

# EXPERIMENTAL DECISION-MAKING AND RESPONSIVE ACCOUNTABILITY

## THE NEED FOR A COMPREHENSIVE AND DIFFERENTIATED EVALUATION OF SCHOOLING

prepared by Hans Brügelmann, Konstanz/ Germany,

for O.E.C.D, Paris<sup>1</sup>

### Contents

- I. The Issue of Accountability: Its Focus and Scope
- II. Accountability and Evaluation: Broadening the Perspective
  - (A) The danger of Accountability Overload
  - (B) The danger of Accountability Freeze
- III. Sources and Concerns of Contemporary Accountability Requests
  - (A) Discontent with Democratic Control and the Public Services in General
  - (B) Discontent with Education and Recent Reform Attempts in Particular
  - (C) The Controversy about Goals and Principles of Education
  - (D) The Competition of Educational Theories and Programmes
  - (E) The Cost-effectiveness of Educational Investments
- IV. Different Styles of Evaluation: Five Archetypes
- V. The Role and Ethics of Evaluation: Towards Checks and Balances in Educational Evaluation
- VI. Activities and Methods of Evaluation: Beyond the Methodical Schism in Educational Evaluation
  - (A) Three Services of Evaluation
  - (B) Comprehensive Designs and Complementary Methods
- VII. Audiences of and Participants in Evaluation:  
Strengthening the Sensitivity to Criteria of Decision-making and Fostering the Capacity of Self-Evaluation
- VIII. Matching Evaluation Approaches to Accountability Concerns:  
On the Potential of Different Modes of Evaluation
  - (A) Accountability through Parliament and Court
  - (B) Accountability through Examinations and Testing Programmes
  - (C) Accountability through Inspectors and Peer-review
  - (D) Accountability through Cost-benefit Analysis and Public Statistics
- IX. Conclusions and Recommendations

---

<sup>1</sup> Slightly edited by the author in April, 2006

# TOWARDS RESPONSIVE DECISION-MAKING AND ACCOUNTABILITY

## IN EVALUATION:

### THE NEED FOR A DIFFERENTIATED APPROACH

### TO EDUCATIONAL EVALUATION

#### SUMMARY

#### I. THE ISSUE OF ACCOUNTABILITY: ITS FOCUS AND SCOPE<sup>2</sup>

Policies and practices of basic education have come under attack in OECD member countries. Requests for formal accountability of the system and its members are topical. They influence both the political role and methodical style of evaluation. State- or nation-wide testing programmes, the introduction of new budgeting and management systems, action-research networks, parliamentary debates and special committees as well as increasing community participation in the governance of schools indicate divergent trends. The notion 'accountability' is not widely accepted, however, and the meaning of the concept differs widely from one country and context to another. Yet, it is functional rather than linguistic or institutional similarities that matter.

In this paper, *accountability* is seen as the social obligation to justify decisions. It has to be interpreted in relation to both specific responsibilities or tasks and defined patterns of control. Evaluation, on the other hand, provides a repertoire of roles and methods for collecting, organizing, assessing and reporting evidence.

Basically, accountability presupposes:

- (i) a clear responsibility of an actor or agency to perform a certain activity or to achieve certain goals;
- (ii) the proof of a causal link between his (its) actions and the (missing) effects;

---

<sup>2</sup> paras. 1-9 in the full version of this report

(iii) the lack of a superseding justification for his (its) action on other grounds;

(iv) subjective fault of action.

In education, however, success or failure cannot easily be established and attributed to specific actions or conditions: criteria are controversial and their specification can often be disputed; effects are difficult to ascertain; multi-causality is the rule.

Accountability is no one-way relationship. Responsibilities and the resources needed for meeting such obligations have to be negotiated between different parties both within and outside the System. Thus, four types of accountability have to be distinguished. The first two represent different levels of *external* accountability, the latter ones different modes of *internal* accountability:

(i) political ('public') accountability of the system to the community or society at large;

(ii) individual ('client') accountability of agencies or persons in the system to those served by the system;

(iii) administrative ('managerial') accountability within the education system;

(iv) collegial ('professional') accountability within the education community.

There is no one evaluation and control mechanism that can cope with the different demands emerging in the four relationships. *Evaluation* provides a repertoire of methods for collecting, organizing, assessing and reporting experience needed for answering specific accountability requests adequately. Evaluation does and should, however, serve a broader range of purposes than defined by contemporary accountability requests.

## **II. ACCOUNTABILITY AND EVALUATION: BROADENING THE PERSPECTIVE<sup>3</sup>**

### ***(A) THE DANGER OF ACCOUNTABILITY OVERLOAD***

At present, attempts to make the education system as a whole accountable to

---

<sup>3</sup> paras. 10-17 in the full version of this report

outside agencies or groups are discussed more extensively in education than the problem of internal or client accountability. The relationship between increasing public accountability and attempts to ignore internal monitoring mechanisms has not sufficiently been explored. Sometimes, the impact of internal, esp, hierarchical, accountability demands is moderated by external pressures, e. g. at the local or school level. Central accountability demands, on the other hand, tend to be passed on internally to the lower strata of the hierarchy. Increasing pressures on the justification of decisions and conflicting expectations will lead to accountability *overload*.

The inclusion of more and more groups or agencies in educational decision-making and control will impair the daily work of the school if new responsibilities are simply added rather than redistributed within the system as a whole. A differentiation of accountability related to specific tasks and levels of accountability is necessary. Internal and external accountability have to be balanced and matched to the respective decision-making responsibilities.

Educational practitioners need some freedom to act responsibly and to take into account the particularities of their respective situation. Experimental thinking and progressive experience in problematic areas of education presupposes that teacher autonomy be protected. On the other hand, being left alone when handling the complexities of the classroom will end in routine and bias. Accountability mechanisms are needed to stimulate experimental thinking and the exploration of problematic areas of schooling. It should be avoided, however, to push teachers into defensive operations of self-protection.

### *(B) THE DANGER OF ACCOUNTABILITY FREEZE*

Different levels of evaluation correspond to specific accountability demands. It should be recognized, though, that the four *levels* below have different functional values in the decision-making structure of more localized vs. more centralized systems:

(i) Investigations at the (*inter*)*national* level can contribute to a picture of the educational system and its development over time (cf. 'public accountability' in para. 7);

(ii) studies at the *regional* or *local* level more specifically investigate relationships between education within the schools and certain context factors feeding, for instance, decisions on the allocation of resources (generally 'managerial', sometimes 'public

accountability');

(iii) evaluation at the *school* or *classroom* level can provide feed-back on the day-to-day practice to the people involved (generally 'managerial', 'professional' or 'public accountability');

(iv) *individual* assessment will serve diagnostic purposes, the certification of achievements or selection procedures ( mainly 'client accountability').

At these four levels, the different types of accountability merge with other functions of evaluation. Four such functions have to be distinguished:

(i) the justification and criticism of past actions ( i. e, accountability as usually understood );

(ii) the accreditation of agencies and individual achievements;

(iii) formative feed-back for improving on-going programmes and practice; (iv) specific predictions leading to a better match of options to given conditions.

Schools would suffer from an accountability *freeze* if public attention and resources became concentrated on summative or retrospective evaluation at the expense of a careful analysis of the causes of certain effects and substantial experimentation with alternative causes of action. Establishing a precise picture of the status quo is useful, only, if it can contribute to better practice in the future. On the other hand, different evaluation tasks cannot be combined without discretion. Often divergent expectations of evaluation participants and audiences severely impair the implementation of a study and the utility of its results. Criteria are needed for the selection of adequate evaluation procedures under defined circumstances (cf. section VI below) and for checking the compatibility of different tasks (cf. section VIII below).

### **III. SOURCES AND CONCERNS OF CONTEMPORARY ACCOUNTABILITY REQUESTS<sup>4</sup>**

---

<sup>4</sup> paras. 18-38 in the full version of this report

Accountability requests in compulsory education stem from quite different sources and express a broad range of distinct concerns. They presuppose different criteria and corresponding methods of investigation.

#### (A) DISCONTENT WITH DEMOCRATIC CONTROL AND THE PUBLIC SERVICES IN GENERAL

Present criticism of schooling can to some extent be explained as a result of widespread dissatisfaction with the effectiveness of the public services and the alleged "legitimation crisis" of liberal democracy. The problem of internal management and control has put an additional burden on the administration. Governments have started to improve existing accountability mechanisms making evaluation "a multimillion dollar industry".

#### (B) DISCONTENT WITH EDUCATION AND RECENT REFORM ATTEMPTS IN PARTICULAR

More specifically, education is traditionally seen as the cause of mis-developments in society at large. The highly visible reform attempts of schooling in the 1960s and early 1970s have disappointed excited expectations of its followers and increased distrust in other quarters of society. On the other hand, the innovation culture itself has propagated rigorous evaluation as intrinsic element of educational planning. It has also provided the man-power and the technology necessary. Thus, both the need for and the response to accountability were stimulated from within the educational reform business.

#### (C) THE CONTROVERSY ABOUT GOALS AND PRINCIPLES OF EDUCATION

The first focus of accountability is on the justification of goals and principles. Controversies on basic assumptions (such as open education vs, programmed instruction), over the role of public schools vs. family education and over the more specific goals of curricula in sensitive areas ( social education, sex education, new maths, etc.) have increased public awareness of divergent values. Cultural fragmentation ( e. g, by immigration, or differentiation of social groups ) and the alleged irrelevance of the school curriculum to both community life and the tasks in the world of work have sharpened goals conflicts. Increasing dissension about the direction and the basic principles of education eventually endanger the (relative) professional freedom of teachers. Evaluation, on the other hand, has to cope with the new task of probing into

the legitimacy of goals; it is asked to contribute to normative decisions about the values and basic assumptions of education.

#### (D) THE COMPETITION OF EDUCATIONAL THEORIES AND PROGRAMMES

The range of educational philosophies, programmes and methodical options has broadened considerably. Educational traditions have lost the appeal of well-established practice. Doubts about the effectiveness of new programmes are raised. Thus, accountability requests concern the goal fidelity, the internal consistency and situational adequacy of educational proposals as well as their impact and effects in practice. Social experiments have gained currency as a response of evaluation to these questions; criterion-referenced testing has been used for monitoring the school system as a whole.

#### (E) THE COST-EFFECTIVENESS OF EDUCATIONAL INVESTMENTS

The efficiency of schooling is doubted. Educational 'output per cost unit' is declining: While costs increase to keep up with salary levels, for instance, it cannot be demonstrated that the yield is increasing commensurably. At the same time public budgets have become tight. Education has to compete with other social services for scarce funds and is not seen as public priority no. 1 any more. Finally, demographic changes are terminating the period of expansion. Evaluators, consequently, have to calculate more precisely the benefits of existing or proposed arrangements against the costs on the input side.

Different evaluation approaches are needed for answering each of the four accountability requests adequately. They are discussed in detail in section VIII .

Secondly, accountability requests have to be answered by different types of information and support from evaluation:

- (i) sometimes people simply need to *know* what a certain programme, activity, etc. looks like in reality; thus, a rich and precise *description* is called for;
- (ii) in other cases people want to *understand* the relationship between different events and conditions, e. g. the functioning of a programme; here, a powerful *explanation* sensitive to the requests of decision-making is demanded;

(iii) finally, people may have difficulties in *assessing* the respective advantages and shortcomings of different programmes, etc.; in this case, a fair and comprehensive *judgement* is needed.

Before discussing these different evaluation activities (cf. section VI (A.)), it is necessary to relate the issue of accountability to a more comprehensive framework of evaluation traditions and approaches.

#### IV. DIFFERENT STYLES OF EVALUATION FIVE ARCHETYPES<sup>5</sup>

Evaluation is part of everyday life. People respond, for example, to existing accountability requests (many of which are rather informal). Accountability mechanisms increasingly draw, however, on more specialized evaluation traditions. These traditions support the present trends towards a *differentiation of roles and a specialization of methods*. Division of labour can help to fulfill specific tasks more effectively and to relieve the workload of the general practitioner in the classroom or administration. On the other hand, translation problems are created: the sensitivity of investigations to everyday problems and perceptions, and the utility of findings for actual decisions may decrease. Consequently, it is important to explore the respective pay-offs and risks of different approaches. Basically, the role of the evaluator and his methodology can be conceptualized within five traditions characteristic of well-established sub-cultures of society (cf. Fig. 1 for an overview):

- (i) following the legal-political tradition of educational decision-making a *hearing model* of evaluation can be developed in accordance with contentious jurisdiction by adversary proceedings;
- (ii) experimental practice and the critical discussion of experiences by professional colleagues would provide a *research model* of evaluation following the tradition of social science as rational argument based on empirical evidence;
- (iii) *personal reviews* by experienced specialists or recognized citizens yields an evaluation based on expertise or personal authority as in art criticism;
- (iv) *standardized measurement* by precision Instruments draws on technical methods

---

<sup>5</sup> paras. 39-47 in the full version of this report

for ascertaining certain effects matching an industry model of educational planning;

- (i) finally, *speculation* based on a commitment to basic beliefs (corresponding to an ideology model of educational planning) still guides much of everyday evaluation.

It is suggested that evaluation designs and approaches

- (i) have to be internally consistent by either adhering to one model or by securing the compatibility of elements combined from different traditions;
- (ii) have to match the educational philosophies, curriculum models, and instructional or innovation strategies to which they are applied;
- (ii) have to match the task at hand and the institutional setting within which they are to be implemented;
- (iv) have to be examined for the specific advantages and shortcomings, respectively, that are connected with the traditions they are drawn from;
- (v) have to be analyzed more deeply if conflicts emerge in the course of a study, i.e. constructive responses to such difficulties can be found if surface frictions are related to more fundamental tensions of philosophy or institutional setting.

Evaluation cannot become effective for accountability when there is a mismatch between the role-concept of the evaluator and the methodical tradition within which he works, on the one hand, and the values of participants as well as the specific expectations of both sponsors and audiences. Thus, it is important to explore in more detail three major aspects of evaluation:

- (i) the role and ethics of evaluation (cf. section V),
- (ii) the activities and methodical options of evaluation (cf. section VI) and
- (iii) the audiences of evaluation, their needs and criteria (cf. section VII).

Two preliminary conclusions can already be drawn:

- (i) The research model and the legal-administrative model of evaluation

match easily the two types of internal accountability described in para. 7. However, the combination of these two approaches, e.g. in the role of an inspectorate, will yield considerable problems as they root in different traditions (cf. section VIII.(C)).

(ii) No comparable correspondence between modes of external accountability and evaluation traditions exists. Accountability both at the system and at the individual level still rely largely on the ideology model. The introduction of new elements, e.g. from the industry or the art model, often suffers from a mismatch between evaluation approach and audience expectations.

## V. THE ROLE AND ETHICS OF EVALUATION<sup>6</sup>

Accountability is defined by *power-relationships*. Evaluation, thus, is delicately related to demands for control and the disposal of sanctions as shown in paras. 10-12, accountability presupposes that the people concerned have a chance to argue the case in question. In education goals and responsibilities rarely are as clear as in the production sector of society. Means-end relationships are not easily defined. The simple analogy to tort in civil law sketched out in para. 10 reveals the difficulties in establishing appropriate accountability mechanisms. Of course, the decision-making and control pattern in a system largely determines who is *responsible to whom, for what* (cf. for the difficulties of a comparative analysis the chapter on Control and Management, however). On the other hand, accountability procedures should not just reinforce the historically grown distribution of power. Accountability and evaluation should rather help to strengthen those voices that at present are not adequately represented and to improve the system of checks and balances often impaired in practice by power accumulating in certain agencies or groups. Finally, evaluation has to protect the interests of those under study and to secure the fairness of the procedure by organizational means.

Basically, three different *political stances* of evaluation can be distinguished:

(i) Evaluation can be organized *bureaucratically*, i.e. as an unconditioned service to powerful agencies or decision-makers. The evaluator keeps detached from the events and people he is studying and acts as an external control. His image in the field often is that of the 'centre spy' creating reservations at the grass-roots level. The work of

---

<sup>6</sup> paras. 48-53 in the full version of this report

inspectorial systems (cf. section VIII.C. below) is often impaired by such suspicions: self-protection becomes more important than learning from external criticism and advice.

(ii) *Autocratic* evaluation assumes independence both from the sponsor and the people/ agencies studied. The evaluator acts as an expert in his own right. Objectivity is the main principle guiding his work and the scientific community his court of appeal. The negative image often associated with this approach is the 'academic voyeur' satisfying his own information needs, only. Educational experiments often suffer denial of co-operation from people and agencies protecting themselves against 'data exploitation'.

(iii) *Democratic* evaluation attempts to represent a broad range of interests, perceptions and judgements recognizing value pluralism as an intrinsic feature of public education. The risk of multiple negotiations with both sponsors and participants lies in the delicacy of the checks and balances to be maintained by the evaluator. In effect, his activities may be absorbed by political management rather than research; or he may be seen as 'partisan' by some groups losing credibility in the eyes of outsiders.

There is no escape from the political consequences of evaluation, especially in the context of accountability. Information implies power. Evaluation has to counteract a wide-spread tendency of sponsors (including internal auditors) to reserve the right to know to themselves. Exclusive feed-back to one party is neither acceptable in a democracy nor feasible in the practice of accountability. Both the credibility of evaluation and the validity of its findings depend on the cooperation of all parties concerned.

Recent attempts of professional organizations in the U.S. to develop a code of 'Evaluation Standards' may provide some safe-guard against the mis-use of evaluation. They will not become effective, however, if state authorities rely on research bodies set up within an administrative hierarchy or when they exploit their strong position on the market for defining the conditions of contracts with outside agencies unilaterally. Evaluation would then become reduced to an instrument of control 'thereby losing much of its potential.

## VI. ACTIVITIES AND METHODS OF EVALUATION<sup>7</sup>

---

<sup>7</sup> paras. 54-70 in the full version of this report

## A. THREE SERVICES OF EVALUATION

Accountability rests on standards or norms as the yardstick for assessing performance. Evaluation, consequently, is a normative activity: It perceives and portrays educational events from a particular point of view; it interprets causes and effects within one theoretical framework rather than another; it has to appeal to shared values for the acceptance of its judgements. All evaluation is value-laden.

Nevertheless, three quite distinct activities have to be distinguished that are related to different accountability tasks:

(i) Succinct and vivid *descriptions* of educational programmes and events can help outsiders to get access to unique and often dispersed experiences. This service of evaluation is particularly important at a time when non-professionals, i.e. parents or members of parliament complain about not understanding the school and its curriculum any more. Evaluation as 'vicarious experience' often has to draw on investigative traditions other than experimental psychology and sociological surveyism: history, journalism, book-keeping, anthropology, documentary film and other disciplines have become prominent in illuminating the salient features of a programme or activity, its outcomes and the setting within which it is located from different perspectives.

(ii) Powerful *explanations* sensitive to the options and conditions of decisions-making in the field can help to detect deficiencies and to take remedial action. Accountability cannot stop with labelling the status quo. Evaluation has to reveal the causes of unsatisfactory performance and to detect alternative courses of action.

The thinking underlying experimental design has considerably influenced this type of evaluation during the past 15 years and, on the whole, strengthened its practice. In recent years, however, a broader framework of validity and credibility has allowed for complementing this specific strand of thinking with concepts and procedures from other disciplines such as law, for instance, and for relating evaluation more closely to everyday interpretations of the problem under study.

(iii) Evaluation as a specialized activity is often restricted to tasks (i) and (ii); decision-makers reserve the assessment of evidence to themselves. The more complicated the issues become, however, and the more the evaluator can claim expertise in the area under study (rather than merely methodical competence) the more often sponsors and audiences are inclined to ask for and to rely upon the evaluator's *judgement* when

taking a decision. Moreover, accountability with its often conflicting expectations sometimes leads to the attribution of an arbiter role to the evaluator.

This raises, firstly, the issue of standards and of appropriate reporting (cf. section VIII). Judgements have to be comprehensive, fair and sensitive to the options open to decision-makers. Secondly, some control of the evaluator is needed which here cannot be achieved by methodological rules, but has to be secured by appropriate social mechanisms (cf. the legal model discussed in section IV and VIII.A).

### *B. COMPREHENSIVE DESIGNS AND COMPLEMENTARY METHODS*

The range of accountability tasks and corresponding evaluation activities puts heavy demands on the methodical competence of evaluators. To cope with these demands evaluation has to exploit the full range of methods developed both within the experimental tradition of psychological testing and social survey science as well as within the so-called interpretive traditions of history, ethnography and law. The *methodical schism* in evaluation has to be overcome. A more differentiated and comprehensive understanding should guide its practice.

Three *dimensions* of evaluation design are of particular importance:

(i) Detachment vs. involvement, i.e. evaluation designs can have built into them either a strong research or a strong action component.

(ii) Standardization vs. responsiveness, i.e. evaluation can be highly naturalistic (favouring external validity) or highly controlled (strengthening internal validity).

(iii) Focus vs. scope, i.e. evaluation studies have to decide between a larger number of cases to be investigated (automatically limiting the number of variables that can be taken into account) and a broader range of aspects and their relationships (investigating them over a limited number of settings).

There is no *a priori* complementarity of the respective poles of the three dimensions. The adequate combination within a design will depend on the goal of the study, the resources available and the circumstances under which it has to be conducted.

No one methodical tradition can cope with all information demands. No method can claim superior validity on principle. Contemporary accountability demands should

stimulate the development of more comprehensive designs and inspire the adaptation of methods from other disciplines. New questions and information demands should be accepted by evaluators as a challenge to their imagination rather than as an excuse for falling back on less controversial methods and inadequate instruments.

## VII. AUDIENCES OF AND PARTICIPANTS IN EVALUATION<sup>8</sup>

### STRENGTHENING THE SENSITIVITY TO CRITERIA OF DECISION-MAKING AND FOSTERING THE CAPACITY OF SELF-EVALUATION

Accountability is dependent on the understanding of the parties involved and on their acceptance of the evidence supplied. Educational decision-making is influenced by four main reference groups with *divergent information interests and criteria* for assessing educational programmes and actions:

(i) Administrators and policy-makers are interested in a comparison of alternatives at hand and look for generalizations based on hard empirical data.

(ii) Educational researchers and developers judge proposals in the light of general principles and favour absolute judgements based on theoretical analysis.

(iii) Practitioners tend to pragmatic judgements based on practical experience and related to the peculiarities of their own setting.

(iv) Students and their parents favour the satisfaction of individual needs and judge education in terms of biographical experience and changing sub-cultural trends. The diversity of standards ( and of the evidence required for judgement ) supports the proposal to differentiate the system of accountability accordingly. Consequently, the focus and the style of evaluation has to take these audience differences into account and to respond to their different value orientations, frame conditions and communication styles. Generally, the evaluation capacity in the field should be complemented rather than replaced by specialist evaluation. Formal expertise has to draw on personal experience anyway. Moreover, it cannot become effective for decision-making unless it responds to the sub-cultural traditions of both programme participants and evaluation audiences. Abilities of and opportunities for practitioners

---

<sup>8</sup> paras. 71-84 in the full version of this report

to monitor their own activities should be enhanced by specialist evaluation and supported by administrative measures. Role differentiation between 'diagnosis' and 'therapy' should be minimized at all levels of the system.

### VIII. MATCHING EVALUATION APPROACHES TO ACCOUNTABILITY CONCERNS<sup>9</sup>

Four major *criteria* have to be related to specific concerns of accountability mentioned:

- (i) the *legitimacy* of goals and principles of compulsory schooling
- (ii) the internal *consistency* of programmes and their theoretical assumptions as well as their goal fidelity
- (iii) the situational *adequacy* of programmes in practice and their effectiveness under divergent circumstances
- (iv) the *economical* use of scarce resources defined by a comparative cost-benefit ratio.

#### (A) EVALUATION BY PARLIAMENT AND COURT

External accountability starts with two issues referring to the *legitimacy* of education:

- (i) Are the goals and principles of compulsory schooling consonant with more principal aims and conceptions of society? (= public accountability at the system or community level)
- (ii) Are the goals and principles of certain activities of compulsory schooling responsive to or, at least, not dissonant with the rights of the individuals concerned? (= client accountability at the individual level).

Accountability for goals and principles of action presupposes, at least in a democratic society, a framework for public debate within which values and assumptions can be

---

<sup>9</sup> paras. 85-154 in the full version of this report

discussed controversially and which feeds a formalized decision procedure. Mechanisms of social control rather than methodical precision is exploited as organizing principle securing the transparency of evaluation.

Three different responses to this demand can be found in OECD member countries:

- litigation before courts
- parliamentary control of basic decisions
- governing or consultative bodies at the community and school level allowing for a broader participation of those concerned.

The dissatisfaction with compulsory schooling has, thus, led to attempts to re-balance

- (i) central and local control of education,
- (ii) administrative and parliamentary and finally
- (iii) professional and political control of education.

The trends in different countries may look dissimilar because of historical differences in the point of departure as well as another ranking of values in the political culture. Yet, most of them show an increasing significance of the political-legal mode of accountability.

Parliamentary and legal control will generally increase pressures on the central administration. Such accountability requests tend to be passed internally to the lower strata of the hierarchy. This can lead to a *legalization* of schooling which often counteracts its own purposes (i. e. making public education more transparent and less arbitrary). Strict adherence to formal rules rather than open discussion of trial and error can be the result. Teachers will tend to make their decisions "court-proof".

Parliamentary control of compulsory schooling is necessary. Education cannot be exempt from supervision by representatives of the people. Such control should be restricted, however, to basic decisions, i. e. the definition of the central aims and principles of schooling, its overall organization and the general criteria for the distribution of funds. Discretion has to be given to local bodies for the allocation of resources to specific purposes, for the specification of aims and principles according to the needs and restrictions within the local community, the school and class, respectively.

Control by the courts has to be restricted to violations of the boundaries defined by central policy. Courts should not neglect the fact that educational encounters are irrecoverable and unique in a significant sense. They have to respect the professional

responsibility of teachers. Other types of accountability are requested for an appropriate response here. Moreover, it should be borne in mind that a uni-lateral formalization of accountability may be counteracted by the re-vitalization of less formal mechanisms of evaluation and control.

### *(B) EVALUATION BY EXAMINATIONS AND TESTING PROGRAMMES*

*Effectiveness* is the second criterion for accountability. Do the effects match the aspirations? will be the leading question?

A short-term measure of effectiveness are examinations and other forms of assessment. Success in learning can be defined in three ways:

(i) one can compare the achievement of a person *relatively* to the achievement of others, i.e. with reference to *group* norms;

(ii) one can establish standards in terms of goals with defined achievement levels and ascertain the degree to which a person has reached the goal, i.e. with reference to *set criteria*;

(iii) finally, one can *record* the *progress* of an individual from one point in time to a later date, i. e. define success in *individual* terms.

All three kinds of information are needed to get a valid picture of individual (or group) achievement.

A careful analysis of information requests leads to two conclusions:

(i) Present examinations - usually *either* norm- *or* criterion referenced - cannot alone provide the information needed for a fair account and constructive criticism of educational activities. Comparative data in particular have to be carefully interpreted in their respective contexts. The limited predictive value of examination marks in the long-term perspective has to be recognized.

(ii) Examinations themselves would benefit from a combination of the three types of information. They then could fulfil more adequately their different functions.

Four purposes of assessment have to be distinguished:

- (i) to determine whether something that has been taught has been learned; this will help the teacher to review his teaching and to meet external accountability requests;
- (ii) to identify the strengths and weaknesses of individual children; this should enable the teacher to take remedial action and to select specific tasks that match the experience and readiness of children with difficulties;
- (iii) to provide information about pupils at points of transition in their school career and when leaving school; such information will feed selection procedures and also lead to certification of qualifications;
- (iv) to motivate pupils through competitive incentives or - more recently - by involving them in a process of self-assessment.

While in former times functions (iii) and (iv) have been dominant, today functions (i) and (ii) are seen as increasingly important. The *diagnostic* function of *assessment* is linked to new guidance Services re-shaping the role of the teacher.

More responsive modes of diagnosis are developed. They are based on informal observation closely associated in its assumptions with the open education movement and the re-appraisal of cognitivism in psychology. It is characteristic of such projects, and the approaches they propose, that they have built into them a strong component of teacher co-operation and in-service education. The assumption is that instruction cannot be programmed from outside the classroom, but has to rely heavily on the judgement of individual teachers. Teachers, then, need two kinds of support:

- (i) more flexible materials developed within a coherent framework of educational thinking;
- (ii) help to develop their own thinking and repertoire through critical examination of their own practice.

An opposite development is the increasing standardization of final examinations connected with the testing movement.

This trend is particularly strong in the Anglo-Saxon countries. Large-scale *testing programmes* with different functions have developed:

(i) National or state-wide testing programmes based on light sampling of pupils aim at achievement profiles over different areas of the curriculum for the whole country or rather broad sub-categories ( e.g. boys vs. girls, rural vs. urban schools ). They also are intended to compare the level of achievement over time. With the aid of matrix sampling they can cover a broad range of competencies and sub-skills. Such programmes focus on

- checking folklore evidence on the effectiveness of compulsory schooling;
- finding out areas of schooling where improvement or additional support is needed.

(ii) Other programmes focus on a smaller number of tests ( usually in the basic skills area ) and are applied to the whole age group. They aim at smaller geographical or institutional units and primarily serve

- to complement less standardized forms of pupil examination;
- to focus educational activities on key goals.

At present only few results are available from these programmes. No consistent trends emerge. The utility and methodical adequacy of large-scale testing is highly controversial. It is difficult to judge their worth generally as they differ in important respects. Some problems are of general significance.

The first issue is the *selection and weighting* of goals and the definition of significant levels of achievement. The decisions about central competencies should be linked more closely to curriculum planning and to political decisions by democratic bodies. Frictions between the curriculum and the testing system have to be avoided. Central tests should not override regional, local or school control over the curriculum in decentralized systems.

Secondly, it is difficult to *operationalize* goals and to *measure* the effects of schooling. The more complex the outcomes of schooling are the less susceptible they are to standardized testing. Testing should not favour particular styles of curriculum planning and instruction (based on behavioural objectives) over more process-oriented styles of teaching and long-term goals of education. In particular, the testing of basic competencies has to be complemented by evidence covering a broader range of possible outcomes and side-effects. This presupposes 'softer' evaluation approaches that have to be included into an overall assessment. More recent testing programmes already attempt to increase their sensitivity. Such internal improvement has its limits, however. Blue-ribbon Panels, reports from advisors and surveys on judgements of clients and their partners in society have to complement testing.

Testing programmes cannot take into account the *conditions* under and the *processes* through which certain aspirations have (not) been achieved. A more ecological approach to evaluation is asked for. The interplay between different factors leading to certain accomplishments have to be investigated. This would increase the fairness of accountability and enable decision-makers to take remedial action.

Finally, no one method or research tradition (such as psychological testing) can cover the whole range of information requests or claim an overriding validity of its instruments. Survey studies of few selected variables over a large number of pupils and contexts have to be complemented by more *intensive case-studies* of few selected institutions covering a large number of variables and yielding more comprehensive 'functional patterns' of schooling under certain circumstances. This will also help to link accountability at the system level to school or classroom accountability fostering teacher development and institutional self-study.

### (C) EVALUATION BY INSPECTORS AND PEER-REVIEW

*Consistency* is a much neglected criterion in accountability. It is, however, an essential one. The question *why* something has gone wrong and *where* changes are necessary can be answered only if one looks at the relationship between different elements of a system or activity. Consistency has to be achieved between

- means and ends of a programme or activity,
- theoretical assumptions and the structure of a program,
- programme rhetoric and practical activities.

Three approaches can be utilized for this type of accountability:

(i) The corresponding evaluation can rely on the public authority of recognized citizens or experts. It then is based on the personal experience of the observers accumulated over a long period of practice and on their credibility as independent guardians over the public good.

(ii) The second approach is based on experimental practice and the critical discussion of experience within the professional community. It is based on special training and shared values and aims at further development of teachers' abilities and the tradition within which they work.

(iii) The third approach is modelled after managerial accountability in large industrial companies or administrative accountability in state bureaucracies. This approach is based on the centralization of decision-making power and a hierarchical organization of information channels.

Inspectorial systems are an attempt to integrate a professional model of evaluation into a framework of legal-administrative control. Its flexibility is the particular strength of this assessment strategy. There are, however, serious tensions between the legal and the professional tradition of evaluation. A balance between the diagnostic feed-back and the authoritative control component of evaluation can be maintained only by differentiating them into different roles and procedures.

Such external assessment of the hierarchical type has to be complemented by (a) internal self-assessment and (b) external peer assessment strengthened by a community control component. Projects of the first type have become influential in parts of the U.K. It has been argued, though, that professional self-evaluation cannot cope with the increasing credibility gap between the lay public and self-contained professions. It is here that accreditation procedures derived from the tertiary Sector come into play. They have the advantage of being externally, but not hierarchically organized. They should be based on broad professional expertise and moderated by community involvement.

#### *(D) EVALUATION BY STATISTICS AND COST-BENEFIT ANALYSIS*

The *efficiency* of public schooling is one of the main concerns behind the present accountability debate (cf. section III.(E) above). It corresponds to evaluation being conceived of as industrial product testing and cost-benefit analysis. The main criteria for accountability in this sense are utility of schooling for well-defined purposes and productivity of given resources. Such evaluation is a necessary counter-part of the other modes mentioned so far.

It has, however, been strongly resisted by educationists for a long time. It is true that this approach has come up against tremendous difficulties in practice: The effects of education are not easily specified in advance; outcomes and input cannot be measured as precisely as the production rate of a factory. Nevertheless, the broader idea behind cost-

benefit analysis demands wider application. More adequately interpreted, costs do not only imply financial expenditure, but also

- the emotional side-effects of a programmed reading scheme,
- the motivational exhaustion of teachers swept by new curricula,
- absenteeism rates in compulsory schooling,
- parents' anxieties about a *numerus clausus* in certain university subjects.

Such side-effects tend to be under-valued or simply overlooked by evaluators. The alleged pay-offs of innovations in particular have to be related to the money, time, energy, goodwill, etc. demanded from numerous people in the system. On the other hand, latent costs of 'well-established practice' are often taken for granted.

It has to be acknowledged that this area of evaluation still is largely unexplored or explicitly avoided by educational evaluation. Attitudinal as well as methodical changes are necessary:

(i) educators have to accept that their work partly is determined by the resources available and that it is necessary to account for their parsimonious use;

(ii) the public and its representatives, on the other hand, have to recognize that economic models and methods can represent only to a limited degree what is important in education;

(iii) evaluators, finally, have to become more imaginative in developing approximate measures for variables deemed important in schooling by educators and their clients.

Moreover, evaluation itself has to become more cost-effective. Data available in official statistics and other documents (e. g. school records) have to be utilized. Schools already complain about over-testing and questionnaire overload. Accountability will further increase this burden if care is not taken to economize the business of data collection.

## **IX. CONCLUSIONS AND RECOMMENDATIONS<sup>10</sup>**

Since the beginning of this century compulsory education in OECD member countries has made progress. It has potential for further improvement. Good practice has to be identified and consolidated. Problematic areas have to be carefully explored. *An*

---

<sup>10</sup> paras. 155-165 in the full version of this report

*experimental approach* to teaching, and *educational decision-making* in general is needed. There is no normative consensus nor an accepted theory on which everyday practice could be safely based. Programmes cannot be pre-ordained in detail.

They are more like hypotheses to be tested out under different circumstances and to be further developed in the light of accumulating experience.

Accountability can and should stimulate the critical examination of present practices and alternative courses of action. This presupposes:

- (i) a sufficient range of options open to practitioners and some freedom for becoming experimental;
- (ii) appropriate support for doing so in a competent way and incentives for accepting this additional burden;
- (iii) an accountability system responsive to the particularities of classrooms and the full range of issues and criteria necessary for a fair assessment of schooling.

The documents on which this paper has drawn provide ample evidence that there exists *no one model* of accountability for different education systems and their respective sub-systems. Nor is there one best method of evaluation to answer different accountability demands. Nevertheless it can be suggested that

- (i) *accountability mechanisms* have to become both
  - (a) more differentiated in *responding to the four audiences* and the corresponding responsibilities defined in section I as well as
  - (b) more comprehensive in judging programmes or activities with *respect to the four concerns and criteria* of assessment defined in section III;
- (ii) *evaluation procedures* correspondingly have to become both

- (a) more comprehensive in *exploiting the full range of investigative traditions* as defined in sections IV and VI as well as
- (b) more differentiated in *matching roles and methods* adopted to the specific accountability task at hand{cf. section VIII).

*Four particular risks* of accountability and evaluation in the present situation have been discussed in detail:

(i) Increasing pressures for justification from different parties within and outside the education system can lead to *accountability overload*. This generally results in self-protective measures of those in charge of education at all levels. Openness presupposes some freedom protected by a balanced system of responsibilities and mutual checks. Changes of established practice depend on the participation of those concerned in the evaluation and decision-making process (cf. section II.A).

(ii) Increasing pressures for an on-going justification of past actions can lead to *accountability freeze*. The education system and its members will exhaust themselves in defending what they have done instead of improving what they are doing and intend to do. Accountability mechanisms have to leave room and resources for pro-active types of evaluation ( cf. section II.B ).

(iii) Increasing specialization of evaluation can be the result of increasing demands for accountability. The resulting division of labour would widen the already existing *gap between diagnosis and therapy*. Responsibilities and competence become separated between persons and agencies creating considerable translation problems. Specialist evaluation should therefore aim at developing the evaluation capacity of decision-makers rather than replace it ( cf. section VII ).

(iv) Increasing *use of evaluation for control purposes will impair its potential for supporting everyday practice*. A more widely accepted ethics for evaluation practice and institutional support for securing its independence is needed ( cf. section V ).

The elements for a more responsive accountability system and the respective strengths and shortcomings of different evaluation approaches have been analyzed in the preceding sections. It is not possible to determine one best combination of elements for all countries. *Accountability mechanisms have to match existing patterns of decision-making in the*

*respective system and to respond to the prevailing philosophy of education and the values of its political culture.* The argument put forward in this paper should help to find a better match. Nevertheless, further research is needed to explore the pay-offs and risks of different approaches and their functioning in different context in particular. The influence of more prescriptive (*ex ante* programming ) vs. more evaluative (*ex post* correction ) approaches deserves particular attention ( e. g. the combination of a soft or a strict curriculum and a soft or strict examination system and implications of the emerging patterns for accountability ).

Since no *one* system or method can be recommended, it is necessary to define some *general criteria that can help to develop a more balanced and responsive accountability system for a particular context.* The following list, developed by John Nisbet (1978) in the British context, provides a useful framework: Accountability

- must operate in a way that is *fair* to all concerned;
- should be *valid* and *relevant* to current concerns;
- should provide *feedback* for decision-making and encourage wider *involvement* in decisions;
- should either be *objective* or make subjectivity *explicit*;
- should be *verifiable*, i.e. open to checking;
- should *not distort* the processes of teaching and learning;
- should be *understandable* and the results *communicable*;
- should be *comprehensive* and take account of the wide variety of aspects of education.

The application of these criteria to contemporary trends would lead educationists to fight for a more balanced accountability system and to *avert*, in particular, a *preponderance*

- of central over *local* control of education;
- bureaucratic over democratic ethics of evaluation;

- administrative and public over *professional* and *client* accountability;
- legal and industrial modes of evaluation over *research* and *artisan* modes;
- effectiveness and efficiency over *legitimacy* and *consistency*;
- justification and accreditation over *diagnosis* and *guidance*;
- judgements over *descriptions* and *analyses*.

Educationists cannot escape accountability. They should concentrate on shaping it in a way that everyday practice can benefit from it. Such attempts should be supported by evaluators developing more adequate methods. If, however, the public and its representatives insist on simplistic measures for the success of schooling, defensiveness rather than improvement will be the answer.

## References

This report is mainly based on the national statements of OECD member countries and some special reports that are fully referenced in its long version.

Becher, A. L./ Maclure, S. (eds.) (1978): *Accountability in education*. Social Science Research Council. National Foundation of Educational Research: London.

Nisbet, J. (1978): *Procedures for Assessment*. In: Becher/ Maclure (1978, 95-112).