosen once during primary education and once again in secondary. The information provided from these assessments is reflected in each pupil's school file and book of school reports. At the end of compulsory secondary education, pupils receive an accreditation which shows the school years covered and the marks obtained. In the event of an overall favourable result, they also receive the diploma of secondary school graduate.

External Support, Control, and Assessment Mechanisms

The model for pupil assessment which is currently being implemented in Spain is, as has been stated, flexible and school-based. As assessment is understood to be a constituent part of teaching and learning processes, this flexibility is very advantageous because it enables an adequate response to the needs of the different schools, teachers, and pupils. A second advantage this model offers is that of involving teachers in decision-making processes. As has been said, teachers have to draw up a school curricular project, which must expressly include the attainment targets for the different areas of each level. The requirement that specific decisions be made to assess what pupils are learning automatically entails teachers taking an active part in considering the teaching to be given and the criteria on which the results are to be assessed. This implies a stimulus for teachers as professionals, although it can present problems in practice.

Together with these advantages, the model presupposes certain risks. The first of these is the possibility of the system splintering and allowing a loss of control in results. Because schools themselves make the decisions on assessment,

promotion, and accreditation of students, criteria applied could be too heterogeneous. In an education system like the Spanish one, in which respect for the right to education on conditions of fairness is a fundamental principle, the aggravation of this situation could mean it did not fulfill one of its main aims. A second problem is that of comparing the results of different regions, schools, or pupils. This is not serious unless it affects the basic rights of the citizen, although, even without doing so, it may make it difficult to steer the education system as a whole.

With the intention of taking the utmost advantage of the benefits of this model and circumventing its difficulties, avoiding the risks and disadvantages set out, education administrative bodies have set a number of different mechanisms in motion. Some of these are geared to controlling the system and providing information, while others are targeted at supporting school development.

The Role of the Inspectorate

Among the control and monitoring mechanisms that should be mentioned are the inspectorate services. Each regional administration has its own inspectorate, the role of which is the control, supervision, and evaluation of the education system within its sphere of competence. Its methods of action cover a relatively broad spectrum of areas. On the one hand, the inspectorate monitors the drawing up of curricular projects on a regular basis, helping the schools in this process and checking that the result corresponds to what has been established. In this way, it supervises the decisions adopted in assessment (among other aspects), drawing attention to those which are not justified or clash with legal provisions. This makes for a first and elementary level of control. Secondly, schools send a copy of their assessment documents to the inspectorate each year, which can ensure that the criteria adopted are being applied, and can then act in cases of incoherencies or irregular practice. Finally, examination of the assessment mechanisms applied in a particular school is explicitly included in the schools evaluation programme which the inspectorate has been developing over the last few years (the Plan EVA).

Until now, more than 500 schools have been assessed by the inspectorate, which has allowed a specific idea to gel regarding the real application of regulations on this and other aspects of

curricular development (Luján & Puente, 1993). It goes into greater detail and depth than the two preceding types of assessment and constitutes a third level to control the system's operation.

External Examinations: Aptitude Tests for University Entry

A second mechanism to monitor the assessment system comprises the analysis of the results of what are known as aptitude tests for university entry. Spanish pupils who have completed their Baccalaureate and wish to gain access to university have to pass these tests, which are held every year and the marks for which, once combined to form an average with those from upper secondary, enable university courses to be chosen. There are a total of eight tests and they are held over two or three days. Some of these are common to all pupils (language and literature, foreign language, text commentary, philosophy) and others depend on the subjects that have been studied at school. Each university designs its own tests, although the structure and characteristics of these are the same. Furthermore, these are nationwide examinations, since they permit entry to any Spanish university, once they have been passed. Although that is not the main objective, they help to mitigate the worst excesses of heterogeneity in the assessment criteria adopted by schools in preparing Baccalaureate pupils for tests that are comparable though not identical.

The results of these tests have come under close scrutiny in the last few years. Social demand for their rigour, reliability, fairness, and comparability have led to the writing of a large number of works which allow interesting conclusions on roles and effects to be deduced (Muñoz Repiso et al., 1991; Muñoz Vitoria, 1993). In fact, these tests have made it possible to obtain a better idea of education results at the end of the Baccalaureate, and at the same time have reduced the risks of fragmenting the assessment system.

The Overall Evaluation of the Education System: The National Institute for Quality and Evaluation (INCE)

A third mechanism takes the form of the new National Institute for Quality and Evaluation (INCE), which started work in 1993. Its basic task is the overall evaluation of the education system.

This new body has begun to develop studies that are aimed at getting to know education system results in greater depth by collecting information in schools. Although its action is not limited to a simple examination of students to assess their performance, the information they are able to provide on this is vital for carrying out synchronic and diachronic comparisons. Thus, the studies currently underway have permitted a better acquaintance with possible inequalities that may exist between regions and types of school, as well as the development of results over a period of time. These studies, which are diagnostic in nature and are carried out on samples of schools and students, must provide the basis for more sustained analysis of the efficacy and efficiency of the education system, supporting decision-making. In addition to this, the Institute's actions constitute an essential component in steering the education system (Tiana, 1995).

In the specific field of assessment, it is hoped that the Institute will make valuable contributions: that it will enable assessment procedures that are in keeping with the general model of curricular development to be tried out, prompting the spread of new practices; that it will provide teachers, families, administrators, and society as a whole with relevant information on levels of achievement in schools and pupils, thus providing feedback to the process of setting standards in schools; and that it should provide information on the implementation of the new education arrangements, allowing decisions to be made that will better guide the process.

It is important to point out that the mission of the National Institute for Quality and Evaluation (INCE) is not to define standards. The Spanish model of curricular development assigns that role to schools. Nonetheless, the information contributed by the studies carried out must enable schools, teachers, families, and society as a whole to reflect on the levels of achievement in the education system and on a possible review of its aims and assessment criteria. From this point of view, it is a crucial mechanism for quality improvement in education.

School and Teacher Support Mechanisms

The educational authorities are not limited to establishing mechanisms for controlling the assessment system. Equally or even more important than these tasks is that of setting up school

and teacher support mechanisms to develop and apply the new curriculum. Among these, the role of the inspectorate should once again be mentioned in its capacity as an instrument of guidance for schools. Far from being limited to controlling the actions of schools, the inspectors work with them in the process of developing the curriculum, offering guidelines and models that they can apply.

Together with the inspectorate, teachers' centres have offered a wide range of training activities for several years now. It is not difficult to see that the quality of the Spanish model of curricular development is due, in large part, to what is done by teachers. It is for this reason that teacher training has been a key strategy within the Spanish education reform process. Some of the activities concerned have been geared to facilitating the process of introducing the new curricular model, especially those activities held in schools for their own teachers. Other bodies dependent on the administration, such as educational programme units or psycho-pedagogical support teams, have also cooperated in this task, although they have had different levels of responsibility.

Overall, education authorities have created a wide range of instruments to ensure the smooth running of the mechanism set up. Some of these have been aimed at controlling it, while others have tried to give support to those who are responsible for or participate in its application. But all of these should be seen as complementary facets of the same strategy.

Final Remarks

Before concluding, it may be worth adding a few general and personal reflections on this subject. All societies show a more or less pronounced interest in the achievements of their education system, although not all express this interest in an identical way. Every education system has its own context, its own history, traditions, and characteristics that define the shape taken by the debate on standards and the decisions that have to be adopted in this respect.

It would seem clear that any approach to taking decisions on assessment procedures should be couched in a coherent model of curricular design and development, respecting its peculiarities. Otherwise, we run the risk of setting up separate or

even divergent mechanisms which threaten the necessary coherence of the system.

The Spanish experience would seem to vouch for an understanding of standard setting as not simply a form of controlling and monitoring the results of the education system, but also as a means or renewing school practice. Social expectations generated by this issue constitute a powerful driving force within education systems.

In the specific case of Spain, assessment should not be understood as a simple means of marking students, but as part of the core of the teaching and learning process, all of the components of which are affected. For this reason, the design and implementation of the assessment mechanisms must be consistent with the curricular model adopted to avoid the curriculum being indirectly affected by the assessment.

The above remarks impose certain requirements when education systems are comparatively analysed. As comparison is both possible and desirable, it should be carried out in a way that respects the characteristics and aims of each education system without distorting the interpretation of the results.

REFERENCES

Coll, C. (1987). *Psicología y currículum*. Barcelona: Ed. Laia.

Coll, C., & Martin, E. (1993). La evaluación del aprendizaje en el currículum escolar: Una perspectiva constructivista. In COLL, C. et al. (Eds.), *El constructivismo en el aula*. Barcelona: Ed. Graó.

Coll, C., & Martin, E. (1994). Aprendiendo de la experiencia. *Cuadernos de pedagogía, 224*, 8–15.

De Puelles, M. (1992). Informe sobre las experiencias de descentralización educativa en el mundo occidental. *Revista de Educación, 299*, 353–376.

Del Carmen, L., & Zabala, A. (1991). *Guía para la elaboración, seguimiento y valoración de proyectos curriculares de centro.* Madrid: Ministerio de Educación y Ciencia-CIDE.

Lujan, J., & Puente, J. (1993). El Plan de Evaluación de Centros Docentes (Plan EVA). *Bordón, 45*(3), 307–320.

Ministerio de Educacion y Ciencia. (1986). *Examen de la política educativa española por la O.C.D.E.* Madrid: Ministerio de Educación y Ciencia-CIDE.

Ministerio de Educacion y Ciencia (1987). *Proyecto para la Reforma de la Enseñanza*. Madrid: Ministerio de Educación y Ciencia.

Ministerio de Educacion y Ciencia (1988). *Papeles para el debate*. Madrid: Ministerio de Educación y Ciencia.

Ministerio de Educacion y Ciencia. (1989), *Libro Blanco para la Reforma del Sistema Educativo*. Madrid: Ministerio de Educación y Ciencia.

Muñoz-Repiso, M. et al. (1991). *Las calificaciones en las pruebas de aptitud para el acceso a la universidad*. Madrid: Ministerio de Educación y Ciencia-CIDE.

Muñoz Vitoria, F. (1993). *El sistema de acceso a la universidad en España, 1940–1990*. Madrid: Ministerio de Educación y Ciencia-CIDE.

Tiana, A. (Ed.). (1992). La Ley General de Educación, veinte años después. *Revista de Educación*, n° extraordinario.

Tiana, A. (1995). Perspectivas españolas actuales para la evaluación del sistema educativo: El Instituto Nacional de Calidad y Evaluación. In M.J. Saez Brezmes (Ed.), *Conceptualizando la evaluación en España*. Alcalá de Henares: Universidad de Alcalá.

CHAPTER 9

REFLECTIONS ON THE CONFERENCE: NEW INSIGHTS AND FURTHER ISSUES

PAUL BLACK, King's College, London
ALAIN MICHEL, Inspecteur Général de l'Education Nationale, Paris

Shocking Results

Both Barry McGaw and Margaret Brown quoted evidence which ought to give rise to serious concern about teaching and learning. It seems to be the case, across many subjects and across all countries, that the spread in achievement of pupils in any one age group is very wide: It is common to find that the bottom and the top percentiles can differ in their scores by 60 percent. International testing programmes show that the ranges within all countries are much greater than the range in the differences of mean scores between countries—yet it is these inter-country differences that seem to hold public attention. Moreover, it takes the average pupils many years—on the order of seven—to improve their scores by the same amount as this spread. How can countries be complacent in the face of evidence that, for example, there are pupils in their classes at (say) age ten for whom there is only a small chance that, by the time they may leave school in six or seven years, they will have reached a level of understanding that the top ten percent of their ten-year-old peers have already grasped? It is all too common that the pupils near the bottom of this range will be labelled as failures, year after year, and will become the school drop-outs and the unemployable youth in adult society. Streaming, setting, and making students repeat a year of schooling are all attempts to cope with this problem, but they seem inevitably too crude to respond to the widespread and diverse needs of students, and there is strong research evidence that streamed groups do not produce better overall performance than mixed groups of comparable ability.

One indicator of the development of formative assessment amongst any group of teachers is whether they are at least facing clear evidence of this problem in their classes and are struggling to find ways to respond. For many this seems not to be the case—perhaps because, as Perrenoud has commented, "There is a desire in everyone not to know about things about which one can do nothing."

Assessment, Pedagogy, and the Curriculum

This striking example serves to highlight one of the outstanding problems of formative assessment—it is not possible to act on the information without quite radical changes in pedagogy. The Brown and Baker papers bring this out very clearly. What they also bring out is that, for this reason, formative assessment cannot simply be added to existing practice as a useful extra. Any initiative to enhance formative assessment will require investment of resources in developing methods and support for teachers struggling with the changes: There is no universal quick solution that could be taught to teachers by setting up a short re-training programmes tomorrow. Indeed, as formative assessment develops, teachers must be involved both in composing the instruments for themselves and for sharing with others, and in interpretation of the feedback since this can contribute to broader evaluation purposes as well as to meeting the immediate needs of individual pupils.

Recent changes in pedagogy reflect the constructivist view—that learning must proceed by starting from pupils' expressions of their own ideas which are then challenged and extended. Interaction with pupils and feedback to them are essential in this approach to learning. Looking ahead, if the use of information technology is to play a greater role in learning, as argued in David Stevenson's paper, programmes will have to be interactive and embody formative strategies of high quality. If teachers are to become more effective in the use of assessment, then in addition to assessment expertise, they will need to have a firm basis in their subject knowledge, access to a range of teaching options, and help in terms of summative assessment and curriculum structures consistent with the practices they are trying to develop.

If pedagogy is going to respond to provide for the wide range, in learning needs and in progress in achievement, amongst

students, then this has implications for the structure of any curriculum, whether it be a national framework or one teacher's own planning guide. A curriculum cannot be expressed as a set of topics tied to the successive years of schooling. It has to have a structure which is not strongly linked to age, but which expresses the sequence of learning achievements through which students might best make progress. Such a framework will set out the targets for any particular student in relation to the level of achievement that students has reached—and thereby help to provide learning work with which he or she can succeed whether the student is in the top percentile, or the bottom one, or anywhere in between. The task of formative assessment is then clear. Of course, we do not know enough to set out this sequence within every particular subject and in different cultures, so we must do the best we can and then mount the empirical work needed to refine those efforts—in ways that Barry McGaw's paper sketches out.

Given such a scenario, summative assessments have to be structured to match the curriculum and the pedagogy. The assessment for each pupil should be tuned to the levels of achievement where that pupil can succeed at a substantial proportion but may also show limitations in going further. This means that the summative system will have to be highly differentiated, and that students may be distinguished, not by marks ranging from ten to 90 percent, but by those levels at which they can achieve, say 80 percent. Unless something like this can be done, we might go on sending pupils out of school carrying only a badge of failure. While such a badge may well be evidence that the school has failed them, it will be evidence to them and to those to whom they might present themselves that they possess little competence or capability.

Several of the above arguments indicate the need for supportive consistency between the practices and pressures of summative assessment and the needs of formative assessment. It must not be assumed that formative assessment should be entrusted entirely to the internal work of schools. External stimulus and support are essential. Externally provided instruments and guides to procedures can help in ensuring equity in assessment, in ensuring that schools are working with a common understanding of national or regional standards and thus in strengthening a criterion referenced basis against the natural tendency to work within class or school norms.

Formative Assessment and Summative Assessment

The distinctions between formative and summative assessment seem to be far more important in some countries than in others. The distinction between the two is not necessarily a distinction between agencies—both can be carried out by teachers, and while teachers must be directly involved in formative assessment, agencies outside the classroom can play a central role. Neither is it a distinction as to methods or instruments of assessment—some of the methods used can be the same for both, although other methods will be more suitable for one or other of the two. The essence of the distinction lies in their purposes: Formative assessment is directed to improving learning of pupils, summative is for reporting, whether on individuals or groups. Where the same methods are used for both, the interpretations of the results will differ according to the two purposes.

Admittedly, a formative test, for example, will have to be designed so as to allow for diagnosis of the mistakes in the way of thinking, or in other words, the "status of mistakes." This type of information will help in deciding upon the relevant remedial action. Such an integration of the use of a test and of its results within the learning/teaching process is crucial for formative purposes. It is not essentially a technical matter but rather an attitude, a state of mind, of both student and teacher in reading, translating, and using the outcomes of the test. One important issue is to aim for the "maieutic" of the Socratic mode of enquiry, i.e., to use the tests to bring out evidence of the successful achievements in understanding as well as the shortcomings. In this respect, it is crucial to motivate students to learn more and to improve their performance.

It may be important to emphasise the distinction between these two main purposes, particularly in discussions of policy, for two reasons. The first is that formative assessment is often used as simply another name to cover all teacher assessment. The empirical evidence is that, while most teachers give tests, mark students' work, and ask pupils questions, very few actually design and use these methods to modify the learning approaches to meet the individual needs that the evidence reveals. For example, a test given at the end of work on a particular task cannot be formative because it is too late—the work is over and something new will be started. One reason for weakness here is

that assessment is given little attention in initial training courses, which focus instead on "methods that work" and on "survival in the classroom."

The second reason follows from the first—the aspect of teachers' assessment skills and procedures that is most in need of further research and development is the formative one—as made clear in the papers by Margaret Brown and Eva Baker. This need will not be taken seriously unless the precise nature of and the need for formative assessment are clearly grasped.

Countries differ in the degree to which they might separate formative from summative procedures. Where there is a strong tradition of high-stakes external testing in which teachers play no part, then there might be very little overlap and there is external pressure to give more importance to summative aspects. One main reason is that the ideology of competition, which is spreading within our societies, calls for competition both between schools and even within the school, even although this is often officially denied. This tendency can be more or less encouraged by the external stakeholders and by the government, who see, in objective data about student achievement, important information for democratic transparency and for gauging the effectiveness, either of a school or of the educational system as a whole.

In such situations, there is a danger that the summative pressure will dominate and this will make it hard to foster formative assessment. Such external dominance is strong in some European countries where external selection tests for higher education are influential. Where, by contrast, there is no external testing, and teachers and schools are trusted to report on their students, then it is easier to establish an intimate connection between the two. Thus, within the freedom to draw up its own curricular project that each school has in Spain, as described by Tiana Ferrer, each school has the flexibility to relate the different functions of its assessments.

The national assessment in France of the students at the ages of 8, 11, and 15, described in the paper by Claudine Peretti, shows that one can use the outcomes of national standardised tests for both formative and summative purposes. The tension between the two purposes is overcome by the fact that the tests are taken at the beginning of the school year—so as to indicate that the priority is formative—but their outcomes can be used for summative purposes. This can be the case for the former

teachers of the students in the previous school year. It is also the case for the school: Indicators of outcomes for self-evaluation are not published but can be discussed by the teachers and the school's governing board. The outcomes can also be used, through a sampling procedure at regional and national levels, to yield interesting comparisons at a given time or over time.

A strategy for combining these different types of assessments should take into account that some types are particularly appropriate at particular levels of the education system (e.g., classroom, school, district). It is important to take into account, also, the risk of perverse effects of summative assessment. Such tests can create feedback, the effects of which are inconsistent with the need for more active learning in a rapidly changing world, and which risk giving more relative importance to skills simply because they can be easily assessed (the predominance of mathematics in some countries is partly due to such an effect), and which, for the same reason, can take attention away from cross-curricular skills and from attitudes and values.

Thus, an exclusive concentration on summative assessment could be inconsistent with the requirement to adapt the curriculum to the changing world. In particular, it would not allow for reinforcement of important aspects of the so-called hidden curriculum, some of which are more necessary than ever in the perspective of life-long learning, and for strengthening of the role of the school in maintaining social cohesion.

Systemic Approaches

Systemic Metaphors

It is clearly necessary to consider the ways in which different aspects of any educational system interact. Thus, formative assessment affects, and is affected by, pedagogy and curriculum structure; summative assessment is framed to meet social expectations and in turn affects formative assessment and pedagogy; and so on. Insofar as the term systemic stands for a determination to consider such interactions in planning, it is unexceptionable.

However, a systemic approach to curriculum reform is not merely an instrumental approach. It is a new paradigm—a new frame of reference—for considering the evolution of the education system in relation to its environment. Thus, it

provides a general conceptual framework for defining a strategy for steering a continuous and sustainable process of change over time, to prepare students to face the new challenges of a rapidly changing world.

It is important to look carefully at the sense in which the term "system" is applied to the problems of educational innovation. The term could invoke a mechanical image in which the system is seen as a set of discrete parts, linked as in a machine by determinate effects of one part on the others, often in a hierarchical chain. This can be a misleading metaphor. It is often linked with an assumption that innovation can only happen through commands from the top which are designed to alter everything at once—a strategy that can be both expensive and risky. Even for such innovations, the metaphor is inadequate because it does not reflect the complex and iterative nature of educational change.

A more helpful metaphor is to see the educational system as an ecological system. Edgar Morin expresses this in as a more complex image of an "eco–self-organising system" seen as "sets of units related through an interactive process aiming at a common goal...linked by multiple relations and capable by interaction with its environment to react, to evolve, to learn and to self-organise." The implication of this metaphor is that an evolutionary model, with its slow pace of change coupled with the iterative adaptation of any new element to its complex interactions with its environment, is a better guide to innovation strategy than a mechanical model. In particular, it is consistent with the belief that the system is one which learns about itself as it evolves.

In the machine metaphor, evaluation of reform may be seen as the quantitative measurement of inputs and outputs: The processes by which these are related are either obvious or uninteresting. In the ecological metaphor, input and output data convey little by themselves, because nothing can be inferred from them given the complexity of the system interactions. Evaluation therefore has to illuminate quantitative data by the results of qualitative study of the ways in which any innovation affects the interactions with, and reactions of, other parts of the system.

Systems and Policies

In the context of the new systemic paradigm, both formative and summative assessments are part of the required cybernetic regulation, which is the more necessary as the system becomes more decentralised and complex. The perception of a common goal by the units and actors of the education system and the stabilisation of tensions and disequilibriums cannot result from any rudimentary laissez-faire type of mechanism.

Assessment and evaluation procedures are necessary instruments for aiming at both efficiency and equity across the education system. However, the roles of assessment and evaluation depend on the nature of the systemic policy in which they play a part. A cybernetic regulation implies the existence of an education policy. Education is a public commodity as much as a private one, and this policy must provide all the units of the system with relevant goals and information. As Prigogine said, the rapidity of the circulation of relevant information throughout an eco-system determines the maximum complexity that the system can reach without becoming too unstable. An efficient information function within any system is the basic condition for giving more autonomy to the system while establishing an overall among the various decisions and initiatives. In a centralised system the same result would be sought by control of the allocation of resources and by requiring close conformity to bureaucratic norms.

The advantage of a decentralised system is that it encourages grass-root initiatives and bottom-up changes. However, an extension of such innovations cannot result from spontaneous generation—it requires a strategic policy of communication in which both formative and summative assessment play important roles, along with ways to give value to and promote those innovations which have proven to be effective. These features are illustrated in Tiana Ferrer's account of the Spanish reforms.

The main challenge is to find a balance between the need for more quantitative information and more rationalisation in the operation of the education system, on the one hand, and the risk of a hyper-rationalism or scientific nominalism, on the other.

Systems and Their Boundaries

The ecological metaphor is also helpful in indicating that it is possible, even desirable, to change only some components of a system in order to promote helpful evolution. This raises the more fundamental question of defining boundaries. In any innovation, some aspects of the system will be inside the boundary and will be acted on directly as part of the change. Others will be defined as outside the boundary: There will be no attempt to change these, and so the innovation will have to work with them as fixed influences, or perhaps with the way they are changing outside the control of the innovation process.

For educational change, an important element that is always outside the boundary is the social and political climate within which changes occur. Thus, it is not possible to understand the evolution of the system of national curriculum and assessment in Britain without studying the evolution of educational thinking within the ruling Conservative party and the shifts in power between different ideological camps within that party. Similarly, in Norway, a central plank of the policy is that schools must serve the social integration of the nation—so that division by ability into separate school classes cannot be contemplated. Again, the very different strategies for reform between France, Spain, and the United States can only be understood in the context of their specific political traditions, the differences between nationally central and federal systems being a particularly relevant feature. Pressures for improved economic competitiveness, or to meet the needs and expectations of the world of employment, or for schools to be agents to promote desirable social change, or reverse the undesirable, are other common out-of-boundary effects. All of these must usually be treated as outside the boundary within which the educational system has to work—a reform that aimed to try to change any of them would be a very different enterprise from one which tried to adapt to them as they are.

It is important to stress these aspects because any assessment plan has to allow for them. Thus, for example, an ideal design that will arouse political or public opposition might be less effective than a more modest alternative with better prospects of adoption.

The Technical and Political Dimensions

Assessment—As Spur, As Stimulus, and As Support

Many educational writers assume that assessment led reform is sure to be unhelpful. There may be a confusion here—for clearly a first priority in any innovation is to formulate its educational aims. Given these, then assessment and testing can be powerful tools in securing the realisation of educational aims, not only, nor even mainly, by exerting pressure on reluctant schools and teachers, but rather by expressing the intentions in concrete operational terms so that the implications for pedagogy and for learning can be clearly understood.

Of course, negative possibilities abound—as shown by the reactions in the United States against the constraining effect of the standardised tests that have been used in some state assessment systems. The testing tool causes concern precisely because of its power, and it has too often been deployed—in part inadvertently—to impose educationally undesirable pressures on the curriculum and on the aims of teaching. Where the instrument is used in a context where teachers are not trusted such effects are more likely than where it is used in a context of negotiation with the teaching profession.

The stimulus through assessment need not only operate through testing of all pupils in conditions set externally. For example, in Scotland the government provides national banks of items with guidelines for their use, but it is up to teachers to decide when best to use them. National monitoring can serve many accountability purposes by sampling rather than by imposing on all pupils—and by so doing can deploy a greater variety and range of items which can subsequently be disseminated for wider use.

The Technical Imperatives—Authentic Assessment and Reliability

Little was said in the meeting about technical issues. However, they hover over all discussion because they can impose strong constraints which are too easily ignored. The outstanding problem with external summative tests is the severe limitation of their reliability and validity. Any test which is to be applied to whole populations can only be affordable if it is short and uses a

restricted range of techniques—notably written papers composed of multiple choice items or items allowing for only a very limited response. The claimed reliability of these is often based only on the internal consistency of responses and does not allow for many factors which cause variability in students responses. Bias in tests is also hard to correct when students are only able to respond in one mode within a narrow range of contexts.

The limited validity of tests is and even more serious problem . It is simply not possible, using as data only the results of reductionist types of tests, to make valid inferences about a student's capacity to (say) produce extended prose, communicate orally, pursue an investigation of an open problem in science or in mathematics, or design a technological artefact. Yet these capacities are often more relevant to future study and employment than those which reductionist tests can mirror. The emphasis in Hans van Aalst's paper points to the need for assessing these broader abilities. However, such assessment is expensive, and has to be extended over many tasks and so over long periods if the results are to be generalisable.

More particularly, if tests are to raise standards, they would be more effective if they were to have a direct affect on the processes of learning and not merely try to work through the backwash effects of summative tests. This means that assessment as a tool for reform might better be seen through its use in formative mode—as in the French initiative described above. A major technical challenge to assessment studies is find ways of resolving the problems of reliability and generalisability in the area of "performance" or "authentic" assessments, particularly where these are embedded in normal classroom work with the intention of using the teachers' assessment for summative purposes as well as for formative help.

Public and Political Understanding

Defects in the public understanding of schools seem to be a matter for concern in many countries. One comment was that classroom realities in schools are very similar across different countries whereas the political and public rhetoric about education can differ widely. A clear obstacle to pursuing some desirable paths to change is that the assumptions on which they are based are not understood, let alone accepted, by politicians and the general public. Thus, in some cultures teachers' own

assessments are not trusted and there is no will to invest in improving their reliability or in working to calibrate standards between schools. In most cultures, a quite unwarranted faith is placed in external tests, while hard-headed technical objections to them are treated as typical defensive evasions by the professionals. Most serious of all, the invalidity of a large proportion of assessment and testing results is not appreciated—if it were, many aspects of present practice might be seen to be indefensible.

Since testing is something that many of the public have experienced and do accept, those seeking any fundamental change are up against a particularly difficult obstacle. It seems clear that educators have to devote more of their effort to raising the level of public understanding of assessment and testing: It will take a concerted effort over many years if such a project is to achieve its aims.

Such issues have to be considered in a wider context of the public expectations of the role of the school. Some look to the schools to preserve present or past culture, others expect them to help society adapt to a world that is changing ever more rapidly, some even seem to want them to pursue both of these aims. There are also different beliefs about the best way to assure that schools fulfill their role—a general, but not universal, approach is to move away from prescription of means but to strengthen accountability for outcomes. Such choices cannot be isolated from general political beliefs about social institutions—which can be unhelpful if they try to treat schools according to some universal model. One notable example here is the application of the market ideology to education by schools. To quote the personal perspective of a participant from Holland:

> It is a misunderstanding of the market that competition leads to higher quality. Examples from industry show that too high a quality leads to bankruptcy. A company is not served well by high quality, but by high turnover/cash flow and sufficient profits. Competitiveness leads to better wrapping for products and marketing strategies. For example, Dutch companies like Philips and Fokker always had a quality strategy. Philips went almost bankrupt but changed its strategy in time, Fokker did not and collapsed.

Neither can the school be treated like a commercial enterprise. Managers of a production industry have a first responsibility to the shareholders who appoint and fire them—power flows naturally from the top. It is at least arguable that school managements have the interests of pupils as their first responsibility—but power cannot thereby flow from the bottom. The optimum model for schools must be unique to them, not a copy taken from the successful management models of commerce. Nevertheless it might be worth pointing out, in the context of assessment, that manufacturers have moved away from quality control by product testing to the use of quality circles to ensure participation of all along the production route in ensuring quality outcomes. The educational equivalent is a shift of emphasis away from summative testing to formative assessment, and might be helpful for similar reasons in the two cases.

Students, Teachers, and the Burdens of Change

Hurrying Slowly—The Pace of Educational Change

A common lesson that emerges from studies of educational change is that significant changes in classroom practice can only be brought about slowly. Given the complexity of any teacher's task in managing the classroom, this is hardly surprising. Matters might be more straightforward if those promoting any educational change were to know exactly what needed to be done to translate their ideas into practice—but this can hardly ever be the case. Any new idea that aims to make a significant impact on practice has to be tried out and then reformulated as the complex realities expose unforeseen implications. Thus, there has to be a process of iteration in order to fashion a workable innovation. Moreover, success in an experimental situation with teachers who have direct contact with the source of innovation does not guarantee success when the innovation has to be disseminated on a wider scale. After initial and thorough trial, there has to be a slow and carefully evaluated build-up of implementation on a wider scale. Thus, the natural time scale for effective innovation is probably between five and ten years—a reality which politicians find difficult to accept.

Students in Assessment

Reforms in assessment give rise to heavy burdens. Any policy should aim to share out these burdens so that they do not fall entirely on teachers. This is one, among several, of the reasons why the role that students might play in their own assessment is emerging as a new and important issue. It seems that students—at least at secondary school level—are capable of assessing themselves in ways that help them to take more responsibility for their own learning. In addition, this innovation can help make it possible for teachers to cope with the daunting task of keeping track of the wide ranging and changing needs of a classroom full of students. However, as with the case of teachers' formative assessments, it is not easy for students to grow into this new role. They have to be guided. They have to first come to understand the aims of their learning and the criteria for judging success in those aims. There is evidence that most students lack such understanding and move through learning routines without grasping their structure or their purpose. Given such grasp, they then have to learn to be realistic in judging themselves, and come to see that they disadvantage themselves if they make judgements that are either too optimistic or too pessimistic. Developments in these directions will make assessment practices fully consonant with, and capable of enriching, the main thrust of improvements in learning methods which have derived from constructivist models of learning. Such models stress that, to be effective, learning has to be meaningful and so has to be based on and develop from the individual's own set of understandings and meanings—not superimposed on these as a meaningless and disconnected overlay.

The Sine Qua Non—Teachers in Classrooms

The focal point for any improvement in pupils' learning is the classroom—yet what happens in the classroom is hard to control or prescribe. It is paradoxical that many reform plans which clearly depend for their success on changes in the classroom seem to have little to say about what their changes imply or hope to achieve at that level. So it is left to others to forge—within the complexities and pressures of classroom life—the essential link between the changed external conditions, conditions that the

reformers have designed both to provoke and to support achievement of their aims, and their actual achievement. This task is almost always difficult and, if the plan is unrealistic, may be impossible. The chances of success may be slight unless teachers fully grasp the aims and means, and share a commitment to them.

The commitment of teachers is an essential condition. There are many examples of state or regional reforms which have withered, either because teachers simply ignored the demands of change, or because they found them impossible to achieve within the conditions that accompanied them. Thus, in addition to their involvement, it is essential that teachers be supported, particularly by being given the time and the peer support that they need when they are trying to solve classroom problems that noone outside could reasonably have anticipated. In such situations, teachers themselves are the only ones who can fashion a solution.

All this is true for reform in any aspect of teaching and learning. In the present case, the reform of assessment, the need to satisfy these conditions is unusually stringent. This is so in the case of formative assessment because current practice is so far away from what can be achieved and yet achievement may require a change in habits of working built up over many years in a career. It is true also for the summative aspect because the interactions between the different purposes of assessment can create difficult tensions, and the optimum mix between external agencies and pressures and teachers' own areas of responsibility is hard to achieve. It is essential here to have understanding and mutual respect in the partnerships that should be evolved.

However, just as formative assessment is not to be conducted by schools in isolation, so its results are not solely for classroom use, as the French innovation referred to above exemplifies. These results can be important for planning at school and at regional or national levels. This would be more evident if a culture of evaluation for improvement of teaching were as firmly established as it ought to be.

Readiness to envisage such a contribution might help in the difficult task of convincing politicians and parents of the importance of formative as opposed to summative assessments. Politicians in particular tend to back away from government involvement in formative work partly from lack of understanding, partly from a belief that they should raise

standards by reliance on external pressures, and should avoid "interference" in the classroom. Insofar as this is done, formative and summative assessment policies are separated, to the detriment of both. However, the corollary of this argument is that teachers must accept that their government's right to play a role in classroom assessment goes with its responsibility to support it.

Who Initiates, Who Drives, Who Supports?

Reforms can originate with teachers, in which case they will have a classroom focus and will be associated with professional development and trust in the profession. They may be driven by formulation of standards and by testing; then they will have a focus on attainment outcomes and with concerns for effectiveness and—possibly—equity. A third impetus may come from new curriculum frameworks, which will focus naturally on the epistemology and educational roles of each subject and may lead to concern with classroom resources and with pedagogy. Any one of these or other approaches may be valuable, even optimum, if they are well matched to their particular systemic contexts. There is no single ideal route or method for systemic reform.

A system which can allow for and can support initiatives which arise at different levels and for different agencies is likely to promote a richer evolution than a system where reform can only come from some central initiative so that creativity of the many stakeholders is inhibited. One difficulty with centralised and comprehensive reforms is that they are likely to be seen by teachers as alien impositions—then the inevitable problems of implementation will be causes for complaint and blame rather than stimuli for creativity in improving the innovation idea. The process of innovation ought to be so managed that it builds on strengths and recruits commitment from teachers. If, for example, a government succeeds in a struggle with teachers to impose a new scheme, they may have won a battle but lost the war.

Thus there ought to be a shift from seeing a systemic reform strategy as one of constructing the master plan for all aspects of education to one of creating a culture of innovation within which all stakeholders feel they can contribute and to which they develop a natural commitment. Researchers, trainers,

publishers, administrators, and, notably, both teachers and students have to be considered in such an approach. If any of these is alienated, the chances of successful change are thereby reduced.

APPENDIX A

INTRODUCTION TO COUNTRY OUTLINES

The nominated delegate from each member country submitted a brief paper which covered these three broad conference themes:

Formative Assessment for Learning

The role of assessment of pupils in the improvement of their learning and the current state of professional practice in implementing formative assessment in the classroom.

Summative Assessment

The use of assessment of pupils for pedagogical purposes; for certification, awarding of diplomas, transfer of pupils between stages of education; and for monitoring purposes.

The use of assessment to monitor and evaluate both individual institutions and the education system as a whole.

Systematic Reform

National policies on curriculum and assessment.

APPENDIX B

COUNTRY OUTLINES

AUSTRALIA

PETER WHITNEY, Department of Education Employment and Training, Canberra

Under the Australian Constitution, the responsibility for school education is shared by the state and territory governments. This includes the development of school curriculum and assessment policy and practices.

Assessment of student learning is mainly carried out at each level within individual schools, except in the final year, Year 12, where a mixture of school-based formative assessment and external summative assessment provides a basis for secondary school exit statements. Each state has developed its own assessment practices for Year 12 though results are recognised across states.

Commencing in the late 1980s, the federal government has played a leading role in encouraging national collaboration to assist schools and systems to develop specific objectives and strategies on school curriculum and assessment matters. The development of a first edition set of curriculum statements and profiles in eight key learning areas is one of the major national initiatives that has provided the impetus for Australian systems, schools, and teachers to review their approaches to curriculum and assessment design and implementation.

A curriculum profile for each of the eight key learning areas was completed in June 1993. The profiles have been used by education systems across Australia throughout 1994 and 1995 as a basis for reviewing their own curriculum and assessment practice.

Another major initiative is the implementation of the key competencies, developed in 1992 by representatives of the Australian education and training sectors, business, industry, and unions. The key competencies are the generic abilities that industry and education considered essential for effective

participation in emerging forms of work and for work organisation, further education and training, and adult life generally. They have been trialed throughout education systems since 1994.

Formative Assessment for Learning

Role of Assessment in the Improvement of Pupil Learning

The implementation of the curriculum statements and profiles in Australian schools has led to a general acceptance, by systems and teachers, of the merits of an outcomes-based approach to assessment. Throughout the trialing of the profiles, many teachers indicated that the strengths of the profiles are that they offer the benefit of a shared language for planning courses and for describing and reporting student achievement, and make student achievement explicit.

Many respondents to the trials thought that the use of the curriculum profiles to assist assessment and reporting will provide valuable feedback about learning outcomes, which would lead to the achievement of better student learning outcomes.

The curriculum profiles for Australian schools provide a progress map against which student achievement can be measured and monitored. The profiles describe the progression of student attainment of knowledge, skill, and understanding through eight levels. The curriculum profiles do not prescribe particular instruments and procedures to estimate or determine a student's level of attainment as they progress through the levels of knowledge, skill, and understanding. They do, however, promote the use of a range of assessment strategies and encourage teachers to rely more on formative assessment and less on summative assessment.

State of Professional Practice in Implementing Formative Assessment in the Classroom: Initiatives to Promote Assessment Using the Curriculum Statement and Profiles for Australian Schools

The federal government recognises that effective implementation of the curriculum profiles to improve the quality of learning outcomes of all students means that teachers, in

particular, need considerable support in using the profiles for assessment and reporting purposes. Professional development support for the implementation of the curriculum statements and profiles is being provided by the federal government through the National Professional Development Program.

Some federal government initiatives to promote assessment strategies to support the use of the profiles approach are listed below.

Assessment resource kit. The Australian Council for Educational Research (ACER) is developing a resource kit for schools and teachers that will:

- draw on a range of national and international experience and research in regard to assessment strategies which support the implementation of profiles;
- identify principles of best practice;
- develop units illustrating sample assessment tasks and approaches relating to a range of curriculum areas;
- address issues related to assessment of students from disadvantaged settings.

The project will publish a set of ten assessment guides and a video on different assessment methods and issues for teachers in using them. The assessment guides will be published by June 1996 and provide practical assistance to teachers. They cover:

- Assessment Methods;
- Developmental Assessment;
- Inferring Achievement;
- Paper and Pencil Tests;
- Performances;
- Portfolios;
- Products;
- Progress Maps;
- Projects;
- Recording Observations;
- Reporting Progress.

Assessment and use of profiles professional development project. This project will develop a series of professional

development materials to support the use of the Assessment Resource Kit and for use in pre-service and in-service professional development programs. It will be informed by the outcomes from the above research project. The professional development project commenced in December 1995 and will be completed by December 1996. It is expected that the project will look at issues of assessment and reporting in depth and make connections between methods, purposes, and contexts for assessment.

Assessment and use of profiles research project. The federal government has commissioned the Australian Council for Education Research to prepare a research paper addressing:

- The current situation for teachers in the implementation of the profiles, including how they are using them, their understanding of them, and problems they may be experiencing with them;
- The use by teachers of standardised and diagnostic testing in classrooms;
- How key competencies are embedded in everyday classroom activities in the compulsory years of schooling, and ways in which they can be incorporated into quality assessment and reporting strategies;
- The use of the profiles for reporting of student outcomes to teachers, parents, and employers;
- Reporting practices and strategies that are being used effectively across Australia;
- The various technologies that have been developed to support implementation of the profiles.

National professional development program. The Australian federal government is supporting reforms to assessment practices in schools through the National Professional Development Program (NPDP). One of the major focuses of the NPDP is to support the implementation of the curriculum statements and profiles for Australian schools. The recent evaluation of the NPDP indicated that in 1994 and 1995 more than 50,000 of the 250,000 teachers in Australia had participated in professional development relevant to the curriculum statements and profiles, across eight key learning areas.

The effectiveness of the NPDP has been underpinned by productive partnerships among teacher employers (government and non-government), teacher professional associations, and universities. Funded projects have involved teachers directly in determining their own professional development needs, in action research, and in establishing on-going support networks. There is evidence that the scope and strategic targeting of the NPDP is encouraging teachers to embrace the change agenda and beginning to affect classroom practice.

Using Summative Assessment

The Use of Assessment of Pupils: Student Assessment and the Key Competencies

Since 1993, funding has been made available under the key competencies program, to support the development, training, and evaluation of the key competencies in Australia's general and vocational education settings. The key competencies are an example of how assessment can be used to report on student progress toward identified skills.

The key competencies are:

- Collecting, analysing, and organising information;
- Communicating ideas and information;
- Planning and organising activities;
- Working with others and in teams;
- Using mathematical ideas and techniques;
- Solving problems;
- Using technology.

An eighth key competency, cultural understandings is currently under development.

Assessment is one of the most complex aspects in the implementation of the key competencies and is still under consideration. A number of the projects funded by the federal government under the key competencies program are specifically addressing assessment and reporting issues and methods. The question is how to assess student performance against the key competencies in such a way that their capacity to broaden vocational education and training remains as important

as their capacity to provide a greater employment-orientation to the general curriculum in schools.

In each state and territory, the authorities responsible for assessment at the end of secondary schooling are closely involved with the key competency program pilots. At this stage, it is expected that reporting on student attainment against the key competencies will feature in all state and territory student exit materials in 1997.

The Use of Assessment to Monitor and Evaluate Both Individual Institutions and the Education System as a Whole

As the curriculum profiles become part of the school's assessment and reporting processes, teachers need to be concerned with making sound judgements about student achievement, for the planning of future classroom activity and for different reporting purposes. School principals and school communities will require information about levels of student achievement for use in school evaluation and planning. In addition, information on student learning outcomes may be sought as a basis for monitoring educational standards across a system.

Systemic Reform

National Policies on Curriculum and Assessment: National School English Literacy Survey

In May 1994, the federal government allocated funds to collect by the end of 1996 reliable national data on English literacy levels of school students at three significant stages of schooling. There have been no reliable data on the levels of English literacy attainment among Australian school students collected since 1980.

After consultations with school systems and authorities and the teaching profession about the nature of the data collection process, it was agreed data could be collected from a light national sample survey and that planning for such a survey be highly collaborative. A steering committee for the proposed survey was set up with members from the key stakeholders—state and territory education systems, peak non-government school authorities, teacher unions, professional literacy

associations, parent organisations, the business sector, and the federal government.

The steering committee agreed that:

- Teacher judgement be central to procedures to be trialed for collecting the data for the proposed survey;.
- Data collection procedures model good practice in assessing English literacy and enhance the professional skills of participating classroom teachers in the assessment of student performance in English literacy;
- The national English curriculum profile (and the way the profile is being implemented in schools across Australia) provides a useful framework for the proposed survey;
- Students in Years 3, 5, and 10 be assessed in the proposed survey.

During late 1995, a trial was conducted by the ACER to explore the feasibility of procedures for collecting data for the proposed survey. One of the fundamental purposes of the trial was to analyse how teacher judgment and the richness of classroom practice can become an integral part of student assessment programs.

After analysing the outcomes of the trial, ACER will be making recommendations to the steering committee about the methodology for the proposed survey. At its March 1996 meeting, the steering committee will decide whether the proposed survey is to go ahead in October 1996 and the nature of its methodology.

AUSTRIA

KARL HEINZ GRUBER, University of Vienna

In the ongoing Austrian discourse on education assessment and standards are not high priorities. (Indeed, there is no German word which carries the same connotations and implications as the English expression "standards.") The issues presently getting the most public and administrative attention are curriculum reform, teacher training, and more autonomy for the individual school. The assumption that Austrian pupils and school graduates perform well in international comparative terms is widely held (although lacking empirical foundation: Austria is a latecomer to the IEA evaluation; only through the Third IEA Mathematics Survey will the performance of Austrian pupils be exposed to international scrutiny).

Being part of the realm of the powerful German tradition of educational thinking, Austria shares an assessment culture which has great trust in the judgements of teachers and strong reservations about the measurability or "testability" of educational achievement. The holistic assessment of the individual pupil's work through the individual teacher has withstood, since the 1960s, massive empirical evidence showing the flaws in its reliability and its inequitable social bias.

In Austria, selection for different types of secondary schools begins at the age of 10, with a second major point of transfer to the differentiated system of upper secondary education at the age of 14. Despite the high-stakes nature of teachers' assessments in the process of early scholastic differentiation and despite their low dependability, there are at present neither plans to make assessment for selection purposes more objective and reliable nor plans to abolish or postpone selection to the end of compulsory schooling.

The Curriculum: Central Regulation and Teacher Freedom

The outstanding features of the Austrian practice of educational assessment are:

- Its high degree of centralised, nationwide regulation;
- Trust in its fair and objective implementation by the individual teacher;
- Absence of external measurement, testing, or even sample surveys of scholastic performance;
- Lack of empirical data in this area; many generalizations are unavoidably based on fragmented or anecdotal evidence.

Assessment and standards are comprehensively and elaborately covered by the *Schul-Unterrichtsgesetz* (SchUG), a major act that regulates all aspects of "the inner life of schools," the *Leistungsbeurteilungsverordnung*, an extraordinarily detailed ministerial decree on assessment and all its eventualities, and they are also built into the national curriculum (or, rather, the various curricula underlying the work of different types of school). The SchUG and the assessment decree prescribe:

- The range of and balance between oral, written, and practical assessment;
- The number and pacing of formal written examinations (usually six per year in the major subjects German, mathematics and the foreign languages);
- The grading scale from 1 (best) to 5 (failure) and the principles to be observed when applying it;
- The length and mode of oral examinations and the number and type of questions;
- The conditions under which unsatisfactory exams may be retaken;
- The consequences of partial or overall failure.

The national curriculum (*Lehrplan*), while describing in great detail the goals, aims, and content to be covered for each subject in each school year or grade, is officially labelled a "framework" (*Rahmenlehrplan*) within which the teacher has the professional responsibility, freedom, and discretion to choose, differentiate, and emphasize with respect to the abilities and needs of a given group of pupils. The potential for variation in the implementation a prescriptive national curriculum is considerable but a number

of factors contribute to the narrowing of the scope of what actually goes on in classrooms:

- Textbooks, which have to be approved by ministerial working groups, are related to the national curriculum; it is widely observed (and criticised) that the sequencing and implicit pacing of the textbooks constitute the "real" (albeit hidden) curriculum.
- School inspectors randomly check pupils' progress of work as recorded in class log books (*Klassenbuch*) and occasionally also check class examinations, especially in cases of parental complaints or when a teacher or a school show atypical grading behaviour.
- Critical decisions, such as the reports at the end of primary school (age 10), on which transfer to an academic secondary school depends, are not made by an individual class teacher alone but confirmed corporatively by the staff conference. In secondary schools any negative assessment that results in non-promotion, i.e., repeating a school year (*Sitzenbleiben*) because of failure to reach a satisfactory level of competence in one or more subjects, requires a collective decision by all the pupil's teachers.

Both initial and in-service teacher training are expected to equip teachers with comparable standards of expectation and assessment. There is some evidence that this may not be so; some teachers and some schools have a reputation for being either "tough" or "soft," and in some schools a high proportion of pupils resort to (expensive and therefore socially divisive) private tutoring to keep up with teacher demands. Assessment in Western Austria seems to be stricter than in the East of the country, therein matching the more rigorous assessment practice of neighbouring Southern Germany.

The focus on the quality of individual schools and the variation among them is a fairly recent Austrian phenomenon. It was triggered not only by an OECD-induced serious scientific concern for school quality but also by some unsophisticated, questionable school ranking attempts by two newspapers. The

recent publication by the Vienna School Board of a list which ranked academic secondary schools in Vienna on the basis of their 1995 failure rates angered many teachers because it was based on raw data without accounting for pupil input and value added, which—given the unsatisfactory data base on Austrian school system performance—is not yet possible.

Formative and Summative Assessment

The aims and purposes of assessment are in Austria the same as in other countries: To diagnose the strengths and weaknesses of pupils and thereby motivate them; to provide teachers with feedback on the efficiency of their work; to certify the pupils' (probable) ability and entitlement to move on to the next stage in their educational careers; and to inform parents, employers, and the world outside the school system.

The practice of making a pupil repeat a year when he or she fails to reach a satisfactory level of competence intertwines the formative and the summative aspects of assessment. "Staying down" has become rare in primary schools. Instead of using the 1 to 5 grading scale, a number of primary schools are piloting a "verbal assessment," which they hold to be a more suitable mode of formative assessment by giving individualized and informative feedback to pupils and parents. In the non-selective secondary modern school now only about two percent have to repeat a year, which is largely due to the introduction of ability setting in German, mathematics and English. In upper secondary education the failure/repeating rate is still approximately ten percent annually.

The final secondary school certificate, the *Reifeprüfung* or more popularly *Matura*, is the high-stakes form of summative assessment. Its standards are monitored three ways: the examination papers have to be submitted beforehand to the regional school inspectorate, samples of the teachers' grading are checked by the inspectorate, and the viva voce part of the exam is chaired by an inspector or the principal of another secondary school. Many teachers feel that it is not just their pupils who are being examined but they themselves.

Although the majority of young Austrians recognize the importance of a good education, the school system as such does not do much to emphasize "the pursuit of excellence." Annual grade reports and *Matura* acknowledge outstanding

performance, but entry into universities is non-meritocratic, non-selective and non-competitive. Any kind of *Matura* certificate—even rather poor ones—entitles the student to the same free choice of university and field of study. Only the newly established institutions of the non-university *Fachhochschul* sector with their limited capacities operate a system of selection.

The consequences of the existing assessment policies with respect to the goal of equity are difficult to prove because the official Austrian school statistics are socially blind, i.e., they do not categorize children, their educational performance, and their school careers according to parental occupational or social status. There is worrying evidence that a disproportionately high number of children from non-German speaking ethnic minorities or migrants families are concentrated in special schools. On the upper end of the educational spectrum it has become evident that despite the massive expansion of upper secondary education the percentage of first year students from working class families has not significantly increased over the past 25 years. (As in Germany, it is middle class girls who have profited most from the growth of extended secondary and higher education.)

Outlook

The ongoing Austrian curriculum reform—although presently focused on the introduction of partial curriculum autonomy of the individual school, on the reduction of "subject matter overload," and on a shift from knowledge accumulation to cross-curricular competencies or key skills (*Schlüsselqualifikationen*)—looks like re-accentuating the prevailing culture of assessment. The idea of improving ("professionalizing") teacher-centred assessment and supplementing it with system-wide sampling for calibrating and monitoring purposes seems to gain acceptability as an option for the future.

BELGIUM: THE FLEMISH COMMUNITY

PETER MICHIELSENS, Ministry of Education, Brussels

Policies on assessment in the Flemish educational system will only be transparent after some key concepts have been clarified, because they explain some recent evolutions in the system.

Freedom of Education

Freedom of education is permitted for those who provide education. Individuals or various bodies may organize one or more schools, provided they meet certain requirements and they are willing to be responsible for their pedagogic project. Consequently, parents and students can choose from an extended and varied range of schools.

This freedom of organizing schools combined with the freedom of school choice results in networks of schools. Public schools, organized by the community, represent only 13 percent of all secondary schools. Official schools organized by cities, communes, or provinces represent 12 percent. The greatest network of schools is the Catholic network : 75 percent of all secondary pupils attend Catholic schools.

The networks and their schools have a large autonomy. They can freely shape their pedagogic projects as they like.

The influence (direct and indirect) of these networks on educational policy is very important and gives an explanation for some key options about assessment and curriculum policy.

Compulsory Education and Consequences

The Compulsory Education Act compels parents to have their children taught at an official or private school for primary, secondary, or special education. School attendance is compulsory for a 12-year period. As compulsory school attendance was extended to everyone, secondary education is now for everyone.

Formative Assessment for Learning and Using Summative Assessment

The way in which the programmes for secondary education are built is a very important consequence of our compulsory school attendance. The essence is a very special relation between government and schools, expressed in specific procedures.

First: Teaching practice is a responsibility for the school. The government has no right to evaluate the pedagogical methods and the didactic practice. Evaluation is an affair of the school itself. The government does not intervene in assessment procedures. Instruments for assessment are elaborated by the teachers themselves. This implies the need for good, professional teachers. Therefore a reform of teacher formation is planned. New ideas about in-service training are also implied in the teacher-formation reform.

Second: As compulsory school attendance was extended to everyone, basic education is fundamental to our educational system. Schools have a large amount of freedom in how they realise and evaluate the aims of basic education.

Linked to basic education is the concept of attainment targets, which have to be achieved by every pupil. Attainment targets concern subject matters and cross-curricular goals such as social aims, learning to learn, environmental education. The attainment targets are points of reference for quality control by the inspectorate.

Third: The government has to guarantee and ensure that minimal aims are met. Therefore, schools have to present their programmes to the inspectorate for approval. Programmes have to contain the attainment goals as well as school-specific targets and contents. Only after approval can the schools implement their programmes. Moreover, the inspectorate has to verify whether schools do accomplish the approved programmes.

Systemic Reform

Context

Quality of education refers to the specificity of each school. Free choice by parents of a school implies that schools have a specific project.

A pedagogical project contains more than curriculum contents. Social attitudes, emotional attitudes, or specific values make one project different from another. This pedagogical difference cannot be object of quality control on the part of central authorities. This pedagogical difference has to be object of regular self-evaluation.

The realisation of essential curriculum contents and minimal social aims offers legitimacy for the efforts and financial investments by the government in education.

The realisation of essential curriculum contents and minimal social aims on one hand, and the realisation of the pedagogical project on the other hand can only be guaranteed in an effective school.

Quality Control

In Flanders, only attainment targets are used, which are controlled by the inspectorate, in addition to the school diagnosis. In contrast, assessment and standards of control for subjects are always school- and teacher-bound.

The role of the inspectorate is to provide an independent audit of the quality of education in general, and of the effectiveness of individual institutions (schools). Inspections are designed to inform education authorities and governing bodies about the quality of the schools for which the inspectors are responsible (the report is to both the head of school and the authorities), and to inform the minister of education about the performance and needs of the education system as a whole (the annual general report).

A school audit results in a report about the school the inspectorate has visited. It is considered an important means of improving quality in education. The concluding part of a school audit report contains one of three conclusions about the school:

A positive report means that the school will be financed or subsidized. It also means that the school's certificates or diplomas are recognised as official documents: an audit resulting in a positive conclusion has a civil effect on the certificates and diplomas. In Belgium (both communities) central examinations do not exist. In the Flemish community, certification is completely the responsibility of the schools: The report from the inspectorate, based on a school audit, is therefore a key concept.

A positive report with reservation clearly formulates the main shortcomings and also fixes a deadline. At the end of the established period of time, a supplemental investigation will take place to determine whether the shortcomings have been corrected. Those deadlines should be realistic and attainable: before the end of the school year following the audit or ten months after the audit.

A negative report will lead to a second examination by the inspectorate. If the second examination leads to another negative report, the school will no longer be financed or subsidized. And, in this case, the school has to close. As new certificates or diplomas will no longer be recognized.

In the case of a positive report with reservation, follow-up must be organized by the school itself. In the case of a negative report, the inspectorate will organize a follow-up.

Each year there is a statement to the Minister of Education, in which the inspectorate produces a general report on important aspects of education, identifying strengths and weaknesses and making recommendations for improvement. This statement is based on the results of all audit reports. This annual statement also contains the results of other specific inquiries made by the inspectorate. This report, which is transmitted to the Flemish Parliament, gives a yearly portrait of the situation of Flemish education.

The connection between school audit results and the annual report of the inspectorate proves how—in an educational system where central assessment procedures are not possible because of the political traditions and context—policy indirectly can be influenced by the situation in educational field.

CANADA

NORMAN MAYER, Ministry of Education and Training, Winnipeg, Manitoba

Each of the ten Canadian provinces and two territories holds individual responsibility for education and training. Canada has no federal education department or authority, although the formation of the Council of Ministers of Education Canada (CMEC) in 1967 has provided provinces with a collective national voice in education, noting that certain trends and emphases are common across the country.

The CMEC facilitates, coordinates, and promotes national and interprovincial initiatives between and amongst provinces. Cooperative efforts are normally entered into through formal protocol agreements signed by participating jurisdictions. In cases where international education-related activities are involved, an agreement between CMEC and the federal Department of Foreign Affairs guides Canadian participation.

While education is purely a provincial matter, there have been a number of initiatives to promote interprovincial cooperation and collaboration on a number of fronts. The four Atlantic provinces formed the Atlantic Provinces Educational Foundation (APEF). The four Western provinces and two territories recently signed a Western Canadian Protocol intended to facilitate joint work in a number of areas. More recently, a Pan-Canadian protocol, modelled on the Western Protocol and the APEF agreements, has been approved by ministers to facilitate work that might be undertaken jointly by any combination of Canadian provinces and territories. Two areas were chosen for initial development: science and information technology. So far the main work done through these arrangements has been on the development of a common framework of science outcomes for K–12.

At the present time, there are numerous assessment programs in place across Canada, some for secondary graduation and certification purposes, and others for curriculum, program, or system evaluation purposes, as can be seen in the details that follow this section. At the national level, in 1989, CMEC

undertook a major national assessment effort for 13- and 16-year-olds in mathematics, reading and writing, and science. These tests are administered on an annual, rotating basis to a sample of students from each province and territory. The information is to be used for program analysis within the provinces, and results are made available only at the provincial/territorial level. The assessment instruments used in the School Achievement Indicators Program (SAIP) are developed to measure a wide range of goals in the subject areas tested. Performance-based assessment is used in addition to short answer and multiple choice testing questions. Because SAIP is a national program, great efforts were taken to see that state of the art theories and practices of assessment were applied as much as was reasonably possible.

At the international level, Canada is participating in the Third International Mathematics and Science Study. Five provinces and one of the territories are also participating as independent jurisdictions. Canada is also a major contributor to the work of the OECD/INES project on the development of student performance indicators. In addition, a network of provincial or territorial contacts has been set up to assist with Canadian participation on the Steering Committee and Area 3, "Systemic Analysis of Assessment" of the OECD project "Teachers and Curriculum Reform." CMEC also participates in the Asia Pacific Economic Cooperation (APEC) education forum work in performance measurement of the education sector.

Discussion

Generally, provincially developed testing systems serve one of two purposes. Assessment systems are generally designed for monitoring, evaluation, or program improvement purposes. Reporting is usually done at the school, division or district, or provincial level although individual students may learn their results. Examination systems are designed to award marks to individual students as part of their overall course mark. While schools generally receive a summary report on their own students' results, the extent to which these results are made available to the public varies widely across the country.

Canadian testing and assessment programs are characterized by a high degree of teacher involvement. Tests are normally developed by committees of teachers. The degree and

extensiveness of piloting varies, with some jurisdictions piloting all items before final administration and other jurisdictions placing greater reliance on professional judgement to develop quality items.

Tests are also normally marked by trained teachers. Most provinces conduct their marking on a centralized basis although one or two provinces do some or all of their marking on a decentralized basis. The weight assigned to exit exams is from 30 to 50 percent, with quite a bit of variation across the country. Whether or not marks on examinations are adjusted varies from jurisdiction to jurisdiction.

The number of subjects tested as part of an exit exam system also varies quite widely. Some provinces have exams in only one or two subjects whereas others include most of the academic courses found in the final year of high school. The extent to which the student body is affected also varies. In some subjects in some provinces, all students enrolled are required to take the exams. In some subjects in some provinces, only those contemplating attending a post-secondary institution are required to take the exams. In one province, only students taught by a teacher lacking a subject accreditation are required to be tested.

Many provinces are moving toward a standards- or outcomes-based approach to education. Outcomes or content standards are statements, expressed in measurable or observable terms, of what students are expected to know or be able to do. Performance standards, called benchmarks or simply standards in some provinces, are statements of how well students are expected to perform in relation to outcomes or content standards. Beginning efforts are just underway in a few provinces to design tests in accordance with such an approach to defining educational expectations.

CMEC, in collaboration with Statistics Canada, has undertaken development work on an indicator resource or database. Six indicator areas were chosen for initial development: Accessibility, student flows, school/work transitions, achievement, citizenship, and satisfaction. These areas represent important goals for Canadian education systems. They are not mutually exclusive, but interconnected parts of a whole that together provide a comprehensive, coherent picture of student and systems performance. It is expected that SAIP and the Third International Mathematics and Science Study (TIMSS) data will be used as part of the achievement component. Many

provinces, in doing development work on their own education information systems, are mindful of the need to provide data for the pan-Canadian indicator resource.

The General Picture in Canada

There is a great deal of diversity within Canada in terms of subjects and grades tested. There is also considerable diversity in the relative emphasis placed on testing for assessment purposes and on testing for examination or grading purposes. Nevertheless, some generalizations can be made about the nature and amount of testing occurring in Canada. Seven provinces and two territories use some form of high school provincial examinations. The Northwest Territories uses Alberta's high school exit exams and the Yukon uses British Columbia's high school exit exams.

All provinces and territories except Prince Edward Island, Manitoba, and Northwest Territories have provincial assessments at various grade levels for program and system evaluation purposes.

All provinces except Ontario administer some form of the General Educational Development (GED). Quebec has been authorized by the Educational Testing Service (ETS) of Princeton, New Jersey, to develop an adapted version of the GED—the Tests d'Équivalence de Niveau de Scolarité (TENS).

All provinces except Saskatchewan participated in the mathematics and reading and writing assessments of the School Achievement Indicators Program (SAIP). Saskatchewan chose to concentrate on its own indicator and assessment programs but retained observer status. Saskatchewan is presently a member of the SAIP science consortium.

The Canadian Tests of Basic Skills (CTBS) is mandated on a provincial/territorial level only by Newfoundland and the Yukon. In the provinces of Alberta, Saskatchewan, Manitoba, Ontario, Nova Scotia and the Northwest territories, districts, boards, schools, or teachers use it at their own discretion.

The Canadian Achievement Test (CAT) is used at the discretion of districts, boards, schools, or teachers within British Columbia, Alberta, Saskatchewan, Manitoba, Ontario, Nova Scotia and the Northwest Territories.

Ontario, New Brunswick, Alberta, British Columbia, Newfoundland and the Yukon are currently committed to take

part in the Third International Mathematics and Science Study (TIMSS). A Canada-wide sample is also being tested.

THE CZECH REPUBLIC

PETR DRABEK, Czech School Inspection, Prague

After 1989, in spite of developing new elements in our educational system, the traditional pupil assessment by the teacher remains the only one. This internal pupil assessment is used for both formative and summative purposes.

Formative Assessment

The purpose of the formative assessment of pupils is to help them to improve their performance, to support and guide them in a positive and encouraging manner and motivate them in their studies. The pupils are entitled to receive feedback on their progress, study skills, and results. The formative assessment is also used to give the pupils feedback on their learning process.

The purpose of feedback is to promote the pupils' growth and development in the direction defined in the curriculum, and to strengthen the pupils' self-esteem and to support their learning process and recognition of their own abilities and skills.

Formative assessment is regarded as a regular part of teaching and learning. Pupils are assessed for their everyday class activity and performance, and for their homework. This may be assessment either by marks or by verbal assessment.

Summative Assessment

The purpose of the summative assessment is to appreciate pupils' attainment at a fixed point and to provide information about their achievement to pupils and to their parents. The behaviour of all pupils and their knowledge, skills, and progress in each subject have to be evaluated twice a year. The assessment of pupils is based on the objectives and on the subject-specific requirements as defined in the curriculum. Assessment and marking (in accordance with the grading code) of pupils is regulated by a state-wide decree of the Ministry of Education. It deals with educational measures, assessment, and marking of pupils, complex assessment of pupils, re-

examinations, and re-examinations overseen by an examination board. Pupils are given official reports in the middle and at the end of each school year. Their achievement in each subject is expressed numerically (by a 5-point scale) or verbally (an alternative used in elementary classes).

The selection of pupils for different types and levels of schools is another purpose of summative assessment. Nevertheless, the pupils have to pass entrance examinations to be accepted to secondary and tertiary schools.

The summative assessment at the end of upper secondary (mainly vocational) school is also used for certification.

An issue yet to be resolved in our country is a possible use of external assessment—e.g., of uniform, nationwide tests—mainly for graduation examinations, in order to increase the equality of educational opportunities. Results of nationwide assessment should also be used for the evaluative purpose—for evaluating the effectiveness and efficiency of education at the level of both the schools and the educational system. Nowadays, some private agencies offer their tests on a commercial basis, but there is no official information about how many schools are using those tests. Czech school inspection uses its own tests for evaluation of schools and for the assessment.

Since 1991, the Czech Republic has participated in IEA projects. Altogether, 510 schools with 810 classes and 18,000 pupils of three populations (9-year-old, 13-year-old, and students of secondary schools) participated in the TIMSS project. The Ministry of Education has also decided to replicate an earlier IEA project, the Reading Literacy Study. The sample included 6,000 pupils from 140 schools. Both projects were finished last year. Finally, the Czech Republic participates in two new IEA projects: the Language Education Study and the Civic Study, which have just commenced.

Systemic Reform

After 1989, a long-term transformation of our educational system began. It has led to enhanced diversity and plurality, and schools have been granted a considerable degree of autonomy and responsibility. Curricular policy is based on the cooperation of the centre and schools. The state will determine educational standards for primary and secondary schools (nowadays, some of them are checked by the Ministry). Standards will serve as a

point of departure for the development of educational programmes and external assessment tools .

The state will support the development of other educational programmes and will efficiently back up efforts of schools to work out or adapt educational projects, thus helping to enhance the pedagogical autonomy of schools.

The key requirement of the transformation is quality assurance. It is necessary to apply evaluation in different forms at all levels (pupil, teacher, school, educational system). It is limited by insufficient qualification of teachers and school managers for the development of their own curricula. Therefore, it is necessary to support the development of the modular educational programmes as a base for the pluralistic offer that can help develop the educational autonomy of schools. This will also help to make comparisons in the regional and international context.

DENMARK

JØRGEN BALLING RASMUSSEN, Ministry of Education, Copenhagen

Introduction: A Brief Description of the Danish Education System

The Danish Parliament (*Folketinget*) pass the laws which regulate education. The laws outline the framework of the different levels of education, but a growing number of decisions must be made at lower levels. These levels are country councils, municipal councils, and the individual educational institutions. Within the framework of the law, the Ministry of Education has the main responsibility for education.

In very broad terms the Danish education system can be divided into three main categories.

Primary and Lower Secondary Schools (Age 6 –16)

The basic school, a nine-year comprehensive school; there is no streaming. The basic school (*Folkeskolen*) is run by the municipalities.

Youth Education (Age 16–19)

Upper secondary education is offered in different forms and at different types of institutions.

General upper secondary education is offered at general upper secondary schools. The *Gymnasium* prepares students for the upper secondary graduation examination. The *studentereksamen* courses prepare students for the higher preparatory examination (HF). General upper secondary education is run by the counties.

Technical and commercial upper secondary education is offered at vocational schools (commercial and technical schools) in preparation for the higher commercial examination (HHX) and the higher technical examination (HTX). Vocational education and training is offered at the same vocational schools.

The basic structure of the vocational training scheme is based on the alternation ("sandwich" course) training principle

consisting of a combination of theoretical and practical training at a school and practical training in a firm. The vocational schools are private foundations financed by the Ministry of Education.

Post-Secondary or Higher Education (Age 19+)

The universities and other higher education institutions are run and financed by the State. They enjoy comprehensive autonomy.

The Basic School (Folkeskolen)

Formative Assessment for Learning

The basic school is decentralised in many ways. The Parliament still makes the decisions governing the overall aims of basic education and determines a nationally agreed framework for curriculum content, minimum numbers of lessons, teacher qualifications, and other such general, basic minimum "input" standards. The Ministry of Education sets the targets for each subject at a general level, but local authorities and schools decide how to reach these targets.

Parents have a great deal of influence, not only as members of the board of governors. They are expected to take an active interest in what happens in the classroom. The fact that the state supports private schools by about 85 percent of their budget makes them a real alternative to the public schools and consequently makes public schools more willing to listen to parents.

Testing does not play a significant role in Danish basic schools. However, the Education Act states clearly that pupils and parents must be regularly apprised of the school's opinion of how each pupil is profiting from the schooling. This is done verbally in the early years. Beginning in the 8th grade this information system is augmented by a written report at least twice a year, giving the pupil's position in academic achievement and effort.

In 1993, the Parliament adopted a new act on the Folkeskole. The act is based upon the principle of steering by targets and frames maintaining and expanding the decentralised governing of the Folkeskole. An essential part of the basis of the new act was created by a four-year innovation programme for the Folkeskole.

The crucial innovation is found in the organisation of the teaching content and in the improvement of the evaluation methods used. The earlier division of the subjects of English, German, mathematics, and physics/chemistry in basic and advanced courses in the 8th–10th grades has been abolished. Instead, the new act requests the schools and the teachers to adapt the teaching to the qualifications of the individual pupils. This has to be done on the basis of a running internal evaluation and the fixing of goals for the individual pupil and groups of pupils.

Summative Assessment

As mentioned, testing does not play a significant role in the basic schools. There is no overall examination; graduation examinations may be taken on a single-subject basis. National tests at age 16 in basic subjects are not obligatory but are taken by practically all pupils. It is the pupils themselves who decide whether they want to present themselves for an examination in a particular subject. There is no pass mark.

In order to vary the evaluation of the pupils' benefit of their schooling a mandatory project assignment will be fixed by a ministerial order, but the topic of the project will be chosen totally by the schools and the teachers.

The Folkeskole teachers decide whether or not the young person is qualified for upper secondary education.

Youth Education

Formative Assessment for Learning

The executive order for the upper secondary schools says that the instruction must be regularly evaluated for internal purposes so that students and teachers are informed of the benefit derived from the instruction. The aim of the continuous evaluation is to guide the student and the teacher with a view to the further planning of the instruction. The individual teacher will thus have a possibility of adjusting the progress and the level. The evaluation also provides a basis for a detailed guidance of the individual student with regard to progress in the subject and in studying methods.

The methods of evaluation are determined by the students and teacher together and are influenced by the form and content of the instruction. The whole course of instruction shall be evaluated by means of tests, special assignments and reports, or conversations.

At the schools there is an increasing interest in strengthening the quality of the teaching through internal evaluations.

Summative Assessment

Marks are given at the upper secondary schools for written and oral work:

- Term marks and testimonials are given by the teachers of the subjects.
- Marks for the year's work are given by the teachers and appear on the student's examination certificate.
- Examination marks for written examination papers are given by two external examiners appointed by the Ministry of Education.
- Examination marks for oral exams are given by the subject teacher and one external examiner appointed by the Ministry of Education.

The written examinations are produced by the ministry, by committees for each subject. The examinations are of the essay type, allowing students to explain a problem, argue a case, etc. Committee members are distinguished teachers. Each year the ministry decides on the subjects in which the students in the *gymnasium* shall be orally examined. The higher preparatory examination is always taken in all subjects. The student's results at the final examination at the end of the upper secondary education is of utmost importance in the competition for university places, etc.

The teachers are responsible for their teaching to the head teacher. At the end of each year they prepare a report of the year's work, which the head teacher submits to the national subject adviser. The head teacher is responsible for the curriculum, the teaching, and the examinations to the Ministry of Education.

Examination results are reported by the schools to the Ministry of Education and are controlled by the national subject advisers,

who use the results from all the schools for improvement and guidance at the national level. This is important, because the examinations are closely aligned to the curriculum and the teaching. National results are published as a total for the system and broken down by subjects, but they are not published for individual schools.

Systemic Reform

Danish education, like the educational systems in many other OECD countries, has been in a state of reform for the last ten years. The reforms have been carried out in accordance with—by now—well-known principles. The main ones are:

- Steering by targets and frameworks instead of detailed regulation;
- Decentralisation to schools and other educational institutions;
- Increased market orientation with user-influence through boards at the institutional level;
- Free choice of schools;
- Focus on quality, internationalisation, and new information technology;
- Establishment of better bridges between the various levels of the educational system.

In recent years much emphasis has been placed on adult education, both at the secondary level and at the level of tertiary education.

In 1993, the Minister of Education introduced a plan for coming to grips with the drop-out problems of youth education. The main guidelines for the plan were:

- The individual pupil should be brought into focus.
- All young people should be challenged.
- All types of youth education should develop the personality and creativity of the pupils.
- The youth education system should make it possible for the individual pupil to tailor his/her own sequence of courses.
- Education should be stimulated by experiments and development projects at the schools.

In 1995, the Minister of Education introduced an action programme, which is bringing the teacher and the future role of the teacher into focus. The program is based on a wish to involve the teachers, pupils, parents, organisations, the trades and industries, and the mass media, etc., in a debate on the role of the teacher and in the development of that role. One of the main targets is to reinforce development projects at the schools and in-service training programs for the teachers. Another significant target is to strengthen the status of the teachers in the society.

This project is important, because the quality of education and the quality of the individual institutions is extremely dependent upon the work of the teachers. It is essential for the educational system that the teachers have the sufficient qualifications and are able to develop an appropriate practice in the classroom in accordance with variable conditions, such as the qualifications of the pupils.

ENGLAND

DIANE SIMMONDS, Office for Standards in Education, London

Current assessment practice is born out of very different approaches in the primary (ages 5–11) and secondary (ages 11–16) phases of compulsory education. Traditionally, primary teachers were engaged in much informal assessment of children's progress and attainment, and records have been varied in style and content. There was very little formal testing in this phase, other than of the basic skills of number, spelling, handwriting, and grammar. In the later stages of the phase, some use was made of standardised tests of these basic skills. Where there was selection in terms of ability for entry to schools at age 11, there were specific tests, for example, of verbal and non-verbal skills.

In contrast, the emphasis in the secondary phase was on regular testing, some use of standardised tests in the early stages of the phase, and annual examinations set by the school. The major public examination experience at age 16, strongly influenced in style by the examination system in the post-compulsory phase and university entrance requirements, catered mainly to the more able pupils.

The national curriculum (NC) and its assessment arrangements—introduced progressively from 1989—maintained and extended the range of assessment approaches in the secondary phase, and have a growing influence on the early secondary years. In the primary phase, the NC is leading to more formality in assessment methods, not only in the statutory tasks and tests in some subjects at ages 7 and 11 but also in attainment records and, in the later primary years, in an increase in formal testing.

The National Curriculum and Assessment

National school testing and assessment was introduced in England as part of the national curriculum which was brought about as a result of the 1988 Education Reform Act. Under the 1988 Act pupils are assessed at ages 7 (Key Stage 1), 11 (Key

Stage 2), and 14 (Key Stage 3), in English, mathematics, and science. There is an intention in the future to introduce statutory assessment in other subjects in the curriculum at age 14, but this has not yet happened.

Pupils are assessed against the standards set out in the national curriculum. For each subject the national curriculum requirements are set out in a document (the subject Order) which lays down the programme of study which pupils must be taught during the key stage, and indicates the standards which they are expected to reach in each aspect (or Attainment Target) of the subject. The standards are defined in terms of the performance which would be typical of pupils working at each of the eight levels in the national curriculum. Pupils are assessed on their performance in two ways: first, by their own teachers, using the level description to come to an overall judgement about their attainment at the end of the key stage; and second, by means of nationally set tests or tasks. The results from the tests/tasks and teacher assessment are reported alongside one another to parents, and in local and national summaries of results.

The body responsible for developing the tests and tasks and for administering the national assessment arrangements in England is the School Curriculum and Assessment Authority (SCAA). In Wales, where similar arrangements apply, the responsible body is the Curriculum and Assessment Authority for Wales (ACAC); at present SCAA oversees the development of the testing and assessment arrangements on behalf of both bodies.

At ages 16 and 18 there are the external school qualifications; the General Certificate of Secondary Education (GCSE) mainly for 16-year-olds, and the General Certificate of Education (GCE) Advanced Level for 18-year-olds. The latter have recently been joined by a new family of General National Vocational Qualifications (GNVQ), which are mainly taken by 18-year-olds.

Purposes of Assessment

Developments in assessment and the introduction of statutory requirements have led to a perceived need to articulate the rationale for assessment. An overriding purpose of assessment in the relatively recent past, particularly in the secondary phase, but to some extent also in the primary phase in the context of

selective education, has been selection for the next stage in education or for employment. The move to widespread comprehensive education and changing patterns of employment, coupled with higher staying-on rates in education after age 16, have fuelled a reappraisal of the purposes of assessment. Over the recent past, assessment practice has moved in two directions.

- from an almost total emphasis in the secondary phase on formal tests, and public examinations for leavers, to a broader range of assessment approaches both during, and at the end of, courses;
- from a very informal approach to assessment and recording in the primary phase to one which involves more formality and, statutorily, includes externally set tasks and tests and thus preparation for them.

These moves illustrate two major purposes of assessment, both of which are now embedded in the widely accepted rationale for assessment: to enhance standards and increase accountability. To many teachers and other educationalists, the first of these is the educationally defensible purpose, recognising that assessment is at the heart of teaching and learning. The second, though, is seen by some as hard-edged and intrusive, creating too high a degree of competition between schools, and moving the focus from pupils' needs to the interest of others. In reality, both aspects are central to the purposes of assessment.

The overall aim of the national curriculum and its associated assessment arrangements is to raise standards of learning in schools by ensuring that all pupils have an entitlement to a broad programme of study and by setting realistic but challenging targets for achievement. Pupil assessment fulfills an essential function in providing the information needed to measure improvement, and in motivating schools and pupils to aim for higher standards.

Within the overall aim of raising standards, pupil assessment has two main purposes:

- the formative purpose, to inform teaching and learning, and to help pupils to improve on their performance;

- the summative purpose, to provide information about pupils' attainment at a fixed point in time, to report this to their parents and others, to provide information about national and local performance, and to evaluate the effectiveness of the education being provided.

The first purpose is fulfilled essentially through everyday teacher assessment, using the framework provided by the National curriculum. It is not formalised, but is regarded as a normal part of teaching and learning.

The second purpose is fulfilled through the end of key stage arrangements, which comprise statutory teacher assessment (in which a level has to be given to each pupil in each subject, using the level descriptions set out in the subject Orders), and statutory tests or tasks that are set centrally and administered under controlled conditions in every school. Statutory teacher assessment is based on the teacher's knowledge of a pupil's performance over a period of time, and the tests and tasks provide a standard, reliable measure of attainment at a fixed moment in time.

Debates in recent years have centred on the respective strengths and weaknesses of these two forms of assessment, with some commentators saying that teacher assessment ought to be the predominant means of recognising pupils' progress because it is based on fuller information, and others saying that the tests provide a more reliable picture of attainment because of their standard nature. The debate was resolved in 1993, when it was agreed that the two forms of assessment were both essential to provide a true picture, and should be regarded as complementary to one another in all forms of public reporting. Most educationalists are prepared to accept this view, but efforts need to be made to explain it to parents and the wider public.

Using Assessment Outcomes

At the national level, the results from both tests and teacher assessment at all three key stages are collected from all schools, and a summary is published by the government, giving the percentages of pupils who were awarded each level in each subject at each key stage. Comparisons of these results with those of previous year show the extent to which overall levels of

achievement are changing from year to year. The national summaries are also published by individual schools alongside their own summaries of results, so that parents and other interested parties can view the school's performance in the light of the national picture. For individual parents, their own child's results are given via an annual report alongside both these sets of figures, so that they can look at their own child's performance in the light of both the local and the national picture.

The publication of results is an important element in the government's desire to make information concerning the performance of the education system available to the public, as part of the Citizen's Charter. It is an essential factor in ensuring that the system as a whole, as well as individual schools, is accountable to those it serves. The government has recently announced its intention to publish tables of results of 11-year-olds in 1996, giving the performance of primary schools in comparison with one another on the national assessments, to parallel those already published for secondary schools in respect of the public examinations at GCSE and A level. There has been opposition from head teachers to the introduction of school performance tables, on the grounds that unadjusted results do not present a fair comparison between schools of different characters and with children from different socio-economic groups. For this reason SCAA has commissioned a national project to develop value-added indicators, which will allow schools to measure their success in terms of the progress made by pupils between two assessments, rather than in terms of the results taken in isolation. It is hoped that a national system of value-added performance indicators of this nature can be introduced, using the national test outcomes, within the next five years.

A further use of the test outcomes is to analyse performance at both national and local levels in more detail, in order to identify specific aspects where performance is weak. For example, an analysis of the mathematics tests for 11-year-olds in 1995 revealed weaknesses in aspects of numerical computation on a national level, which need addressing in schools. SCAA now publishes annual reports on pupils' performance on the tests, which give an account of the national strengths and weaknesses, to help policy makers, schools, and teachers target areas for improvement.

FINLAND

MARTTI APAJALAHTI, The National Board of Education, Helsinki

Formative Assessment for Learning

Comprehensive School

In Finland, the nine-year comprehensive school is intended for the whole age group of 7–15. Schools and teachers are responsible for pupil evaluation and assessment. The national framework curriculum includes general guidelines concerning assessment; mainly regulations on grading and on the forms of certificates and school year reports.

According to the national guidelines, the purpose of assessment is to motivate, encourage, and guide the pupils in their studies. A written school report, using a numerical scale of 4–10, must be given at the end of each school year and at least once during the school year. Descriptive verbal assessments (written) may be used instead of the numerical report for grades 1–4.

The more detailed principles of assessment are determined in the curriculum of the school. Many schools are active in developing pupil evaluation and assessment as part of their new curricula. Self-assessment and portfolios are used in some schools.

There is no national testing system. Many textbooks include tests on the topics of the book. It is up to the teachers whether to use these tests.

Senior Secondary School

The senior secondary school has a three-year syllabus that can be completed in two to four years. Studies consist of courses, and most schools are non-graded. How each course is graded, and how each subject is graded on the basis of the course grading, is determined in the curriculum of the school. The curriculum of a non-graded school also defines the grounds that will cause the interruption of study in a subject. This kind of

situation occurs if a student repeatedly gets failing course grades in a subject.

Summative Assessment

Comprehensive School

After completing the comprehensive school curriculum, pupils are given a graduation diploma. There are no national tests or examinations, so the grades given by different schools and teachers are not comparable. The comprehensive school graduation diploma is, however, used as a selection criterion to further studies. This controversy is the main reason why pupil assessment will probably be reformed in the near future.

Senior Secondary School

The national framework curriculum for senior secondary school includes some general guidelines on pupil assessment (graduation certificate).

A student who has completed the syllabus of the senior secondary school is given a graduation certificate that includes the subjects studied, the number of courses taken, and the grade for each subject (on a scale of 4–10).

A national examination, the matriculation examination, is organised in senior secondary schools. The matriculation examination is set and pupils' answers are marked by an independent board appointed by the Ministry of Education. A separate certificate is given for passing the examination.

The completion of the senior secondary school and matriculation examination gives general eligibility for university studies.

Systemic Reform

Systemic Evaluation

The National Board of Education is responsible for the evaluation of the whole educational system. A large evaluation project was carried out on senior secondary schools in 1994 and on comprehensive schools in 1995.

A new model for continuous evaluation of the educational system is being designed by the National Board of Education.

Policies on Curriculum and Assessment

New national framework curricula were confirmed for comprehensive and senior secondary schools in 1994. The central aim of the curriculum reform was to dismantle centralised regulations and to devolve decision power on the local level, the municipalities and the schools. The actual curricula are designed locally according to the national guidelines.

As mentioned above, the framework curricula also include national guidelines for pupil assessment. The new framework curricula stress the importance of shifting the focus on learning. A strong emphasis is laid on motivation and on the active role of the student as the organiser of his or her own structure of knowledge.

Pupil assessment is an integral part of curriculum development.

Comprehensive School

In the beginning of 1995, the Ministry of Education set up a working group to make proposals how pupil assessment should be reformed in the comprehensive school. The working group proposed that, because of the different role and function of continuous evaluation (everyday evaluation and regular school reports during the school years) and the graduation diploma, they should be viewed separately.

According to the working group, continuous pupil assessment should be based on individual objectives and study programmes. It should be reformed in a way to promote flexible (non-graded) teaching and learning arrangements. The main purpose of this kind of assessment is to motivate and guide pupils in their studies and to help them to build up a positive self-concept. The working group wants to diminish comparison, ranking, and competition among pupils during the school years.

No strict national guidelines or regulations are needed on continuous pupil assessment. The principles of assessment should be determined in the curriculum of the school.

On the other hand, a more standardised graduation diploma is needed. The working group proposed that there should be

national criteria on grading in different subjects for the graduation diploma. These criteria should be with optional tests that schools can obtain from the National Board of Education. The final grading is the duty of the teachers, but they can use tests as means of adjusting their grading to the national scale. Thus, the grades given in the graduation diploma would be comparable.

How can continuous evaluation and assessment based on individual objectives be combined with the standardised graduation diploma? When pupils plan their study programmes and set their learning objectives (with their parents and teachers) for the whole period of comprehensive school, it is natural that when graduation gets near they also set individual target grades in different subjects for the graduation diploma.

In Finland there is a discussion going on concerning the proposals of the working group. Decisions will probably be made this year. Most criticism has focused on standardising the grades of the graduation diploma.

Senior Secondary School

The matriculation examination has been much criticised during the last few years, but at the moment there is no strong criticism. Due to the increased options in the new senior secondary school curriculum, it has been necessary to develop the matriculation examination in such a way that it better relates to the studies in the senior secondary school. The individual examinations of the matriculation examination are developed so that they will better measure the skills and abilities of processing information and knowledge.

According to new legislation concerning the matriculation examination the individual examinations in different subjects may be taken separately in three different, successive examinations. This means that when a student has completed all the compulsory courses in a subject he or she can take the examination in that subject. The matriculation examination is organised twice a year.

FRANCE

CLAUDINE PERETTI, Ministry of Education, Paris

France has had for a long time a full system of pupil assessment, both in the form of teacher assessment conducted informally in the classroom, and in the form of external examinations. The best known of these is the *baccalauréat*, the graduation certificate for 18-year-olds that also gives access to higher education. A lesser known examination for 15 year olds is the *brevet des collèges*, for which pupils sit on a voluntary basis and which is made up of both continuous assessment and proper examination papers. This examination is seen more as benchmarking and does not affect pupils' progress to a higher class.

It was, nevertheless, felt that these systems of pupil assessment could usefully be complemented by other types of evaluation procedures.

As a result of France's choice in favour of mass education to meet the expectations of the country and guarantee its economic competitiveness, the French Education Minister decided in 1987 to create within his Department an office responsible for assessment and forecasting, *la direction de l'évaluation et de la prospective (DEP)*. Its brief is twofold: to give the nation an account of the effectiveness of the education system, and to provide administrators and agents of the system with the tools and the indicators necessary to monitor and manage it.

Measuring the Effectiveness of the Education System

To measure the effectiveness of the education system, four approaches of increasing complexity have been selected:

- Assessment of pupils' competencies and knowledge, including a comparative dimension both historical and international; the national curricula make easier the organisation of assessment and the French experience in this field dates back 20 years.

- Evaluation of education policies used to promote a more favourable global learning environment and to help specific groups of pupils to overcome particular difficulties; the aim of these assessments is to measure not only the agents' satisfaction but also the effects of policies on pupils' achievements.
- Evaluation of the agents in the system, namely those responsible for implementing education and training policies; the main approach is researching into efficient teaching practices.
- Evaluation of teaching units, more specifically trying to understand the impact of organisational constraints in schools and of the interaction of the people in them on pupils' learning and achievements. The added value of schools is a major subject of study in this area.

System Tools and Indicators

To provide administrators and agents of the system with tools and indicators for monitoring and managing it, the DEP is working at the following levels:

Classroom Level

In 1989, the DEP created national assessments at the beginning of *CE2* (age 8) and of *6ème* (age 11). The purpose of these assessments is to provide teachers with a measure of pupils' difficulties and deficiencies and, as a result, to help them to find remedies. In the same way, the DEP produces and disseminates assessment tools—diagnostic, formative, or certifying—for use by teachers at will. At the moment, these tools exist at the primary school level for every branch of learning except arts and physical education; at the lower secondary school level (*6è* and *5è*), for mathematics and French language; at the upper secondary school level (*2nde*) for mathematics, French, foreign languages (English and German), history, and geography in the general and technical branch, and for mathematics, French, economy, and management or industrial subjects in the vocational branch.

School Level

In 1995, the DEP disseminated an indicators system to all secondary schools. The objective is to provide schools with tools for working out a school project and managing it. This system proposes some 20 standard indicators common to all secondary schools, enabling them to calculate their own specific values for these indicators, and thereby assess their own performance rating; furthermore, this helps them to take stock of the way they are run, of their available resources, and of the various constraints which they have to take into account.

Policymakers' Level

The widespread diffusion of work conducted by the DEP helps policymakers to regulate the system. The link between DEP work and policymakers' decisions is close, even if other factors come into play. In particular, some synthetic publications play a regulatory part. These are: *l'Etat de l'Ecole*, which uses thirty indicators to analyse the cost, the activities, and the results of the system; and *Géographie de l'Ecole* which brings to light the disparities between the *académies* (regions) and their evolution with thirty-six indicators. These indicators are calculated and published every year.

Spreading an Assessment Culture

The main obstacle to the spread of an assessment culture throughout the education system is the lack of training. Teachers and school managers are not familiar with this new approach to assessment. The inspectors who are working at the local level with teachers are not always themselves convinced of the advantages of using new concepts and new tools for assessment. As a result, a major effort is currently being made to develop training. For example, national and regional training sessions are set up at the same time as the indicators system is extended to all secondary schools.

Nevertheless, it is reasonable to think that the new assessment culture which is promoted is exerting an increasing influence on teachers' and school administrators' practices and that in this way changes are implemented in the system, thereby improving its results.

GERMANY

PETER MUNK, Ministry of Education, Bonn

Over the past twenty years, the German *Lander* have elaborated in a step-by-step process of revision new curricula, which are generally characterised by their educational objectives. But even today, the core of school curricula continues to be a plan that establishes the basic framework for the material to be taught under each subject for each age group. Regulations governing the approval of textbooks complete the administrative control of schools via curricula. A second major set of administrative regulations relating to classroom tests, the awarding of marks, the transfer of pupils to other classes and to other schools as well as to examinations provides a standardised basis for decisions concerning a pupil's school career. The regulations ensure the formal continuity of the assessment of pupils' performance as well as standardised assessment procedures—in most cases leaving some flexibility for teachers in making their assessments of pupils' performance. In addition, a third set of regulations establishes the organisation of instruction, for example by determining class sizes, pupil-teacher ratios or permissible—or necessary—forms of differentiation of instruction.

In principle, all types of schools in Germany are based on age groups. In other words, as a rule, pupils move up each year into a higher grade. Pupils can be moved up, however, only if they have achieved sufficient scholastic attainments, which are continuously assessed. Twice each school year, pupils are handed reports on their school work, in which their achievements in the individual subject areas are assessed with marks ranging from one to six. This marking system applies for all age ranges and types of schools providing basic schooling. If a pupil's performance is poor or unsatisfactory, he or she is required to repeat an entire school year. In general, the sets of data compiled by the German *Lander* display a continuous drop in the number of pupils who have had to repeat a school year. Over the past years, important new regulations have been introduced concerning the normal promotion of children at

primary schools (in particular, from Grade 1 to Grade 2) and the introduction of verbal assessments to replace assessment by marks in the first two grades. This procedure is intended, above all, to facilitate a clear-cut recommendation as to what, and how, a child should best continue to learn and study. However, studies analysing the practical implementation and focusing on the effects of verbal assessments reveal that teachers frequently continue to award marks in their verbal assessments and find it difficult to provide useful hints on the nature and content of further learning by giving an accurate, detailed picture of the level of learning achieved by any one child.

Assessments and School Transfers

In the individual German *Lander*, regulations governing the transfer of pupils to secondary education (second-level general school, intermediate school and grammar school; comprehensive school) differ. Transfers are usually effected on completion of the fourth school year. Occasionally, transfers are effected on completion of the sixth year. Transfer to the intermediate school and grammar school often places great strain on pupils attending the final class at primary schools. Empirical studies on the reliability and accuracy of the decisions on the streams of pupils selected for transfer reveal that this issue has not yet been satisfactorily solved. Average marks and the results of test papers reflect merely a fraction of the skills required to achieve success at secondary schools.

In order to ensure the best possible transfer procedure from the educational point of view from primary to secondary education, a diagnostic stage was introduced as far back as in the 1970s, which today still represents the basis for assessing the aptitude of pupils. The diagnostic stage is intended to:

- enable the individual pupil to recognise his or her specific skills and interests, as well as to familiarise him or her with the learning requirements of the school type chosen;
- promote the pupil's willingness to learn as well as his or her learning ability;
- optimise the decision to be taken with regard to the pupil's continued schooling (type of school) by means of observation and consultation;

- eliminate, if possible, the influence of other than individual attainment differences on the choice of the pupil's continued education, in particular the influence exercised by sex and social background.

Over the past few years, the educational programmes provided for Secondary Level I have expanded their curricula and, at the same time, they have become more differentiated than in the past. This means that final examinations and admission prerequisites can no longer be exclusively linked to a specific type of school. The development of educational programmes at Secondary Level II, which, to some extent, also provide certificates making pupils eligible for studies at *Fachhochschulen* in addition to vocational qualifications, have also helped to open up the system. In addition, manifold activities are designed to help bring about the equivalence of general and vocational education.

Promotion of Pupils in Accordance with Attainment

In addition to promoting all children and to promoting disadvantaged pupils, the school oriented to equality of opportunity endeavours to provide appropriate support for those pupils who are considered particularly talented. Special curricula for such pupils do not exist. There are opportunities for enrolling children with a high degree of intelligence in school before they reach school age. They can jump classes—this is, however, seldom—and a number of grammar schools are conducting experiments designed to enable groups of particularly high achievers to sit for their *Abitur* examination one year sooner than the rest.

Independent of the existing organisation of the *Mittelstufe* (school years 5–10) in different types of schools, the *gymnasiale Oberstufe* (last three years of secondary education) no longer offers the traditional varieties such as scientific or foreign language courses, but a flexible system of compulsory and optional subjects, instruction in which is offered at two levels. The intention is to continue to provide joint basic education on the one hand and to facilitate special forms of pupil promotion on the other hand. The choice of subjects and of level enables students to select widely differing subject combinations and focuses, which pupils determine to a large extent based on their

own assessment of their abilities and aptitudes. By exercising self-assessment, pupils exert a direct influence on their learning process. The traditional assessment of the pupil's attainment is supplemented by a system based on the awarding of points with a complicated procedure for re-conversion to marks. The points gained during the second stage of secondary education and the *Abitur* examination mark are combined in calculating the pupil's average *Abitur* mark.

The major forms of differentiation adopted by comprehensive schools are cross-sectional differentiation of levels for cooperative comprehensive schools and individual differentiation through course levels for integrated comprehensive schools respectively. Pupils are assigned to courses depending on the assessment of their performance. It is thus also possible for pupils to switch courses every six months.

Trends in School Policy

Current endeavours to achieve systemic reform in schools focus not merely on new learning contents (curriculum) and new forms of teaching and learning, but also increasingly on new ways for schools to exercise self-responsibility and self-organisation. The enhancement of independence and responsibility for the further development of the school system calls for new forms of quality development and quality assurance in schools. Increased independence and accountability on the basis of internal and external assessment of the curricular and extracurricular activities in schools go hand in hand. This process must be accompanied by appropriate qualification of, and support for teachers.

In this connection, new forms of the evaluation of procedure at schools serve to ensure the maintenance of quality as well as quality development. Increased responsibility entails the need for local as well as regional cooperation on the part of educational institutions among themselves, as well as with other institutions. During the next few years, pilot schemes will have to develop and test various forms of such cooperation.

Evaluation and quality assurance concepts have gained considerably in importance within the context of current processes of development. Schools are increasingly required to examine the effectiveness of resources employment and to regard quality assurance as a permanent task. The autonomy of

schools means that schools must engage in continuous scrutiny of their own work. In this connection, consideration of the assessment of their school by the pupils themselves furnishes important information about the "client's" point of view, which represents a useful supplement to the quality standards applied by teachers. Questionnaires are already being distributed among pupils at many schools. It is, of course, important that endeavours do not stop at surveys, but that appropriate conclusions are drawn and that pupils participate in discussions as to how the current situation can be improved.

In Germany, school researchers regard evaluation as a field for investigation in the years ahead. To date, little practical experience is available for inclusion in debates on school development. For this reason, currently available evaluation concepts must be refined in the course of systematic trials and examined as to their efficacy and efficiency.

GREECE

JOSEPH SOLOMON, Pedagogic Institute, Athens

Despite a long-standing expression of interest in the decentralization of education, both from teachers' professional organizations and from major political parties, only recently has there been a move towards the partial decentralization of administration and funds allocation, due to the recent laws for decentralisation of some government responsibilities and activities. In practice, however, education in Greece is still highly centralised and schools, both primary (Grades 1–6) and secondary (lower secondary, Grades 7–9, and upper secondary, Grades 10–12) are so far characterized by strong uniformity in teaching contents and methods, following detailed curricula and guidelines emanating from the Ministry of Education with a mandatory character for all schools.

Within this context, education evaluation and monitoring, given the centralized prescription of educational practice, has until recently been undervalued. Even when information and data on several aspects of the education system were collected, they were, for reasons of low priority and financial deficiencies, either never published or published too late. Overall, collecting information on education practice and achievement (performance of institutions, quality of the teaching force, effectiveness of education policies, students' achievement), let alone using it for reasons of education policy and decisionmaking, has not yet become a priority with the education authorities.

Nevertheless, in the past few years there has been a growing concern among policy makers about assessment and the need for a system of educational evaluation. In fact, the first steps towards the institutionalisation of such a system have been made, such as the constitution of work groups on evaluation by the Ministry, the appearance of a prospect for a system of evaluation in legal texts, or the recent establishment of a Department of Evaluation at the Pedagogic Institute (the Ministry's advisory organisation).

Assessment of Students

Students' assessment in Greece is, so far, a practice that generally complies with guidelines emanating from the Ministry of Education, as does all educational practice in both primary and secondary education. The assessment guidelines form part of the teaching guidelines included in the existing detailed curricula for each subject taught, in the Teachers' Guidelines publications as well as in students' textbooks .

Traditionally, student assessment methods still widely used by teachers include individual oral testing on a daily basis, short written ad hoc tests, periodic hour-long tests (e.g., every trimester) and final tests on the contents covered. These assessment procedures can be said to have been of a predominantly summative nature with only an indirect formative character. Their outcomes were announced to both students and parents; however, there were no specific feedback and formative procedures in action, apart from teacher-parent meetings.

The 1994 Presidential Decrees (PDs) on Students' Assessment aimed to (re-)define the approach and methods of assessment. The approach adopted by these PDs was at the same time summative and formative with a greater emphasis on the latter. The assessment methods or procedures suggested in the PDs are:

- individual oral testing;
- short tests during classroom periods;
- one or more one-hour long tests following specific "re-capitulation" procedures each trimester.

The Presidential Decrees also proposed the introduction of some more qualitative assessment methods and formative pedagogic procedures. These included:

- descriptive evaluation sheets depicting the progress and effort of each student;
- assessment of individual or group creative work projects;
- pedagogic diaries recording students' achievement and progress through a year or stage of education;
- teachers' council meetings, and teacher-parent counselling meetings.

For the lower primary school grades (1 and 2) assessment is effected only through descriptive evaluation sheets. For the middle primary school grades (3 and 4) characterisations are awarded not by straightforward marks but rather through verbal characterisations, i.e., "excellent," "very well," "well," "almost well." For the two upper primary school grades (5 and 6) grading is done according to a 10-degree numerical scale (1–10).

At lower as well as upper secondary schools assessment is effected, as in the past, on a 20-degree numerical scale, in addition to descriptive evaluation procedures for the lower secondary schools.

The student assessment procedures throughout primary and secondary education, are run by the individual teachers, and are classroom based.

Summative assessment is practiced only within each single grade. Assessment outcomes in different grades are not summed up at any level. School graduation certificates are based on the mean grade obtained through the final school year of each level only.

Primary school students, with few exceptions, are promoted automatically throughout primary education (Grades 1–6) and on to lower secondary education (Grades 7–9). Promotion in lower secondary schools is controlled through each school year's assessment results. Failing students are obliged to repeat the year.

The PDs and the specific ways of implementing their measures are undergoing revision and fine-tuning after a pilot period in 1994–1995, based on the feedback from teachers and school advisers. The PDs produced strong reactions from teachers associations, concerning mainly the material conditions for the implementation of the new assessment methods (time and salary allowances for the extra work demanded, etc.), although some reactions referred also to the actual need for or nature of an innovation, the complexity of some measures, etc.

Evaluation of Education

As mentioned above for the time being assessment is not used towards the evaluation of educational institutions or of the education system as a whole. So far, this type of evaluation is not practiced in a systematic manner by the education authorities

or any other official evaluation or inspection organisation. In general, monitoring of educational processes and outcomes is not effected at a national level, except indirectly through higher education entrance examination results.

The only case in which external assessment was used towards this end was in the context of the studies conducted by the IEA Greek National Centre. However, these outcomes have not yet been used for drawing conclusions and designing educational policy.

Currently, evaluation policy is considered and elaborated by the Pedagogic Institute. The aim is the constitution of a National Assessment and Evaluation System that will promote and regulate local formative and summative assessment procedures, as well as external standardised testing used mainly for formative and monitoring purposes. To that effect, the Pedagogic Institute and the Ministry of Education are considering the constitution of a National Testing Service and a National Item Bank. Also, a system of self-evaluation of educational institutions is being elaborated and a system of education quality indicators at school level is being developed.

In addition, the Pedagogic Institute is in the process of constituting a series of new services such as a bureau for quality control and certification of educational material, a standards bureau, and external committees of curriculum and textbook evaluation and development.

These projects are linked to the major reform for the primary and secondary levels of education currently being designed by the Ministry of Education and the Pedagogic Institute. The main tool of this reform will be the unified curriculum. A new approach to curriculum design is planned, aiming at the production of a relatively concise national curriculum for all compulsory education, containing the axes and aims for education in general and for the various content units. The unified curriculum will aim to regulate approximately 80 percent of the overall teaching time, and 20 percent will be left for local determination, at either the school or local education community level. The unified curriculum will be published yearly with only slight modifications when necessary, and may be reformed every five years.

At the same time the Ministry of Education promotes the transformation of the upper secondary school (*Lykeion*), and the introduction of external examinations for the upper secondary

school graduation certificate, thus allowing for the transformation of higher education entrance procedures.

IRELAND

CARL O DALAIGH, Department of Education, Dublin

The White Paper, "Charting our Education Future," was published in April 1995. This White Paper enunciates national policy on curriculum and assessment in primary and secondary level education.

Curriculum in Primary Education

The primary school curriculum is based on the following principles:

- the full and harmonious development of the child, with due allowances made for individual differences;
- the central importance of activity and guided-discovery learning and teaching methods;
- teaching and learning through an integrated curriculum and through activities related to the child's environment.

These principles were embodied in the 1971 child-centred curriculum, which is being reviewed at present while the basic principles adopted in 1971 are being retained. The review is aimed at a more precise statement of objectives in terms of student behaviour and attainments.

Apart from the core areas—mathematics, Irish, English, social and environmental studies—special emphasis will be given to new programmes in the arts, in science—seen as part of the social and environmental programme—in a European awareness programme including European languages, life, and culture, in Irish language, in social, personal, and health education and in religious education. The revised curriculum will reiterate the right of schools, in accordance with their religious ethos, to provide denominational religious education and instruction to their students.

There will also be emphasis on students with special needs—all students will have a right of access to, and participation in, the education system according to their potential and ability. This will entail positive intervention at all levels for those minorities who experience particular difficulties. Another policy objective is that all traveller children of primary school age be enrolled and participate fully in primary education, according to their individual abilities and potential, within five years.

Curriculum in Second Level Education

Junior Cycle

The junior cycle curriculum is based on the principles of breadth and balance, relevance, quality, continuity and progression, and coherence.

Schools will be expected to provide students with experience in the areas of language and literature; mathematical studies; science and technology; civic, social, and political education; arts education; religious education, guidance counselling, and pastoral care; physical education; and health education including personal and social development, relationships, and sexuality education. A curricular framework will apply to all students and schools and there will be a range of full courses (3 years) or short courses (1 year or dispersed over 3 years). In particular, a new civic, social, and political education (CSPE) is being introduced as a compulsory part of the curriculum from September 1996. Relationships and sexuality education is also being introduced from September 1996 as part of a broader social, personal, and health education (SPHE) programme.

Senior Cycle

Retention rates to the end of senior cycle are 77 percent at present. The target is that the percentage of 16- to 18-year-old age group completing the senior cycle will increase to at least 90 percent by the year 2000. This will be achieved through a major restructuring of the senior cycle which is underway in the context of a full three-year programme in addition to the on-going and gradual developments of the curriculum at junior cycle that will encourage further participation at senior cycle level. The Senior cycle changes are:

- the Transition Year programme is now available as an option for all second level schools. In 1995/96, 579 schools (75 percent) and 30,000 pupils (45 percent) participated;
- the revision of the established graduation certificate programme;
- the development and expansion of the graduation certificate vocational programme—a number of vocational modules, Preparation for Work, Work Experience and Enterprise Education, must be taken as well as six graduation certificate subjects in this programme;
- the introduction of a new graduation certificate applied course which is based on three main strands—general education, vocational education, and vocational preparation—and is modular rather than subject based.

This new structure with three separate orientations provides a coherent framework to broaden the scope of educational provision in the senior cycle and to facilitate schools in providing appropriate and relevant programmes for all students. The aim is to provide an effective general education with an expanded, enhanced vocational orientation and diversity of provision to meet differing aptitudes, needs, abilities, and career options.

In addition to updating the content and relevance of the syllabuses for the thirty subjects in the graduation certificate curriculum, the National Council for Curriculum and Assessment has been requested to increase, where relevant, the vocational orientation of each subject.

Assessment in Primary Education

Assessment is of central importance in monitoring and enhancing the quality of education at school and national levels. Assessment practice ranges from observation, classroom discussions, and homework, to the use of standardised tests, both norm- and criterion- referenced. Standardised assessment is currently being used to evaluate reading and mathematics in

many schools on an annual basis from first class to sixth class (6–12 age group).

In accordance with the proposals in the White Paper, all primary schools will be required to develop a policy on assessment within the framework of the school plan to ensure uniformity and continuity of approach between classes and within the school. Students will be assessed at the end of first and fifth class in order to evaluate the quality of their learning and to identify any special learning needs that may arise.

The objective is to ensure that, having regard to the assessment of their intrinsic abilities, there are no students with serious literacy and numeracy problems in early primary education by the year 2000. The Educational Research Centre and the advisory National Council for Curriculum and Assessment will develop appropriate standardised forms of assessment for these core competencies to be applied to all levels in the primary school.

Assessment should be diagnostic, formative, and continuous, and geared towards providing information for teachers, schools, students, and parents to help improve the quality of education and educational outcomes. There should be a judicious level of informal teacher assessment and standardised tests without serious encroachment on class time and activities.

Assessment should cover all parts of the curriculum and all the various elements of learning—the cognitive, creative, affective, physical, and social development of students; their growth in self-esteem; the personal qualities being acquired; and the acquisition of knowledge, concepts, skills, attitudes, and values.

Assessment data for each student will be recorded on standard student profile cards. Parents will be guaranteed statutory right of access to their own children's school records and will be informed of their children's assessment outcomes. At the national level a system of monitoring standards, based on the regular assessment of the performance of a representative sample of schools, will be established to provide information on an aggregate basis in the Department of Education and to the general public.

Assessment in Second Level Education

There are two principal stages of summative assessment in second level education, the junior certificate examination at the

end of the third year (15 years of age), and the graduation certificate examination at the end of the fifth (or sixth year) (17–18 years of age).

Both examinations are characterised by a reliance on the written terminal examination. In some subject areas, there are also oral and aural examinations (languages) or practical and project examinations (mostly technical subjects, e.g., engineering and construction studies). At present, almost all components are externally assessed, that is, there is no direct involvement of the class teacher in assessment and, where this does occur to a limited degree, the assessment is externally moderated.

The White Paper indicates that the assessment methods should fully support the achievement of the full range of curricular objectives in the new junior certificate programme (introduced on a phased basis between 1989 and 1992). A special group is to be established to draw up a detailed timetable and programme for a fundamental restructuring of the assessment carried out at the end of the junior cycle to ensure that the full range of curricular objectives are evaluated, including an increased role for school-based (internal) assessment.

While external examinations will continue to be the main instrument for assessing achievement, a wider range of assessment techniques will be necessary. The techniques will evaluate the range of skills now demanded in a modern society, including thinking and innovation skills, practical skills, and research and problem solving skills as well as the ability to apply knowledge within the established graduation certificate.

There is a stated policy objective to increase the proportion of the marks awarded for oral and aural competence in Irish and the modern languages up to 60 percent of the total marks available.

The two other strands of the graduation certificate—the graduation certificate vocational programme and the graduation certificate—will use a wider range of assessment techniques. The former will use the assessment of a portfolio of course work as well as a terminal written examination for assessment of the vocational modules. In the case of the graduation certificate applied, there will be continuous assessment of tasks as well as terminal examinations. In addition, students will have an on-going report of their progress during the two years of the programme rather than, as is normally the case, having to wait until the programme has been completed.

The White Paper also indicates that a full evaluation of the standards achieved in specific subjects in the graduation certificate, by comparison with the highest standards in a representative sample of developed countries, will be undertaken over a ten-year period. This international benchmarking exercise is important for students, society and the economy in general.

The Inspectorate

There are approximately 170 inspectors and psychologists working in the Department of Education. The inspectorate will continue to be recruited nationally, but most inspectors will be seconded to ten regional education boards, to work under the executive direction and authority of the directors of these boards. These inspectors will constitute the regional inspectorate. A small number of inspectors (the central inspectorate) will be assigned to and organised within the Department of Education.

The central inspectorate's primary purpose will be to establish, evaluate, and promote the highest national standards of quality in educational provision. To achieve this purpose, the central inspectorate's core functions will be to evaluate and report on the standards and quality of the education provided, and the effectiveness of policies and their implementation; to advise on policy formulation; and to supervise the operation of the national examinations system.

The inspectorate will be reorganised into three units: an audit unit, a policy unit, and an examinations unit, under the direction of the Chief Inspector.

The audit unit will concentrate on evaluating and reporting on the educational services at regional and national levels. It will carry out this function through the preparation of reports on major curricular and other issues; monitoring and advising on the quality of teacher pre-service, induction; and in-career development programmes; monitoring and evaluating the effectiveness and outcomes of the national examinations system; systematic educational audits of the effectiveness of the education hoards; and commissioning surveys of attainment in curricular areas.

This work will be carried out on the basis of selective audits and a limited number of annual in-depth inspections of schools. The purpose of these evaluations will be to provide information and data on overall educational outcomes and standards

nationally and to provide benchmark national data for the evaluation programmes carried out by education boards.

The policy unit, a small group of expert personnel, will play a key role in contributing to the formulation and development of national educational policy and in designing policy implementation strategies.

The examinations unit will manage the professional and academic aspects of the operation of the national certificate examinations. It will continue to ensure that the examinations reflect the aims of the curriculum. and will also continue to monitor standards.

ITALY

CHIARA CROCE, Ministry of Education, Rome

In recent years, as part of its reform process, Italy has been involved in several efforts that parallel the themes of the three clusters of the OECD project, "The Curriculum Redefined," i.e., curriculum, teachers, and assessment. This paper summarises the efforts of this reform.

Curriculum

The process of reform involved the pre-primary education level (1991), primary education level (1985, 1990), and upper secondary education level (1990). New curricula have two fundamental common aims:

- To provide students with a sound basic education that can equip them to tackle problems in a creative and independent way; to find, in various situations, the most suitable solutions; and to acquire new knowledge in a discerning manner and to manage it in a focused way;
- To foster student attitudes and behaviour which favour socialisation, comparison of ideas, tolerance, and critical appraisal.

In upper secondary schools the most meaningful reform was to overcome the traditional dichotomy between general and technical education by strengthening and developing scientific subjects in classical schools (*Licei*); strengthening and developing general education in technical and vocational curriculum; developing the teaching of foreign languages; including philosophy, economics, and law in all courses.

Teachers

The reform process has been implemented gradually through experimentation that has involved, year after year, more and

more schools. Many in-service training courses have been addressed to teachers' needs with the aim of supporting them in implementing the new curricula, which require—in many cases—considerable changes in teacher behaviour. Key words such as team work, cross-curricular objectives, student-centred teaching can provide a sense of the changes that took place in teaching methods.

In implementing the reform, the autonomy of schools is deemed a very important element that must be strengthened. This autonomy, however, is not intended to isolate each school, but to recognise its individuality in relation to other schools. For this reason, the Ministry of Education promoted networks of schools at the national, regional, and local level working with focus schools, which are responsible for the development and production of models and instruments to be nationally disseminated.

The network structure is based on the idea that schools are able to introduce innovation only if they are successful in developing a communicative dimension. The process of communication and the exchange of information deeply influence teacher behaviour, foster innovation, and lead—by comparison—to self-evaluation.

Assessment

The curriculum reform in compulsory education introduced some important changes in the assessment of students (see below). Regards the monitoring and the evaluation of individual schools and the education system as a whole, there is a large debate on new strategies that are needed in relation to the autonomy of schools.

Formative Assessment

Formative assessment is recognised in Italy as a fundamental component of each educational activity, but in practice, it is implemented in different ways at different school levels.

In primary and lower secondary education formative assessment is found in all schools. As a consequence of school reforms, traditional marks were abolished and replaced by judgements which now account for the process of growth of pupils. In primary education, the Law of Reform introduced the principle of the "joint nature" of assessment, which gave a

strong impulse to a shift from summative to formative assessment, even though the shift has been anything but easy.

Teachers, in fact, were requested on the one hand to modify models of behaviour historically well established, and on the other to manage complex instruments such as new report cards which were perceived, at the beginning, as a heavy bureaucratic burden with minimal usefulness.

The Ministry of Education, aware of these difficulties, did not impose a sharp turnabout but supported the gradual adaptation of teachers to the new reality of assessment, for example through experimentation with the report card to which teachers contributed make them more simple and meaningful for pupils and parents.

In upper secondary schools, the implementation of experimental programmes and projects has led to a revision of the school organisation, with consequent adoption of new assessment methods.

The most motivated and sensitive teachers have also been implementing formative assessment where summative assessment, in the meaning that will be clarified below, persists and resists.

Summative Assessment

The meaning and importance of summative assessment in Italy can be understood only if one keeps in mind that it has been of fundamental importance for the transfer of pupils from one school year to the following and the attainment of the diplomas at the end of primary (11 years), lower secondary (14 years), and upper secondary (17–19 years)—diplomas that all have legal value.

In primary and lower secondary education the situation, as mentioned above, has been largely modified by Laws of Reform which introduced formative assessment. In upper secondary education, on the contrary, summative assessment is predominant. At this school level, teachers are requested to collect during each term (three to four months) of the school year a "proper amount of data" that are decisive for the final assessment of students, on which depends the transfer to the following year; and that are taken into account in the final examinations leading to the attainment of diplomas. Every

teacher collects data through recurrent assessment of the competency levels attained by each student in each subject.

Thus, summative assessment becomes a synthesis that very often does not produce feedback on the teaching-learning process and is, for this reason, ineffective at a pedagogical level. This situation, however, has been changing both as consequence of experimentation and because many activities have been put in place in order to create a new "culture" of assessment in which the "fetishism" of the mark will not influence negatively the teaching-learning process.

In particular, in the last two years the situation regarding formative and summative assessment has been modified by the abolition of the autumn examination session that gave students the opportunity to rectify their negative marks in the June session. Now teachers are asked to justify their negative summative assessment since the students no longer have a "second chance." Consequently, school reports and other traditional vehicles of summative assessment have been altered to include elements which place greater emphasis on the elements of formative assessment.

Student assessment is not used in Italy to monitor and evaluate individual institutions or the education system as a whole. Nevertheless, for the last five years the Ministry of Education has been implementing a project of pupil assessment at the secondary school level at the beginning and at the end of the first two years. This activity is important because it has been providing interesting information on the implementation of the new curriculum and the performances of pupils and of schools throughout the country.

GRAND DUCHY OF LUXEMBOURG

DOMINIQUE PORTANTE, Ministry of Education, Luxembourg

The traditional school system in Luxembourg reflects the location of the country, which finds itself situated between two great cultural movements; francophone and germanophone. All our children learn two foreign languages at primary school—German and French—alongside our national language, which is Luxembourgeois. After primary education they learn a third one (English), or even a fourth.

In terms of basic professional training, we offer "sandwich courses" (work placement and school) and full-time training in school.

However, too few young people obtain a basic qualification that gives them access to higher and university education, and too many young people leave school without any basic qualification. Additionally, the traditional teaching of languages has proved to be very selective for most foreign pupils and for many Luxembourgeois. These are the reasons why, over the last few years, several reforms have been introduced, or are being prepared in order to create new approaches to teaching.

With regard to evaluation in schools, the general orientation of educational policy had been defined in the consultation paper, *Demain l'école*, published by the Ministry of Education in 1991.

Evaluation, Advice on Choices, and Certification

Through their concentration on learners, by offering diverse and flexible learning pathways, and by diversifying their strategies, schools must be able to offer each child, each adolescent, and each adult an individual route in education. Evaluation of performance and choice of subjects are seen as tools which allow each individual to understand their position within their own personal learning plan, and to discover the routes which accord best with their aspirations, as well as with their skills and learning capabilities.

Evaluation and Advice on Choices

Evaluation should not be limited to the judgement expressed in a grade. To play its diagnostic and forecasting roles, it must, first of all, describe positive elements in terms of degrees of competence attained at a given time in a course of learning. Done in this way, evaluation allows for the necessary adjustments to be made during a course of learning; in this sense evaluation is formative. Pupils, advised by their teachers, learn how to evaluate their performance and so fix their own learning plan objectives. This process implies that pupils will be more responsible and motivated because they will have an opportunity to participate actively in the construction of their future; determining the rhythm of their progression, their choice of options and their courses. Such involvement will only be possible if the evaluation process is transparent; the learner must know the objectives to be achieved.

Evaluation must therefore become a tool for advising on choices; so that the choice of a course is based not on the elimination of other alternatives, or on failures, but upon positive criteria. The choice of course and the pace of progression become subject to continuous discussion on choices, in which the pupil is directly involved. In this manner, pupils become responsible for their options and for their results, advised by the teaching team who follow them within the school context.

The Role of the School Community

Education can be thought of as a responsibility shared between the learners, their parents, and the school. At the school level this responsibility lies with the local school community as a whole. The school community delegates part of this responsibility to its teaching team, made up of teachers, psychologists, and subject advisers. It falls to these teams to structure student choice and organise communication with their parents.

Certification

Evaluation is equally utilised in the giving of awards. Evaluation should be done, in general, on the basis of

attainments and not related to areas of weakness. Certification should conform to this new approach and be expressed in terms of the sum of units that have been passed. All individuals have the right to objective certification which lists the elements that they passed. No young person should leave school for the employment market without certification of his or her real achievements. This type of award should be seen in the context of life-long learning, within which each individual has the right to complete or continue their learning based on previous awards and professional or work experience.

The actions undertaken since 1991 are dependent on this general framework:

Primary School Level

At primary school level, a project of action research was set up in the area of language learning (DECOLAP, 1994–1996) with the aim of defining quality criteria for identifying good practice in learning. In this context, conclusions on formative evaluation (as a regulatory element of this learning process) have led to the formulation of a project aimed at the development of evaluation methodologies that are based on the activities of the pupils and on the development of their competencies (learning record). Particular importance is attached to communication among teachers, between teachers and pupils, and between teachers and parents.

Progression from Primary to Post-Primary

At the level of progression from primary to post-primary, a research project (1994–1997) was established with the aim of developing processes and tools that will allow, in the future, for pupils at the end of the primary cycle to be given advice on post-primary education routes, based on a record of competencies developed throughout primary education. Currently, this progression (selection) is still done through entrance examinations to secondary education and secondary professional education.

Secondary education

Since 1990–1991, a new system of progression has been introduced in secondary education and in Luxembourgeois secondary technical education. The new progression criteria, including the possibility to compensate occasional failures under certain conditions, had in effect a double objective:

- to motivate pupils to work regularly in order to obtain a higher overall level of achievement that would guarantee them, if need be, the possibility of compensation;
- to compensate for certain weaknesses in order not to penalise the occasional failures of the pupils.

An evaluation of the impact of this system on the progression of pupils has highlighted the fact that these objectives have only partially been achieved.

Within the same set of concepts, another reform, introduced in 1993–1994, concentrated on the reorganisation of the national examination at the end of secondary education. The main objectives of this reform were the following:

- to take into account, alongside written work, the oral communication skills, considered important for the social, cultural, and professional future of the pupils;
- to motivate pupils to work regularly;
- to encourage pupils to perform well by taking into account the whole year results and not over-penalising isolated failures (compensation of near-pass marks).

An evaluation of the impact of this system for the progressive acquisition of awards by the pupil allows one to conclude provisionally that, following this change, pupils produce better results, both during the course of the year and during the final examinations. However, with the objective of planning a more substantial reform, a research project (PERICLES) was set up in 1995 to analyse teaching programmes in secondary education in

the light of economic, ethical, and cultural realities of society and to redefine, if need be, the aims of secondary education.

Secondary Technical Education

In secondary technical education a very large reform project (project PROF) has led, since 1990, to a redefinition of objectives and methods. In order to adapt the evaluation modes for competencies leading to the certification of applied methods, a research project (Project PROOF) was begun in 1995. This project aims at elaborating, experimenting with, and testing new procedures of evaluation in a logical order.

Educational Indicators

In relation to the development of educational indicators, policy-makers have articulated for several years the desire to base decisions concerning these reforms on solid and systematic information relative to the state of health of the educational and learning system. In light of this, a research project was put in place at the end of 1995 with the aim of enabling policymakers in the Ministry to identify which indicators need to be provided. The project must lead, at the end of this current year, to a definition of the hoped-for indicators. These must concern the effect of teaching on pupils, on the quality of the infrastructure, on the curriculum; the development of quality in teaching establishments and among teachers; the cost; and the relationship between education and the jobs market. Later on, those indicators that are retained will be employed beginning in 1997.

MALAYSIA

NIK FAIZAH BTE MUSTAPHA, Ministry of Education, Malaysia

The 1995 Education Act

The vision or mission of the Malaysian education system is dictated by our national education acts and bills. Basically, the Malaysian Education Act 1961 is the core for the mobility and dynamics of the Malaysian education system. Besides this act, there were reports such as the Rahman Talib Report (1959) Razak Report (1963) and the Cabinet Report (1979), which provided the foundation for the formulation and realisation of the national education system. Our Parliament had revised the Education Act of 1961 and gazetted a new bill called the Education Act of 1995 (RUP 1995). The future changes and challenges were taken into account in this latest bill. The 1995 Education Act is formulated to reform the Malaysian education system towards our national goal, that is to take our nation into a developed country by the year 2020. In this act, there are provisions regarding the general policies and procedures for national testing, assessment, and evaluation.

The Structure of the Malaysian Schooling System

The new Education Act of 1995 provides the means for reformation in our school system. The democratisation of education, national unity through education, producing a progressive society through education, producing a well balanced (physically, mentally, emotionally, and spiritually) future generation are some of the core tasks with which the Malaysian schooling system has been charged. To achieve the national aspiration, the structure of the schooling system is being revised and reformed through the latest Education Bill.

Changes in Learning Assessment in Malaysia

The need for change in our learning assessment is influenced by the change in the concept of "learning," the change in the nature of what the students should learn, the expected role of assessment, the need for teachers to be knowledgeable and skilful in assessment, the need for more meaningful information, and teachers' extra workload. Changes include:

- Change in the emphasis of assessment—from subject-based to element-based;
- Change in what is assessed—from outcomes to process as well as outcomes of learning;
- The introduction of criterion-referenced measurement;
- The idea of giving more emphasis to internal assessment;
- The introduction of continuous assessment;
- The introduction of a separate reporting system.

Standards and National Testing, Assessment, and Evaluation

The homogenous curriculum of the Malaysian educational provisions brought about a national based testing, assessment and evaluation policy. There is an evaluation system that is centrally administered by the Malaysian Examination Board and the Malaysian Examination Council. Included are:

- Primary School Evaluation Test;
- Lower Secondary Evaluation;
- Malaysian Certificate of Education;
- Malaysian Higher School Certificate;
- Other Foreign Certificate Examination administered by Malaysian Examination Board.

The purposes of the national testing system include:

- Assessment of student learning outcomes at the end of learning cycles;
- express mobility, retention, and cooling-out;

- National monitoring;
- Certification and accreditation;
- Entry qualification for tertiary education.

School-Based Assessment (SBA)

School-Based Assessment is a formative testing system that is part and parcel of the teaching and learning process. SBA is implemented in all schools across Malaysia to make the practices of teachers' assessment of their pupils more systematic, refined, and reliable. SBA is conducted by schoolteachers for all subjects that are taught in the new integrated primary school and secondary school curriculum. SBA is a move towards achieving the aspirations of national education philosophy.

The scope of evaluation in SBA includes academic achievement, and behavioural development in terms of interest, attitude, values, and practical and psychomotor skills.

The SBA is meaningful in that it aims to improve the quality of classroom assessment; enhance learning in schools and minimise the emphasis on examination orientation; develop a conducive learning environment in schools, and improve the professional status of teachers in assessment and testing techniques.

Assessment methods include both formal and informal.

The SBA will assess the cognitive domain, psychomotor domain, affective domain, and a demonstration of skills in subjects like art, life skills, religious practices (PAFA), and other course works.

Implementation of SBA faces issues such as a lack of clarity about it among schoolteachers; a lack of expertise to develop effective teacher-made assessment instruments; too much emphasis still given to central examinations; and the use of professional techniques—such as observation—is still limited.

THE NETHERLANDS

MELIS MELISSEN, Ministry of Education, Culture and Science, Zoetermeer

Formative Assessment for Learning

Primary Education

The schools are responsible for pupil assessment. Many schools use a set of tests developed by the national testing institute Cito, called "pupil-monitoring system." This system measures growth all through primary education, and describes strengths and weaknesses of individual pupils. For pupils with a non-Dutch background there is another set of tests which measures school performance as well. Most schools with large numbers of such pupils use results to improve their teaching/learning processes.

Lower Secondary Education

Schools are responsible for pupil assessment. Most schools give written and oral tests that are marked on a 10-point scale. Pupils with marks 5 and below know that they should perform better, which in most cases is a very effective method of formative assessment, although more successful for able pupils than for less able pupils.

Apart from that, there is a system of national testing of the core objectives for basic education (12–15 years). This system is still developing. Schools can use national results of similar schools to compare their own performance in all (15) subjects, and draw conclusions from that comparison. Pupils can gain insights on their performance in basic education and draw conclusions from that as well.

Upper Secondary Education

Assessment in schools is the same as for lower secondary education. Additionally, recent proposals are under review to introduce a system of portfolio-assessment, to be introduced in

1998. These proposals have been well received widely, and they will be gradually introduced in some schools, also before 1998.

Summative Assessment

Primary Education

There is no obligatory summative assessment in primary education. However, most schools make use of summative assessment. About 70 percent of all pupils take the Cito–primary school completion test, for which they have to pay. Most other schools use something comparable to this test. Some schools are fundamentally opposed to this kind of testing and do not use school completion tests at all. Cito's school completion test covers Dutch language, arithmetic, information management and recently also *Wereldorientatie*, a test in which history, geography, and social studies can be recognised.

Lower Secondary Education

The assessment of pupils for grade transition is a responsibility of the school. Grade repetition in lower secondary education is about five percent per year.

Intermediate general secondary education and preparatory vocational education have a four-year course and receive one third of the cohort each. These types are completed with a school completion examination of six subjects for all pupils. Results are determined for 50 percent by a school-based examination and for 50 percent by a national examination in each subject. More than 90 percent of the pupils pass this examination.

Upper Secondary Education

Senior general secondary education (five years) and pre-university education (six years) are completed with the same kind of examination as lower secondary education. Pre-university education has seven subjects. Pass rates are 80 percent for higher secondary education and 85 percent for pre-university education.

National examinations are a very important feature of Dutch secondary education. There is no statutory national curriculum for contents or methods in the Netherlands, neither for primary

nor for secondary education. So the core objectives for basic education, and examination programmes (both set by the Minister for Education)—the national test for basic education and the national examination papers (both developed by Cito and set by a committee under the responsibility of the minister)—are the only normative national standards available. They have a very pronounced influence on textbooks, educational practice, and learning by pupils ("this is for your examination, this is important").

In primary education Cito's school completion test has a similar albeit less important influence. The inspectorate uses examination results, e.g., differences between results on the school-based and the national examination parts of the examination in their discussions with schools.

Systemic Reforms

Systemic Evaluation

In the Netherlands, all major educational reforms are evaluated. The introduction of a unified primary school for 4- to 12-year-old children in 1985 was evaluated by a committee headed by the inspectorate. This evaluation was followed up by policies for improvement of primary education along lines recommended by this committee's report (1994).

The introduction of basic education in 1993 for all 12- to 15-year-old children will be evaluated before 1998. In this evaluation the inspectorate will have an important role, but there will be other evaluations, based on scientific research and testing.

Apart from that, the inspectorate carries out other evaluations according to a yearly agenda. The national examinations are evaluated every year, and there have been evaluations of teacher training, special subject areas, and so on.

Another major focus of evaluation is the research commissioned by the National Foundation for Educational Research. This foundation does not carry out research itself but commissions projects to research institutions linked with the universities.

Policies on Curriculum and Assessment

As has already been stated, so far there is no national curriculum in Dutch primary or secondary education, in the way this concept is understood in other countries. Dutch education is for two-thirds a responsibility of private school authorities, and for one-third a responsibility of municipalities. The whole system is subsidised on an equal footing by the central government, with (small) additional subsidies by the municipalities, still on an equal footing. Although the minister is in a general way responsible for the quality of the educational system, it is not acceptable for the minister to impose a national curriculum. For primary education, the subjects are prescribed, and for secondary education, the subjects plus number of weekly periods. The content of the subjects is not described.

But, gradually, the relationships between schools and state are changing. Schools are held to be more autonomous in their inner functioning and more responsive to their environments. The number of weekly periods is no longer prescribed but advised. Schools are becoming freer to develop their own policies. But there are other developments as well.

In primary education core objectives were prescribed for the first time in 1993, and in the debate that followed the evaluation of primary education by the inspectorate these core objectives received a weighting which was not prescribed, but was not easy for schools to ignore. To sum up: more basic skills and less fringe.

In the second part of lower secondary education, a system of learning routes will be introduced in 1998, which will be described in ways which resemble national curricula in other countries. And, in upper secondary education, reform is being planned by a national steering committee. The reform of upper secondary will be very thorough: contents will be updated, most teaching will be a support for independent student learning and the system of free choice of six or seven examination subjects will be replaced by a choice from four different programs, with a common core of 50 percent of student hours, a concept to replace the concept of the "weekly period." This development of the school from a teaching environment to a learning environment is called *studiehuis* (house of study) in Dutch. All this will be legally introduced in 1998, but most schools have

already started activities to prepare or introduce this kind of learning.

NEW ZEALAND

DAVID PHILIPS, Ministry of Education, Wellington

As in many countries, New Zealand education policy makers encounter a major dichotomy: a regime of self-managing or autonomous institutions using assessment information to promote better learning, contrasted with increasing central control through a national common curriculum, to ensure system accountability. New Zealand's assessment policies reflect this dichotomy.

Systemic Reform

Developments in assessment in schools need also to be seen in the context of major economic reforms and ongoing dramatic changes in New Zealand's education and training systems, including restructuring of the administration of education, and the governance and management of schools, since 1989; complete reform of the national curriculum from 1991, due to be completed in 2001; and a new system for recognising national qualifications due to be fully implemented by 1998.

Curriculum and Assessment Reforms

The curriculum and assessment policies currently being implemented emphasise the improvement of student achievement, and as part of that broad objective, the clearer specification of what students need to learn and more rigorous assessment of their achievement. The New Zealand Curriculum Framework (NZCF), released in April 1993, is the foundation policy statement covering teaching, learning and assessment for New Zealand schools. It sets out the principles that give direction to the curriculum, specifies seven essential learning areas (language and languages, mathematics, science, technology, social sciences, the arts, and health and physical well-being) which describe in broad terms the knowledge and understandings which all students need to acquire, establishes eight groupings of essential skills (communication, numeracy,

information, problem-solving, self-management and competitive, social and co-operative, physical, and work and study), and indicates the place of attitudes and values in the curriculum. The NZCF also outlines policies for the assessment of student achievement.

Supplementing the NZCF, new national curriculum statements—replacing existing syllabuses—are being developed which describe the national achievement objectives in the essential learning areas. These statements include implementation suggestions such as teaching and learning experiences, and examples of assessment activities. Schools use the statements to ensure that their programmes enable all students to meet the requirements of the New Zealand curriculum.

New national curriculum statements have been finalised for the essential learning areas mathematics, science, language and languages (English and Te Reo Maori) and technology, are in draft for social studies, and in development for health and physical well-being, and the arts. Maori versions of the statements are also being written. A phased programme of implementation (one statement per year) is underway.

National Education Goals

Schools are subject to a legislative framework within which the country's over-arching education goals are defined. These include goal 1: "The highest standards of achievement, through programmes which enable all students to realise their full potential as individuals, and to develop the values needed to become full members of New Zealand's society"; and goal 6: "Excellence achieved through the establishment of clear learning objectives, monitoring student performance against those objectives, and programmes to meet individual need."

Assessment and Accountability

Assessment requirements for schools are spelt out in the National Education Guidelines which outline the responsibility of school boards, through the principal and staff. They include the requirement that:

Boards of trustees must foster student achievement by providing a balanced curriculum in accordance with the national curriculum statements[1] (i.e., *The New Zealand Curriculum Framework* and other documents based upon it).

In order to provide a balanced programme, each Board, through the Principal and staff, will be required to:

- implement learning programmes based upon the underlying principles, stated essential learning areas and skills, and the national achievement objectives;
- monitor student progress against the national achievement objectives;
- analyse barriers to learning and achievement;
- develop and implement strategies which address identified learning needs in order to overcome barriers to students' learning;
- assess student achievement, maintain individual records and report on student progress.

Formative Assessment for Learning

A fundamental tenet of the NZCF is that the primary purpose of assessment is that information from assessment of individual needs is taken into account in teaching and learning programmes. This necessarily implies that most assessment activity is classroom based and teacher-owned and -operated.

The required learning in the national curriculum statements is organised according to a progressive series of achievement objectives, usually specified in eight levels covering years 1–13, which provide a basis for assessing achievement as students progress through the system. This level model is not a narrow, linear one, nor are the objectives at each level tightly age or class-related. These objectives identify, in levels of increasing cognitive complexity, outcomes that developing learners should be able to display.

As each new curriculum statement is introduced, it will enable more coherent, and systematic reporting to students, parents,

[1]Existing syllabuses are to be regarded as national curriculum statements until they are replaced.

and the wider community about what the students actually do. Assessment is seen as an integral part of the curriculum and of the teaching and learning process.

To assist teachers implement the new curriculum statements and develop new ways of assessing students' progress, a range of professional development programmes is under way. The Ministry also has a role in research into such issues as recording and reporting, for example, on data aggregation. An assessment handbook (*Assessment: Policy to Practice*, 1994) has been developed for schools, to help them develop their policies and this is being supported by a series of curriculum and assessment guides in various subjects (e.g., *Developing Science Programmes*, 1995). A further statement on assessment, issued in the *Education Gazette* in March 1995, made it clear that schools were not expected to assess student achievement against all the national achievement objectives in the same year.

Using Summative Assessment

The National Education Guidelines require schools to monitor student progress against national objectives, to maintain individual records, and to report their progress. In any event, communities expect schools to report progress of individual students to parents. The guidelines also require schools to collect and analyse data at a school-wide level in ways that will enable them to identify and respond to the needs of under-performing sub-groups, as well as diagnosing and responding to individual needs.

Institutional Monitoring

New Zealand's Education Review Office, which regularly audits schools for evidence that they are fulfilling legislated requirements, reports that few schools have good systems in place at present. To assist schools to analyse barriers to achievement and to report to their communities, the Ministry of Education proposes to develop a standard recording and reporting format for schools to use.

The Ministry is looking at ways teacher-collected school-wide information can be collated centrally to assist with outcome trend analysis and population sub-group analysis. The interpretation possible with the relatively broad achievement objectives

specified in the curriculum poses a number of challenges here. The lack of validity and reliability checks on the tools which teachers use to collect data on students means that this school-based information cannot be used for external accountability or school comparison purposes, notwithstanding a common recording and reporting format.

The government has determined that there should be limited reporting of some aspects of the curriculum in a comparative way. The arguments advanced in favour of school by school achievement-based reporting include the ability to gain information about value added by individual schools, to target additional resources to under-performing schools, and to provide information to assist parental choice of school. Tensions surrounding the issue of local autonomy and central control are evident here. The proposed system consists of school entry level assessment and assessment at transition points where the majority of students transfer between schools. The first national assessments will be made in 1997.

School Entry Level Assessment

The system aims to assess all children within their second month at school and will be administered by classroom teachers. Its development focuses on simple numerical concepts—through the development of a board game based on research into counting, numeration, and classification; observations of reading readiness based on a commonly used diagnostic tool, *Concepts about Print*; and a language assessment rubric based on story re-telling.

Transition Point Assessment

The majority of New Zealand students attend primary school in Years 1–6, intermediate school in Years 7–8, and secondary school in Years 9–13. Transition point assessments in mathematics and science, and eventually other curriculum areas such as English, are to be administered to all students at the interfaces between Years 6 and 7, and Years 8 and 9. Resource banks of assessment items, based on the published achievement objectives of the curriculum in mathematics and science, have been developed and trialled by the New Zealand Council for Educational Research. Teachers can use these at any time for

diagnostic purposes or to gauge comparative performance within a school or classroom. Electronic access to the items is also being investigated.

It is also planned that, each year, schools will be provided with a set of items to be administered under standardised conditions. As a result, the Ministry is investigating the extent of teacher professional development and support required to ensure accurate administration of the items. A current issue of intense interest is the possibility that results of school entry level assessment and transition point assessments will be published in "league tables." Such a move would be controversial, and would heighten debate over possible uses of the data, particularly if the information assumes high stakes. The Ministry is therefore also investigating different systems of comparing school performance with indicators of national performance, or with schools of similar intake characteristics.

System Monitoring: National Education Monitoring

The principal aim of national education monitoring is to show changes in national achievement over time. It will not provide information on individual students, teachers, or schools. National monitoring took place for the first time in October 1995 in science, art, and information skills. All areas of the curriculum, including skills and attitudes, will be covered on a rolling three year cycle.

National monitoring involves a 3 percent sample of 8- and 12-year-old students each year. Tasks are administered by visiting teachers, trained as test administrators. Each pair of visiting teachers remains in a school for one week working with 12 students over three days. The tasks use a variety of formats and are designed to show the full range of capabilities of the students and to assess students' abilities to apply essential skills across the curriculum. The tasks are not tied explicitly to objectives of the curriculum as one of the important goals is to provide evidence for curriculum review.

Certification of Student Achievement

National examinations. The New Zealand Qualifications Authority (a Government agency) administers national awards for the secondary school sector at the end of years 11, 12, and

13. These are respectively: New Zealand School Certificate; New Zealand Sixth Form Certificate; and New Zealand University Entrance, Bursaries, and Scholarship. The possible placement of School Certificate at the end of year 10, as the National Qualifications Framework is implemented, is being considered.

National Qualifications Framework. This new qualifications system, being developed by the New Zealand Qualifications Authority, will integrate all qualifications into a single framework. Some qualifications will be registered in their entirety, such as university degrees. Others will be based on a unit standard structure, i.e., the key elements (skills, knowledge, and understandings) and the standards to be met, are tightly specified. Assessment will be based on evidence that the learner has met all performance criteria stated in the unit standard. Unit standards for conventional subjects are being written from national curriculum statements and syllabuses.

International Surveys of Achievement

Since 1968, New Zealand has participated in international tests and now participates in the TIMSS, to compare its students' achievements compares with international norms .

Conclusion

The challenge for national policy makers is to ensure that the assessment reforms reach an appropriate balance between improving student learning and monitoring institutional performance.

NORWAY

OLA MOE, Central Government Office of Education, Sør-Trøndelag

Norway—A Systemic Reform Case

Norway is in the middle of its most comprehensive educational reform period anytime. During the years 1993—1995, new laws were passed concerning every stage from kindergarten to higher education. In 1994, pupils in upper secondary education and vocational training started with a complete new set of curricula for every step and subject. In 1996, a new common law for the universities and regional colleges was put into action together with a completely reorganised system for the regional colleges. (Approximately 110 different colleges were restructured into 20 institutions.) In 1997, primary school and lower secondary education will change to a new set of curricula. Beginning August 1997, six-year-old children—rather than seven-year-olds—will enter school for the first time.

One of the advantages of the extensiveness of the reforms is that politicians, bureaucrats, schools, teachers, and the public have been engaged in an important and rather thorough discussion on national educational policy and its elements. The problem, of course, has been that a lot of people have been rather busy trying to keep time-schedules and produce the needed new material.

Since education in Norway is predominantly public and nationally regulated (very few private schools and relatively few institutions for higher education; national curricula for primary school, secondary education, and vocational training), it has been possible to try to build the reforms on some basic principles and guidelines:

- Education must be updated in view of life and problems in the society of our time—a rather broad concept of learning, seen as a whole:
  A. A need for new knowledge and the ability to use knowledge in problem solving tasks;

B. The handling of modern technology, internationalisation, and the importance of reflection about "national culture and identity";
C. A need to recognise and resolve environmental problems;
D. The need to solve growing social problems;
E. Strengthening of ethics and democratic and human values.

- Norway has a strong tradition in trying to realise "equal opportunities" and "equity:" Everybody should have an equal opportunity to get an education:
  A. That does not differ in quality
  B. In a comprehensive school system for pupils from 6 to 19 years of age
  C. That follows a "national standard" of content with common subject elements, but at the same time tries to differentiate the education to match the individual pupil's talent and need—realising an individual right to "meaningful education"
  D. That, within the framework of a centralised, national curricula, provides room for local initiatives, local adoption of plans, and local examples.
- Education from primary school through upper secondary school and vocational training is to be seen as a whole, and curricula should be written to take care of this.

A Systemic Reform

The reform is planned and managed by the Ministry of Education. The ministry has stated the underlying principles in a general part of the curriculum. This general part is common for primary school, secondary education, and vocational training. It is called the "Core Curriculum" but core curriculum here means general principles and not core content or national curriculum in the usual sense. A government White Paper, *Stortingsmelding nr. 29, 94–95* (*St.meld 29*), discusses the principles and governing lines. Both the core curriculum and the White Paper have been presented to the Parliament. The members of

Parliament have stated their views on the White Paper in *Innstilling S. nr. 15, 95-96.*

As a supplement to the core curriculum, national curricula for the different subjects are being formulated for primary, secondary and upper secondary education, and vocational training. The Ministry of Education has organised expert groups to draft the curricula. The drafts are subjected to a national consultation. After the consultation, the Ministry of Education will make the necessary adjustments and authorise the curricula.

The ministry has also devised a strategy for implementation of the reform, including:

- information about the reform;
- spending government money on in-service training;
- support to municipal school-leadership programs;
- a research program of evaluation, including reports from the Central Government Offices of Education among others.

Quality of Education: Assessment and Evaluation

Norway has been less occupied with the discussion of national standards and assessment than many other countries. For instance, there is no formal assessment during the first six years of primary school. To understand this, it is necessary to link back to the Norwegian focus on "equity" and the tradition of a compulsory, comprehensive school (*enhetsskolen*). It is widely believed that a too strong focus on national standards and formal assessment will result in a stratification of schools and possibly a redistribution of pupils, and that this can endanger both the principle of an education adapted to each pupil's ability and the attempt to keep the school and classroom a place for social integration where pupils of different background and talent work together.

On the other hand, it is recognised that the Norwegian school system feeds back very little exact information to administrators, politicians, and the public. An OECD report in 1988 made this very clear. The ministry is now working to develop a system that should:

- give such information
- define areas and methods for evaluative analysis

- establish a governing structure on both national and local levels that uses the information for further development.

This evaluation system follows three lines of approach: Assessment of pupils, school-based self-evaluation, and a national system of evaluation (*St.meld 29*, ss 47–49).

Assessment of pupils

In Norwegian terminology *formative* and *summative* assessments are covered by the terms *informal* and *formal* assessments. The two aspects should be kept together, however, and also complement each other, so that the whole range of attainment goals can be covered.

The intentions of assessment are to:

- Inform pupils, parents, teachers, and schools
- Guide, motivate, and help the pupils to find new directions
- Inspire teachers and schools to find areas and ways of improvement
- Certify through diplomas and provide information to other parts of the educational system, employers, and the public

The assessment is learning-oriented. As a whole it is expected to be broad and to cover:

- Knowledge, understanding, and ability in problem-solving
- Skills
- Participation and engagement in different kinds of activities.
- Initiative and concentration
- Creativity
- Independence in work
- Co-operation, reliability, and a feeling of responsibility towards other pupils
- Social contribution
- Punctuality, structure, order, and tidiness in work
- Behaviour

Informal assessment is provided from the first day of school, in the form of reactions to work, guiding, conferences with pupils and parents, written notes to parents on certain dates during the year, and so on. An important intention with the informal (formative) assessment is to develop the pupil's own ability to plan and evaluate his or her own learning.

Informal assessment uses attainment goals as standards, but is oriented more towards individual progress and talent than ranking pupils within groups or according to national standards.

Formal assessment in Norway (summative assessment), is introduced in the 7th grade. During the first six years all assessment is informal and does not include the use of marks. Beginning in the 7th grade, pupils are given marks two or three times a year in the different subjects. The marks are written down in specified booklets and parents have to sign that they are seen the marks. At the end of upper secondary school, some subjects are allotted each year for nationally given written examinations, or locally handled oral examinations with external examiners.

Formal assessment (with marks) reflects attainment goals, yet is group-oriented (ranking), and the marks should overall correspond to a normal variation (Gauss).

School-Based Evaluation

Assessment of pupils is seen as only one method of ensuring quality of education. Since 1974, school-based innovation, and school-based evaluation as part of an innovation strategy, have been looked upon as most important methods. Both school leaders and teachers have been given opportunities to take courses in innovation processes, and government money has been used as an incentive to establish innovative projects. It is an important element in the "quality thinking" that the schools have to see themselves as, and act as, "learning organisations" if the reform policy is to succeed. The school-based innovation activity is important for the development of the local supplements to the national curriculum, and is a way of making teachers more professional.

The ministry has provided a guidebook on school-based evaluation called *Underveis (On the Road)* that is sent to all the schools in the country free of charge.

A National System of Evaluation

During the last decade, the national regulation of the educational system has become less detailed and more oriented towards goals and achievements. The content of education (curriculum) is a centralised, national matter in Norway, while local authorities have the duty to put the national prescriptions into action. The national evaluation system is meant to be a way of getting feedback, to ensure that the local authorities are loyal to the national policy and to yield necessary information as a basis for analysis and adjustments of national policy.

The national system of evaluation includes assessment of pupils, but not results from school-based evaluation, which is seen as the school's own matter. In addition the system includes:

- Statistics of input of resources, money, numbers of lessons;
- Organisational data: Numbers of schools, classes, pupils;
- Specific area surveys defined by the ministry; for instance teaching of foreign languages, science, or music;
- Nationally administered tests of reading and writing abilities, mathematics or other subjects;
- Nationally administered questionnaires dealing with school planning and leadership, cooperation with parents and so forth.

PORTUGAL

MARIA DO CÉU ROLDÃO, Institute for Innovation in Education, Ministry of Education, Lisbon

The three themes of the conference are obviously intertwined and all of them relate to the present situation of a curriculum reform being implemented at a national level in Portugal, from 1989 until now (Act n° 286/89).

The concept of assessment within the philosophy of the curriculum reform is guided by several main principles:

Assessment is perceived as a necessary tool to improve learning, centred upon a philosophy of promotion of success for all which informs the curriculum reform. The whole philosophy of curriculum and education reform lays on the General Principles of Educational System—Bill n° 46/86. According to Act n° 286/89, "assessment should meet both the need of monitoring the quality of educational system and to guarantee the actual implementation of the educational principles established by Bill 46/86."

Formative assessment is particularly emphasised as a means to adjust and improve learning all over the educational process, with implications at the level of teachers education and parents' attitudes towards the role of assessment in education.

Summative assessment is to be considered as a necessary step within a global process of evaluation of students learning and outcomes. This view opposes the traditional emphasis on school marks as the critical issue in school, from the point of view of parents, students, and teachers.

The recognition of the need to assess the system as a whole at a national level is increasingly a critical concern for the Ministry of Education, that is now implementing some modes of national monitoring, namely in mathematics and language, at the elementary level (Grades 1–9), and national exams for every subject matter at the secondary level (Grades 10–12).

Formative Assessment for Learning

The Portuguese model of assessment for elementary schooling—Grades 1 to 9—that was established in 1991 (Norm n° 162/ME/91, later withdrawn and replaced by Norms n° 98-A/92 and n° 338/93) emphasises the formative role of assessment and the principle of individualisation of learning.

At the elementary level, students progress through each of the three cycles (Grades 1–4, 5–6, and 7–9) without a decision on passing or failing at the end of each school year—which was the previous practice. Teachers are told to respect differences in the mode and pace of learning of individual students, and provide constant feedback and differentiation of activities in order to enable every student to follow his or her own process of learning. Teachers decide at the grading meetings which students are in particular trouble in reaching the curriculum objectives. For those students, a personal plan of work and a report on their achievements are prepared to support their progression in the following school year. This has been a major innovation in the Portuguese assessment system, and some difficulties and resistance to its implementation are still in place, both in schools and in families.

Another area of emphasis on formative assessment relates to the increasing need of taking into consideration the cultural diversity that characterises our society in these days. Phenomena of social exclusion are also a relevant problem, particularly in the larger cities such as Lisbon, Oporto, and Setúbal. The response of schools to these particular needs relies largely on an adequate use of formative assessment, which is expected to enable teachers to analyse the difficulties of each student from a social or a cultural minority in order to adjust learning materials, activities, and expected outcomes to their diversity.

A major novelty, though practised before the curriculum reform, refers to the inclusion in curriculum objectives of skills and attitudes—namely those required for citizenship—in addition to content objectives. Assessment is expected to consider all of these dimensions and to de-emphasise a content orientation that was predominant or even exclusive in the past. Progression in all those domains is to be considered and improved through formative assessment.

Summative Assessment

The current situation on summative evaluation is different at the elementary and secondary levels:

At the elementary level, summative assessment is expressed in quantitative value judgements on a 1–5 point scale; teachers test their students' performance each two or three months, and there are grades in the second and third cycles, which, however, do not imply failure at a given grade. At the end of each cycle of basic schooling, it is decided whether the student will be promoted or not to the next cycle. After the first cycle, a summative evaluation is expressed in qualitative terms, in a descriptive report, based on a set of criteria.

At the secondary level, summative assessment is expressed on a 0–20 point scale, and implies a decision that affects the student's progression to the next grade at the end of every school year. At Grade 12 there are (starting in the present year) national exams. The previous system had included national exams in Grade 12 for one nuclear discipline for each student, defined according to the university courses the student intended to pursue later. Summative assessment is, in every level of the system, criterion-oriented and referenced to pre-defined instructional objectives. As described above, even summative evaluation at elementary level has a strong formative component, given the need to establish an individual plan for those students with low achievement.

Certification and progression at the end of each level depend on summative evaluation marks. Some studies of achievement, as perceived through students' marks, have also been developed at the Institute for Innovation and by some other departments of the Ministry.

Monitoring Assessment

As mentioned before, there is now an increasing concern with the assessment of the proficiency of the educational system related to increasing societal demands and expectations towards schools, in order to respond to the needs of present time. National exams—though limited to Grade 12 at the end of schooling—will provide a picture of the way the system is working.

The Portuguese Ministry of Education has assigned the task of managing the assessment of the system, both at the elementary and at the secondary level, to the Institute for Innovation in Education (IIE), an institution within the Ministry of Education.

The Institute developed a study to build a conceptual framework in order to design tests as one of the instruments of national monitoring. As a first step, a three-year project—that is now being implemented—was then launched. It focused on the level of attainment of students in mathematics and language at Grades 4, 6, and 9 by designing standardised achievement tests in those subjects. A curricular analysis of these subjects was developed previously to build a coherent framework for the study. The expected results of the study relate to teaching quality, curriculum structure and organisation, interrelationship among educational contexts, teaching and learning conditions, and products of the educational process.

Systemic Reform

As described above, it is clear that current curriculum issues integrate a systemic view of the educational system that, in turn, relates to the broader social system. Portuguese educational policy reflects the awareness of the systemic nature of educational decisions and practices, as in the following concerns referred to above:

- The interrelationship among assessment, curriculum, and educational principles;
- The dependence of achievement on teaching, context, learning conditions, and social and cultural conditions;
- The role of education in developing skills and attitudes for citizenship.

Legislation

Educational Bill 46/86. *Lei de Bases do Sistema Educativo, lei 46/86 de 14 de Outubro.*
Act n° 369/90. *Decreto. Lei 286/89 de 29 de Agosto.*
Norm n° 162/ME/91. *Despacho n° 162/ME/91 de 31 de Julho.*
Norm n° 338/93. *Despacho 338/93 de 29 de Setembro.*

SCOTLAND

IAN BOYES, Scottish Office, Edinburgh

Formative Assessment for Learning

In the 5–14 programme, assessment is regarded as an integral part of planning, teaching, recording, and reporting; and the starting point for obtaining assessment information is the curriculum and the process of learning and teaching. Planning requires teachers to know and share what is to be learned; assessment is part of effective learning and teaching; recording summarises success and progress; and reporting provides useful feedback to learners and parents. Assessment evidence can also be used to evaluate learning and teaching. National guidelines were published in 1991 that emphasised formative assessment and covered the assessment of all aspects of education such as the knowledge, skills, and attitudes specified in curricular advice and also the values, interests, and talents to be fostered and encouraged in young people from the age of 5 to 14. These guidelines were in three parts: part 1 described the role of assessment and its place in the school; part 2 provided guidelines for assessment in schools; and part 3 was a staff development pack for teachers. These documents were based on the levels and targets of attainment in the 5–14 national curriculum guidelines.

In order to assist teachers in formative assessment, further advice was provided on diagnostic procedures, initially covering mathematics, language, and science; a further document on social subjects is in preparation. This advice was in the form of a general overview that described why teachers should from time to time take a closer look at pupils' attainments, and the strategies they can use to promote success in learning. Specific advice on the three subjects was also issued, describing learning in the subject area, the practicalities of collecting information, gathering and interpreting evidence, advancing pupils' ideas and skills, and giving examples of diagnostic procedures in action. This advice was designed to give teachers more insight into the ways in which pupils learn, and to encourage them to be more

systematic in their interventions as they plan "next steps" in learning with their pupils. It was also decided to place this advice in the daily routine of learning and teaching by encouraging teachers to use the everyday evidence of pupils' work for diagnostic or formative purposes, rather than providing them with ready-made diagnostic tests.

Quite separately, a system of national testing operates in reading, writing and in mathematics, using the levels and targets outlined in the curriculum and assessment national guidelines. Test items are prepared by a national assessment unit and are written by teachers to reflect the normal classroom experience of pupils. Teachers choose items from a catalogue and mark the tests themselves.

Using Summative Assessment

Summative assessment provides systematic information about the attainment of pupils over a course or programme of work. It can contribute to pedagogy, for instance the national tests mentioned in the previous section define whether pupils have reached particular levels of attainment, thus enabling their teachers to plan appropriate future teaching programmes. Many teaching schemes, commercially produced and school produced, contain assessment tasks which determine whether pupils have attained the knowledge and skills required at particular stages to enable them to proceed to further work.

Summative assessment also provides the basis for the certification of pupils' achievements. In Scotland, all pupils take Standard Grade examinations at age 16 which provide a summative assessment of the courses they have taken in their third and fourth years of secondary schooling. These examinations are at three levels: credit, general, and foundation, to cater for the wide range of ability of pupils. These examinations are set by the Scottish Examination Board (SEB), which also sets higher grade examinations for fifth and sixth year secondary pupils and the Certificate of Sixth Year Studies (CSYS) for sixth year pupils. The higher grade system is being reformed as described later in this paper. In the final years of schooling, pupils can also take modular courses by the Scottish Vocational Education Council (SCOTVEC), which are certificated based on assessments by the institutions offering the courses.

There is also a national monitoring system in Scotland, the Assessment of Achievement Programme, which carries out tri-yearly surveys of representative samples of pupils at age 8, 11, and 13 in mathematics, English language, and science using wide-ranging batteries of assessment tasks, both written and practical. Published results describe national levels of attainment, gender differences in performance, performance by ability of pupils, and changes in performance since the previous survey. There is a longitudinal element because pupils are tracked and reassessed at two of the grades and the intention is to track them through to certificate examination results at age 16 and beyond.

There is no national collation, analysis, or reporting of pupils' performance in national testing inside the 5–14 programme. Its main purpose is to support teachers' continuous assessment and to convey information about the achievement of individual pupils in relation to nationally agreed standards that teachers can use to check their own assessments and that will assist teachers' consistent interpretation of attainment targets and levels. There is no set, fixed time at which pupils are tested using national test items. Tests are set at the five levels of attainment which describe the performance expected of most pupils at various points across the seven years of primary school and the first two years of secondary school. Pupils in one classroom may therefore be tested at different levels, reflecting their different learning styles and their different rates of progress. The decision is for the teacher to take when he or she judges that a pupil is consistently and confidently demonstrating that he or she is performing at a given level. The test should then confirm the teacher's judgement.

Systemic Reform

Two major reforms are in progress in the Scottish school system: the 5-14 programme and the Higher Still programme.

The 5-14 programme was introduced in 1989 and its implementation is still ongoing. Its aim is to provide a broad and balanced curriculum for all pupils and to ensure, through improved assessment practices, progression in their learning from year to year and from primary to secondary schools. Curriculum guidelines have been prepared for pupils between the ages of 5 and 14, based on assessment levels, outcomes, and

targets in all subject areas. Guidelines have also been provided on assessment and reporting. Steady progress has been made in schools in implementing the guidelines, which are based on the best practices in Scottish schools.

The aim of the Higher Still programme is for students aged 16 and older to achieve the highest standards of which they are capable in a wide range of courses, both vocational and academic, to provide qualifications for all students with competence in core skills and to ensure an even gradient of progression. The Higher Still programme brings academic and vocational courses into a unified curriculum and assessment system. New curricular guidelines are being developed for modular courses which will be assessed internally and externally. There will be an exit point for pupils at Higher level (normally after five years of secondary education, aged 16–17) but there will also be Advanced Higher courses, normally after six years of secondary education, aged 17–18, which will incorporate the current Certificate of Sixth Year Studies courses. The new examinations will be introduced in 1998–1999.

SWITZERLAND

JACQUES WEISS, Institut Romand de Recherches et de Documentation Pédagogiques (IRDP), Neuchatel[2]

In 1986, the Conference of the Directors for Public Instruction in Switzerland published 22 theses for primary school development. Five of them concerned the evaluation of pupils.[3] The conclusions of this national research are the source of most of the reforms in assessment undertaken in compulsory schools. Furthermore, two working groups, one for all Switzerland and the other for the French and Italian speaking populations, received mandates to elaborate proposals, the first group to work on the certification at the end of compulsory school, the other on a comprehensive review of evaluation for compulsory schooling.

This present report describes the current reforms, mainly at primary school level. It also describes some of the most representative trends in evaluation in the schools of the Swiss cantons.

---

[2]Urs Vögeli Mantovani, Edith Wegmuller and Andreas Dick contributed to this paper.

[3]The appreciation of the pupil's work is useful when its forms and functions are clearly defined. It is particularly important to distinguish formative evaluation from others, such as summative evaluation and prognostic evaluation. Conditions of learning are optimal when the pupil knows clearly the objectives pursued and receives regular information on his progress. In this perspective, the most adequate form of evaluation is formative evaluation. Evaluation allows the teacher to discern the specific learning difficulties of his pupils, and to foresee and adapt his teaching in relation to their possibilities of learning. Responsible and efficient learning by the pupil depend upon realistic knowledge of his own possibilities. One means for encouraging this personal knowledge is to lead the pupil to manage his learning by self-evaluation. A pupil's formation is a task for the school and the family. Frequent information on his progress and difficulties permits his parents to help him with his learning and to give him appropriate guidance.

Preamble

The school is entrusted with the development of all the child's potentialities by offering him different modalities of teaching information. It is further entrusted to certify the pupils' acquired competency and to regulate the educational system itself. Its instrument is evaluation.

Considering the different functions, it is fitting to note that evaluations are centred more on the pupils' learning and on choosing the best methods of training. In this case, the evaluation has educative and formative aims. On the other hand, evaluations also focus on the certification of the pupils' proficiency and knowledge, and on the regulation of the system. In this case, the evaluation has summative and institutional aims.

All possibilities exist between these two kinds of evaluation; the pupils' personal development and the control of the system are two permanent aims, and are in constant tension within every educational institution. The research for coherence can, nevertheless, be attempted.

The theses of the Conference of the Cantonal Directors for Public Instruction in Switzerland, as well as the most recent innovations undertaken in the cantons, give priority to evaluations with educational and formative aims.

Evaluation with Educative and Formative Aims

In fact, this is the first approach taken for evaluation in Switzerland. This evaluation permits the pupils to learn and develop. It is actually the school's first task. Educational organisation must be dominated by a "logic of the subject," a logic based on the "elaboration of learning projects,...a step that aims to develop a sense of responsibility, a personal judgement, a critical mind." It is important that the pupils "may progressively be able to give a sense to their learning and elaborate a formation project that agrees with their interests, their aptitudes, their ambitions."

This educative and formative evaluation, by self-evaluation and co-evaluation, will be examined, according to two different but interdependent perspectives: training and informing. They will be illustrated by example from the Canton of Geneva and the Canton of Fribourg. These focus respectively on the portfolio record of the pupil and on an interview for evaluation.

Evaluation for the Pupils' Formation: The Portfolio

Formative evaluation is a part of teaching. It is interactive, centred on the learner but upheld and animated by the teacher according to a practice entitled "directed self-evaluation" or "assisted self-evaluation." It favours awareness by the pupil of his or her progress and capacity on the basis of the quality of the schoolwork, which is assembled in a school record or formative portfolio. This awareness is guided by referring to the objectives of the curriculum and of the "didactic contract," and requires the teacher's support. The teacher's role is decisive in the analysis and appreciation of the work done by the pupil. The teacher compares the estimations of the pupil with the objectives of the curriculum or contract, as well as the criteria of "sufficiency." The teacher suggests complementary courses, and adjusts any over- or under-estimation.

Illustration

The pupil's portfolio or record in the Canton of Geneva:[4]

> In the primary school classes of Geneva, the practice of having a pupil's record is developing, the aims, functions and utilities not being clearly established. The portfolio is a record of the pupil's "authentic" production, chosen and commented on by the child. The self-evaluative commentary concerns the choice of each production as well as a recapitulatory flyleaf of the record. They focus on success and difficulties met in learning their subjects (subjects often considered to be a priority by the school), on the attitudes of motivation, on the perception of oneself in learning in general.
>
> The objective which has priority is the development of each pupil, according to his capacity, the approach of self-evaluation and meta-cognition. These tests, in the long run, aim at activating the process of anticipation, verification and adjustment, which will lead to auto-regulation. It is more a question of an approach to

[4]Extract from an article by Edith Wegmuller, responsible for the "evaluation" sector of the direction of primary school teaching.

learning than a means of evaluation. With 10–12 year old pupils, that means having the capacity:

- to be aware of what he has learned;
- to select among the equipment used and developed while learning in the classroom;
- to choose the production that is representative of his strength and weakness;
- to add to and assume his choice;
- to go beyond a global evaluation of the kind "I like, I don't like";
- to go into an analysis of how he works, how he proceeds;
- to confront his opinions with those of the teacher.

This teacher-pupil discussion, based on the portfolio, aims at the process of co-evaluation, the source of regulation and the planning of actions.

The elaboration of a portfolio enters into a certain dynamic of the class which enhances the capacity of being autonomous, responsible and sociable. The terms of the contract, negotiated individually, have led the pupils to assume their choice, to manage their time, and to work according to their possibilities. This record enters resolutely into the perspective of formative evaluation, useful for the pupil to begin with, as a means, among others, of bearing results on the development and learning of the child. It remains, therefore, an inside tool for the classroom.

Evaluation from the Point of View of the Family: Evaluative Interview

In several of the German-speaking cantons (e.g., Lucerne, Soleure) and in some having a Franco-German culture (Berne, Fribourg, and Valais), the tendency is to substitute written information for oral information through individual discussions (in certain cases, as a complement to the discussions).

In these, the teachers and the family appraise the progress of the child and the quality of the work that is in the portfolio and that represents current competence. These discussions help the

pupils to evaluate themselves, the families to choose the best measures to be taken, either educationally by internal differentiation, or structurally by external differentiation, in order to build up the best programme for the child.

These acts of advice and guidance are as much a professional responsibility of the teacher as is the teaching itself, and one of the tasks included in the specification of the teacher's contract. It is the responsibility of the teacher to guide a pupil and the family, to stimulate them in the development of a project of formation that is both ambitious and realistic. The teachers claim this role in the name of their professional status. The teaching task becomes richer by guiding the pupil with learning and by helping him or her to choose itineraries of formation. The joint estimation made together, with other colleagues in the case of secondary school teaching, but above all with the pupil and the parents, leads to decisions that give the pupil the best prospects.

Illustration

The informative interview in the German-speaking part of Fribourg:[5]

> An interview that takes place between the child, his parents and the teacher is the most important source of evaluative and promotional exchange.
>
> At the time of these interviews, some information will be exchanged concerning the help, progress and educational objectives that refer to the sphere of the child's personal, social and cognitive competence. The parents will be able to compare the child's own evaluation and the one drawn up by the teacher (auto- and hetero-evaluation). These interviews with the parents, provoked by the school, will take place at the end of the first half year.
>
> The principal partners of the learning process are the teacher and the pupil, and, because of this, with a periodic evaluation interview, two aims are attained: on the one hand, an exhaustive and differentiated presentation of the results, attitudes and development of

[5]Excerpt and adaptation of a text by Andreas Dick, pedagogical counsellor, translated by J. B. Thévoz.

the child, and on the other hand, the optimal opportunity of a profitable collaboration between school and family, in view of the child's global development.

The evaluation interview must take place at least once a year for each child, but may be repeated in order to follow better his situation. In theory, the interview lasts 45 minutes, in the classroom at the pupil's desk, and may develop in the following manner: the child tells and shows his parents how he works and learns; the teacher talks to the child about his contacts with his comrades and the parents listen to how he collaborates with them and to what purpose; the teacher and the child evaluate the results of his learning and the difficulties he has met with; the parents look at his exercise books, note any information concerning their child, can express their feelings and make observations; it is together that the teacher, parents, and child find solutions for a better support of the process of learning in the future and fix new objectives.

Evaluation with Summative and Institutional Aims

Switzerland, because of its federal structure, has no national summative tradition of evaluation and still less of national certification. The curriculum framework for the *Maturité générale* exists only since 1994. The *Maturité* (a recognised national title corresponding to a general baccalaureat) is the responsibility of the cantons, or even of the establishment, on the basis of a "contract of trust" between the evaluators and the Swiss Confederation.

At the present moment, the feasibility of a national certification at the end of compulsory schooling (end of secondary school I, 15–16 years) is still under consideration. Unanimity is far from being attained and no decision has yet been made.

National Reforms in Progress

The major reforms undertaken at present in the educational field are clearly situated in the domains of professional formation at tertiary level. The aim is an evaluation of qualifications by the creation of a professional Maturité and by creating specialised high schools (professional universities) in the fields of

technology, management, communication, music, plastic arts and education. The establishing of these new routes to formation has involved Switzerland in an important reflection on the structures, curricula, and evaluation procedures of education.

UNITED STATES OF AMERICA

DAVID STEVENSON, Department of Education, Washington, D.C.

The United States makes a substantial investment in the use of formal tests and assessments in its educational system. The purposes of these assessments include admission to private schools and post-secondary institutions, placement in programs designed to serve special needs, program evaluation, certification, and the monitoring of the systems at local, state, and national levels. But it is usually agreed that far more assessment is informally used in classrooms on a daily basis than all these formal tests put together.

The decentralised educational system in the United States is comprised of a complex of local schools, local school boards with considerable autonomy, and state departments of education. Because localities differ markedly in their practices, summing up a national picture must rely more on impressions and occasional empirical studies than on a documentable sampling of practices in these diverse settings.

Formative Assessments for the Improvement of Learning

While great efforts are made in the assessment of the educational progress of children, through teacher-made measures and informal judgments, there is little agreement on the quality of the processes used. In part, the disagreement on quality is based on unreconciled beliefs about the extent to which a rational model—involving planning, teaching, testing, adjusting, and retesting—is at all suited to the regular life of classrooms. Nonetheless, there is undeniable pressure on many teachers to reach external standards—whether these are set by numbers of students admitted to prestigious universities, or by the performance in annual state assessments.

Regular assessments of student progress are stimulated in many ways. Local educational systems may require regular reports on progress to parents, which may be in the form of

informal notes, or letter grades, and may include marks for the achievement in subject matter and in student effort. Some reports take a more narrative form and describe the nature of students' accomplishments and remaining challenges to be met. The audience for such reports is the parents and students, and the documentation in support of the grades is drawn, for the most part, from reviews of student work, performance on teacher-developed tests, and observations of classroom behavior. There is concern here that there is considerable grade inflation resulting from the use of marks for student motivation rather than for competence. The comparability of grades from teacher to teacher or from school to school is expected to be low. Such concerns undermine credibility of school grades and give rise to public pressures for "objective" external measures.

Classroom practices for assessing student progress vary widely. Many teachers rely on the use of assessments or tests embedded in the curriculum. A majority of texts and other instructional materials provide exercises, tests, and worksheets for students to complete. These assessments may often be accompanied by prescriptions for additional work that depend upon student proficiency levels. For the most part, neither the tests and assessments, nor the prescriptions connected to them, have been subjected to quality controls of any sort.

In the late 1970s, considerable effort was undertaken in the area of written composition to assist teachers in judging the quality of their students' work. For example, journals were kept by students to document their progress. In most states. universities continue to support such projects in order to create a cadre of specialists in schools who can help their colleagues to assess student work more systematically. These efforts laid the foundation for subsequent developments.

About ten years later, the movement towards project-based teaching and more integrated performance assessment was revived in the United States. Various strategies have been promulgated to assist teachers in helping students develop their proficiencies in complex, demanding tasks involving deeper knowledge of subject matter. One of the most popular of these approaches is the use of portfolios. These may provide a developmental record of student learning and involve periodic teacher-student review and reflection on student accomplishments. Other approaches involve creating projects or extended tasks that teachers judge as evidence of progress.

Teachers learn or create scoring criteria and review samples of students' work in order to build up a scale of reasonable expectations for student products. For the most part, such approaches have not been subjected to rigorous analysis or empirical study. Where data are available, they reveal great variability in the interpretation of judgment criteria used by teachers. In case studies where systematic effort was made to encourage development of teacher assessment practices, progress of teachers was fitful and costly.

Teacher professional development efforts are being mounted nationally by many commercial or quasi-commercial providers. The recent emphasis on making more explicit the curriculum goals (in content and performance standards) has supported the focus of these providers. Many agree that open-ended measures, rather than formal tests (particularly, multiple-choice tests), should be used regularly as measures of progress. Other providers seek greater congruence of interim progress measures with outcomes of external examinations, notably the annual state assessments. Increasing such congruence could strengthen the "alignment" or convergence of the system.

In the pre-service teacher preparation, project-based teaching and relevant assessment models are gaining some purchase. But for the most part, university coursework still focuses predominantly on abstract concepts of validity and reliability rather than considering the needs for teachers to become more skilled developers and critics of assessments. Another common concern of teachers is the absence of effective help for their task of managing the knowledge and time requirements to conduct regular project-based assessments. It is partly for this reason that the use of projects involving collaboration of small groups of students is supported.

Two technology-based avenues are being explored. One involves the provision of easier formats for data collection, management, analysis, and interpretation. Another strategy depends upon the use of unobtrusive measures of student progress in the course of engagement with interactive multimedia programs. For these systems to work, either embedded in a given instructional system or used as stand-alone measures, automated real-time scoring of complex student work is needed. Progress is being made on these fronts and could supply teachers with a valid basis for making more elaborate judgments of student work.

Summative Assessment

Summative assessment for the certification of students occurs in public education at a minimum competency level in some locations. Other programs have created formal systems that award special diplomas to students succeeding at particularly challenging examinations. A few states are exploring the use of a certification of mastery for all students, which will depend upon the completion of multiple tasks, some standardized and some original to the student. None of these programs is fully operational as yet. Other states are focusing on the use of certification in special areas, such as workforce preparation. These efforts will encounter technical difficulties similar to those met with portfolios. A persistent difficulty is the standard-setting procedure used to classify students' achievement.

For the most part, no fixed levels of performance are required for public school students to progress from elementary to secondary schools. The most prominent use of measures for admission occurs in the transition from secondary school to higher education, and from higher education to graduate study. Commercial tests are used nationally, are very well known, and are designed to be general measures of aptitude and achievement rather than of an explicit curriculum. The press and the public continue to infer that these measures give reasonable indication of school effectiveness, so that the scores generate intense public debate. There are efforts to use student portfolios to supplement or replace standardized measures for admission. The technical acceptability of the measures will no doubt be assessed in court cases brought by unsuccessful candidates.

The most widespread use of summative assessment is for monitoring schools. Most states have developed or are in the process of creating annual assessments matched to their standards and curricula. The use of results of these measures varies from public reporting of state averages or school rank, to the assignment of systematic sanctions and rewards for achievement. The manner in which adequate progress is determined, related once again to the standard-setting process, remains a technical concern. The measures themselves may be purchased from commercial providers, developed under contract, or created in part by the educational community in the states. In the early 1990s, many of the measures tried out to assess new curricular frameworks called for complex, open-

ended performance rather than for multiple-choice formats. In certain states, there has been a reconsideration of the wisdom of using these performance-based approaches for a number of reasons. These include the technical difficulty in obtaining reliable individual student scores and the limited content that may be sampled by any assessment. Nonetheless, many states continue to explore the use of these assessments alone on a sampling basis, while others explicitly augment them with traditional standardized tests that yield individual scores for each student.

The use of suites of disconnected measures assembled to meet the disparate demands of public beliefs and technical quality signals a new technical requirement: that systems of information, rather than individual measures, need attention to assure that the validity of inferences derived from multiple measures is warranted.

Systemic Reform - National Policies

Three major initiatives affect recent reform efforts at the national level. The first is the development of standards and models of assessment by national subject matter professional groups. These standards have no official status but there is political consensus on their value and they are being used in part or as a unit by states and local school districts. Test and text publishers have also revised their offerings to the public to make them more compatible with these standards. Standards-based reform and standards-based assessment are realities in many school districts. Some schools and states are taking a decentralised approach, encouraging a bottom-up generation of standards by teachers.

A second major initiative involves the federal government's effort to support these state initiatives. Legislation enacted has supported states and professional groups in the development of standards and assessments, by offering financial support to encourage the process. In 1995, the United States consolidated its education programs for economically disadvantaged children. In a major shift, the evaluation standards for this program (affecting more than 70 percent of schools) were changed. Now, determining program success is primarily a state function, and assessments given to students in the program should be consistent with the standards in the state expected for all

children. This provision was intended to dismantle any mechanism allowing disadvantaged students to be held to lower standards than other students. A corollary of this decision to hold students to the same standards is an expanded requirement for inclusion in testing. Students who do not perform as well as others are not exempt from assessments because of particular learning disorders or of low proficiency in English language. The resulting requirement for test developers to develop comparable measures for all students presents both technical and financial challenges. The degree to which these requirements will be met will be tested during the next four years.

A third initiative involves the use of regularly collected information about the national status of American student achievement. The major program for this monitoring purpose is the National Assessment of Educational Progress (NAEP). NAEP is undergoing change as it attempts to adapt its measures to make them consistent with state curricular reforms while maintaining enough existing measures to study trends across time. A second innovation in NAEP is the use of state-by-state reporting for comparative purposes. Findings from such reports have great influence on the policy debate in states. A third innovation has been the attempt to use continuing international comparisons to calibrate American students' performance.

Summary

The fabric of United States investment in assessment is both strong and multi-patterned. The dual threads of United States assessment are technical quality and fairness, but concerns for public credibility of measures are also essential.

APPENDIX C

DELEGATES

Chairman of the Conference

PROFESSOR RICHARD DAUGHERTY
Head of the Department of Education
University of Wales

Rapporteur Generals

MR ALAIN MICHEL
Inspecteur Général de l'Education Nationale
France

PROFESSOR PAUL BLACK
King's College
School of Education
England

OECD Project Expert

MR JOHN TOWNSHEND
Her Majesty's Inspector of Schools
England

Speakers (in order of appearance)

MR DAVID THOMAS
Counsellor
Centre for Educational Research and Innovation
OECD

PROFESSOR BARRY MCGAW
Director
Australian Council for Educational Research

PROFESSOR MARGARET BROWN
King's College
School of Education
England

DR EVA L BAKER
Center for the Study of Evaluation
CRESST
University of California, Los Angeles
United States

MR PETER OWEN
Director General of Schools
Department of Education and Employment
United Kingdom

MS CLAUDINE PERETTI
Deputy Director
Direction de l'Evaluation et de la Prospective
France

MR HANS VAN AALST
Consultant
(formerly Director of the Scientific Bureau of the ARO, Netherlands)
Centre for Educational Research and Innovation
OECD

DR ALEJANDRO TIANA FERRER
Director
Instituto Nacional de Calidad y Evaluation (INCE)
Spain

MR DAVID STEVENSON
Senior Advisor to the Under Secretary
Department of Education
United States

Delegates from OECD Member Countries

AUSTRALIA
DR PETER WHITNEY
Assistant Secretary
Quality Schooling Branch

AUSTRIA
PROFESSOR KARL HEINZ GRUBER
University of Vienna

BELGIUM
MR KLAUS HAUPT
Commission Europeenne
Direction Generale XXII/A/1
Education—Formation et Jeunesse

MR PETER MICHIELSENS
Inspecteur Général S.O.
Ministère de l'Education

MS VALERIE PIRON-COLLIGNON
MERF/Inspection Enseignement Fondamental

CANADA
MR NORMAN MAYER
Coordinator, Assessment and Evaluation Unit
Education and Training - Winnipeg

CZECH REPUBLIC
MR PETR DRABEK
Director - School Inspection

DENMARK
MR JORGEN BALLING RASMUSSEN
Ministry of Education

FINLAND
MR MARTTI APAJALAHTI
National Board of Education

FRANCE
MR GERARD BONNET
Ministère de l'Education Nationale
Direction de l'Evaluation et de la Prospective

GERMANY
MR PETER MUNK
Bundesministerium für Bildung,
Wissenschaft, Forschung und Technologie

GREECE
MR JOSEPH SOLOMON
Chairman, Department of Educational Evaluation
Pedagogic Institute, Ministry of Education

IRELAND
MR CARL O DALAIGH
Deputy Chief Inspector

ITALY
MS CHIARA CASTELLETTI CROCE
Inspectrice du Ministère de l'Instruction Publique Direction
Échanges Culturelles

MS GABRIELLA PAVAN DE GREGORIO
Researcher, Evaluation Projects
European Center of Education

LUXEMBOURG
MR DOMINIQUE PORTANTE
Directeur, Ministère de l'Education Nationale et de la
Foremation Professionnelle

MALAYSIA
MS NIK FAIZAH BINTI MUSTAPHA
Ministry of Education

NETHERLANDS
MR MELIS MELISSEN
Ministry of Education, Culture and Science

MR PIETER SCHEURS
Media Services Co-ordinator
Katholiek Pedagogisch Centrum - KPC

NEW ZEALAND
MR DAVID PHILIPS
Ministry of Education

NORWAY
MR OLA MOE
Director
Central Government Education Office

PORTUGAL
MS MARIA DO CEU ROLDAO
Ministerio da Educacio
Instituto de Imovacao Educacional

SWITZERLAND
MR JACQUES WEISS
Chef du Service de la Recherche
Institut Romand de Recherches et de Documentation
Pédagogiques (IRDP)

MR URS VOGELI-MANTOVANI
Centre de Coordination pour la Recherche en Education

MR MATTHIS BEHRENS
Institut Suisse de Pédagogie ISPFP

UNITED KINGDOM

Office for Standards in Education
MRS DIANE SIMMONDS
Her Majesty's Inspector of Schools

MR JOHN HAMER
Her Majesty's Inspector of Schools

MR CLIFF GOULD
Her Majesty's Inspector of Schools

MS SUE HANDS
Her Majesty's Inspector Of Schools

MR NIGEL WARWICK
Her Majesty's Inspector Of Schools

MR SANTOSH SADANAND
Executive Officer, External Relations Team

The Department for Education and Employment
MR MICHAEL STARK
Head of the School Effectiveness Division
Schools Funding Branch

MR SIMON JAMES
Head of the School Intervention Team
Schools Effectiveness Division

The School Curriculum and Assessment
DR NICHOLAS TATE
Chief Executive

MR DAVID HAWKER
Assistant Chief Executive

MR JOHN SINGH
Chief Educational Consultant British Council
(formerly Her Majesty's Inspector of Schools)

MR ALAN LOUNDES
Her Majesty's Inspector of Schools Wales

MR IAN J BOYES
Her Majesty's Chief Inspector of Schools Scotland

DR MARY JAMES
Tutor in Education
Cambridge Institute of Education

PROFESSOR CAROLINE GIPPS
Dean of Research
Institute of Education, University of London