

# PROSPECTS

Quartely review of  
comparative education

Issue number one hundred and five  
105

The evaluation of education systems:  
the view at the end of the 1990s

Vol.XXVIII, n°1, march 1998

ISSUE NUMBER ONE HUNDRED AND FIVE

# PROSPECTS

quarterly review of comparative education  
Vol. XXVIII, no. 1, March 1998

Editorial

Juan Carlos Tedesco 3

## VIEWPOINTS/CONTROVERSIES

Education for all: access to learning  
and achieving usable knowledge

Robert Glaser 7

## OPEN FILE: THE EVALUATION OF EDUCATION SYSTEMS: THE VIEW AT THE END OF THE 1990s

Introduction to the open file *Alejandro Tiana Ferrer* 23

Student assessment and learning for a changing society *Caroline Gipps* 31

The potential of international comparative studies to monitor  
the quality of education *Tjeerd Plomp* 45

The OECD educational indicators: purposes, limits and  
production processes *Norberto Bottani* 61

Dialogue evaluation: a possible contribution of evaluation  
to school improvement *David Nevo* 77

Value-added measures of school effectiveness  
in the United Kingdom *Sally Thomas* 91

Institutional arrangements for evaluation *Ernest R. House* 109

How can evaluation contribute to educational policy?  
The uses of information in Australia *Barry McGaw* 117

## TRENDS/CASES

Assessing teacher effectiveness in India:  
overview and critical appraisal *J.S. Rajput and K. Walia* 137

## PROFILES OF FAMOUS EDUCATORS

Florence Nightingale (1820–1910) *Alex Attewell* 153

UNESCO  
3474  
-9 JUL 98  
LIBRARY / BIBLIOTHEQUE

Authors are responsible for the choice and presentation of the facts contained in this publication and for the opinions expressed therein, which are not necessarily those of UNESCO:IBE and do not commit the Organization. The designations employed and the presentation of the material in *Prospects* do not imply the expression of any opinion whatsoever on the part of UNESCO:IBE concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Please address all editorial correspondence to:  
Editor, *Prospects*, International Bureau of  
Education, P.O. Box 199, 1211 Geneva 20, Switzerland.

To find out more about the International Bureau of  
Education, its programmes, activities and publications, see  
the IBE's home page on Internet: <http://www.unicc.org/ibe>

All correspondence concerning subscriptions should be  
addressed to: Jean De Lannoy, Avenue du Roi 202,  
1190 Brussels, Belgium. E-mail: [jean.de.lannoy@infoboard.be](mailto:jean.de.lannoy@infoboard.be)  
(See order form at the end of this volume.)

Published in 1998 by the United Nations Educational,  
Scientific and Cultural Organization, 7, place de Fontenoy,  
75352 Paris 07 SP, France.

Printed by SADAG, Bellegarde, France.

ISSN: 0033-1538

© UNESCO 1998

---

# EDITORIAL

---

During the last decade, numerous education systems have begun to design and introduce tools to measure the outcomes of learning. The origin of this trend is to be found in the State's reform processes and, more precisely, in the need to confront the problem of low levels of responsibility for the outcomes (i.e. accountability) typical of education systems, particularly in the public sector. The assumption on which the setting up of measurement systems is based is that a more accurate and clear knowledge of the outcomes of learning is an important factor in strategies intended to improve the quality of education.

It is necessary to take note of this process in order not to lose sight of the ultimate objective of systems for evaluating and measuring the outcomes. In the final analysis, the presence of these systems is not an end in itself and is only justified if they form part of the process of improving the quality of education. To use the metaphor of John Rawls, evaluation systems should lift 'the curtain of ignorance' which has affected traditional education systems and which should then allow educational activity to be directed according to the specific characteristics of each person, each group and each social and cultural situation. In other words, systems for evaluating outcomes make sense when they allow strategies for action to be put forward, both institutional and personal, that are intended to build on the accomplishments already achieved.

On this point, the experience of recent years has been extremely fruitful. Nevertheless, new problems are constantly arising. It is no longer sufficient to measure learning achievements in cognitive terms; performance in the work market and in society together have led to growing demands to measure achievements in the other areas of personal development as well. As far as institutional aspects are concerned, the experience of these years has highlighted the enormous importance that questions concerning the role of the government and non-State bodies play in the administration of tests and in the dissemination of the results. Confidence in the bodies that produce information, their transparency and legitimacy represent basic factors for the success of these systems. Information for the general public, and not only for specialists and administrators, has also been shown to be an issue of enormous importance about which we have only recently become aware.

In summary, what we measure, who measures it and how the information is disseminated represent three key questions in discussions about systems for evaluating the



quality of education. In order to sum up the present state of discussions, UNESCO's International Bureau of Education and the Brazilian Ministry of Education organized an international seminar in December 1997 in the city of Rio de Janeiro. The most important outcomes appear in this issue of *Prospects*, which—like the seminar itself—has been co-ordinated by Mr Alejandro Tiana, to whom we must express our immense gratitude for his most capable work. We are confident that this publication, accompanied by a more substantial one produced by the Brazilian authorities, will serve as a stimulus to all those involved with this theme.

\* \* \*

With this issue of *Prospects* I complete my mission as editor-in-chief. My experience of these years in editing *Prospects* has been most enriching and I hope that the readers of the review will agree that the three principles featuring in the editorial of the first issue of this new series have been satisfactorily fulfilled: intellectual rigour; respect for diversity; and the ability to adapt to the profound changes taking place in modern society and education. I am certain that *Prospects* will continue on this road, on which we will meet again.

JUAN CARLOS TEDESCO

---

# VIEWPOINTS/CONTROVERSIES

---

# EDUCATION FOR ALL: ACCESS TO LEARNING AND ACHIEVING USABLE KNOWLEDGE

---

*Robert Glaser*

---

## **Introduction**

We all know that people are educated in accordance with the prevailing cultural and political conditions. However, the conditions of older times can remain in effect, become maladaptive, and deter progress even as a culture attempts to change. As we work to evaluate and improve our education systems, we need to look at our goals and accomplishments in light of modern knowledge; we must identify the constraints and barriers that hold us back and that make change difficult.

With this theme of change in mind, I will consider two fundamental aspects of an education system: the first is how access to education is determined; the second is how learning is evaluated. These two issues, access and accomplishment, are reflected in how students are selected to enter school, how their abilities to learn are facilitated, and how their quality of learning, competency and educational goals are assessed.

---

*Original language: English*

*Robert Glaser (United States of America)*

Ph.D. from Indiana University in psychological measurement and learning theory. Founder of the Learning Research and Development Center at the University of Pittsburgh. His current work focuses on the application of cognitive theory to the assessment of competence and the design of education based on modern knowledge of learning. Former president of the American Educational Research Association and the National Academy of Education. Charter member of the European Association for Research on Learning and Instruction. Honorary degrees from the universities of Leuven, Goteborg and Trondheim. E-mail: glaser@vms.cis.pitt.edu

The first issue, access to education, has been influenced by social practices that attempt to identify those who appear to have the ability, aptitude or social class to profit from formal schooling. Access to instruction is managed through tests, as well as through methods of classification and placement.<sup>1</sup> As we work to improve our education systems, we must ask ourselves whether the methods of selection and classification that have developed function well and if they are acceptable for a society that aspires to maximize the capabilities and achievements of most of its citizens, almost all of whom are a concern of the education system.

The second function, assessing achievement and outcomes of learning, has evolved into a system of standardized achievement tests for many countries. The results of a student's performance is referenced to norms which give comparative information, but which may not recognize necessary standards of achievement. What is frequently evaluated are subjects convenient for established testing formats and forms of instruction that rely on older theories of learning, where remembering information is predominant and where processes of thinking and problem solving—the effective use of knowledge—are less evident. For progress and improvement, the results of teaching and student achievement need to be assessed against standards of active mentality and on criteria of competent performance rather than only group norms.

## **Access to education**

Practices of selection and placement in schooling can have mixed effects. As a result of these procedures, some individuals are included into further education with significant benefit to themselves and to society. Others are excluded and may be placed in less advanced or watered-down instruction that does little to enhance their educational opportunities. On the surface, the process can appear to select the best talent; it can, however, be based on beliefs of aptitude and intelligence that set in motion a cycle of undeveloped ability. Too many students can be excluded from certain experiences presumably because they lack the attributes and background that promise successful learning; we must be aware that these very same attributes could be learned in appropriate educational environments. Too often, individuals are being excluded from the very learning experiences that could develop the capabilities they need to be included into effective education.

The background experiences of low-achieving students frequently do not expose them to the information and modes of cognizing that are useful in school learning; yet these skills can be learned if students participate in environments where such knowledge and abilities are exercised. Evaluation tests and other criteria used to make readiness and retention decisions may identify a disproportionate number of disadvantaged children as not ready for regular schooling. As a result, they are placed in lean curricula that are not likely to promote learning skills. Pulling these students from mainstream instruction often excludes them from fruitful educational experiences and from using the knowledge and ability they have to improve their educational attainment (Glaser & Silver, 1994; Oakes, 1985, 1990; Oakes & Lipton, 1990; Shepard, 1991).

In order to provide access to educational opportunity and to lessen the exclusionary aspects of student evaluation, future practices must be designed to survey possibilities for student growth. Assessment and instruction must be more integrated and based on standards of performance. This enables teachers to recognize and support children's strengths, and to provide environments that develop their abilities to learn. Low-achieving students must be assigned to forms of education that do not curtail their opportunities to learn or to develop learning skills.

Clearly, testing for selection and placement is an important practice at certain times in the development of a society, but over-reliance on it can come close to writing a person's biography in advance, determining both life and career. This deterministic philosophy needs to be softened for it is no longer the only option. We need to ask how access to education can best contribute to the growth and potential of students. The requirement (and the research-and-development problem for the future) is to design an educational society that helps, that can be 'judged in terms of its ability to facilitate constructive adaptations of educational programmes for individuals' (Thorndike, 1975) where usable and socially recognized knowledge is achieved by all students and high levels of competence are attained by many. The goal is to increase access to educational opportunity in ways that maximize both a person's and the community's well-being.

#### IMPROVING ACCESS TO EDUCATIONAL OPPORTUNITY

Let me now turn to several areas where the use and development of learning abilities can be considered for the purpose of facilitating adaptive learning environments. School education can build on the following three strengths: (1) community-based knowledge; (2) abilities for self-regulation; and (3) beliefs about effort and fixed ability.

#### COMMUNITY-BASED KNOWLEDGE

Education that is adaptive to the student can take account of the high levels of performance that result from the demands of problem solving and reasoning that face young people in everyday life. Outside of formal schooling, individuals develop competence in solving verbal and quantitative problems through community activities. This can also occur with the need for early employment or apprenticeship to a trade (Lave, 1988; Rogoff, 1990). Participation in these social practices is a fundamental form of learning and its relevance to the design of school pedagogy should be considered. This available knowledge can be related to school knowledge so that individuals become confident about their own learning abilities. Schools need to identify the strategies that have been learned by students in their everyday life. These strategies can then become the basis for future learning. The forms of competence and expertise that many students have acquired are sophisticated in their settings of use outside the school building. While these skills may be less immediately applicable to the objectives of formal schooling, they do provide

knowledge and skills that can become a basis for effective education and training (Carragher, Carragher & Schliemann, 1985; Greeno, 1991; Rogoff & Lave, 1984; Saxe, 1990). Co-operative skills used in working with others can be applied to school relationships for working with peers and teachers for success in school (Sternberg, et al., 1995).

#### ABILITIES FOR SELF-REGULATION

Learning and instruction are influenced by the ability of individuals to monitor and regulate their own actions. As human development and learning have come to be viewed as the result of constructive activity, as well as memory, there has been greater awareness of the self-regulatory abilities that children and adults use to perform a task or to solve a problem. These regulatory activities enable students to monitor and exercise control over their performance as they carry out their work (Brown, 1978). These abilities include control strategies such as: predicting outcomes of one's efforts, planning ahead, appropriately apportioning one's time, explaining to one's self in order to improve learning and understanding, taking notice of failures to comprehend, and calling on background knowledge relevant to the situation. Learners can use such monitoring skills, as they are required, in the course of carrying out a new task or attempting to comprehend a situation.

Good learners often use such skills, but many individuals need to be taught to exercise these capabilities. So, these regulatory and monitoring skills become important for assessing the ability to succeed in school and are important practices that need to be taught in school to increase ability to learn.

#### BELIEFS ABOUT EFFORT AND FIXED ABILITY

In certain societies (for example, the United States), there is a strong view that aptitude or fixed ability predetermines progress and achievement. School administrators, teachers and the students themselves believe that they have been born with a certain level of ability that determines the nature of the goals they should strive for. In school systems associated with this belief, students and teachers act in accordance with alternate programmes which are designed to be most suitable to getting the best performance for a student with a certain level of ability (Dweck, 1988; Dweck & Leggett, 1988; Resnick, 1995; Resnick & Nelson-LeGall, in press).

In contrast to this, comparative studies within nations and societies have described different patterns of belief in aptitude and ability, resulting in different school practices (Holloway, 1988; Peak, 1993; Stevenson & Lee, 1990). To improve aptitude for learning, certain school environments place a major emphasis on habits of learning and ways of interacting with others. Students believe that developing abilities to improve their learning is equally important to showing what has been learned. In short, learning goals are equally as important as accomplishment goals. The individual not only strives to show how well a particular subject matter has been learned, but to develop new ways of improving one's abilities to learn. In these

environments, students engage in learning behaviours such as testing their own understanding and being aware of standards for their performance.

Access to education might be improved by moving away from traditional systems that place too much value on fixed ability toward educational environments that emphasize the importance of expending effort to develop learning ability. Effort-based systems can motivate learning and restore a sense to young people such that they can indeed make positive, productive contributions to their own learning and to society.

Let me restate this general point. Many present systems of education and their modes of entry into school and the work force are designed primarily around the belief that talent and ability are largely inherited and fixed. As a result, educational practices are designed to select the talented for highly demanding curricula; a challenging curriculum is avoided for too many students, confirming their belief that they do not have the talent for advanced thinking. Students do not try to break through the barrier of low expectations because they, like their teachers and parents, accept the judgement that their ability is fixed and that they do not have the right level of aptitude. Not surprisingly, their performance remains low and the system becomes a self-sustaining one.

On the other hand, it is not necessary for education to continue in this way. Schooling and preparation for life and work can be built around the alternate assumptions that effort actually creates ability to learn; and that the development of learning strategies and expectations of effort and accomplishment can create ability. An environment that encourages the development of learning skills and increased personal effort will influence the nature of competence in individuals. This is not just a matter of being rewarded for working hard, but rather a result of linking efforts toward learning and standards of achievement in ways that are likely to improve not only specific knowledge, but abilities to learn effectively.

## Outcomes of education

I have considered only one side of the equation so far, access to education—that is, the strengths and abilities that children and youth bring to schooling. These involve competencies that can be used and developed to ensure more effective education. I turn now to the output side of the equation—that is the outcomes or results of schooling and learning.

In order to point to possibilities for progress here, I must again refer to the weight of current dysfunctional practice. For the assessment of educational outcomes, there has been an over-reliance on 'norm-referenced' measures of educational achievement where, as I said before, the effectiveness of education is judged in terms of an individual's position with respect to group performance (Glaser, 1994). The student is judged in terms of percentile rankings, as being in the twenty-fifth or seventieth percentile, or so many units above average. Indexes of this kind can provide information that a school, a community, or a state might be doing relatively better than others, but less information is provided about the actual nature of the

performance acquired and quality of the achievement attained. For improving education, it is necessary to judge the success of schooling and the performance of students in terms of what they actually know and can do, as well as in terms of standards of performance.

Furthermore, we can assess how well students read words, perform arithmetic operations and remember historical or literary events. But in modern times, this is no longer enough. More than ever, our work and society will require the ability to use this knowledge for thinking and reasoning, and to use what we know in the context of changing information. We must read in order to comprehend, solve mathematical procedures with understanding, and reason with historical and civic knowledge for effective living as citizens in developing societies. In general, the evaluation of an education must assess not only the content of what is learned, but also the ways in which this information is used—the forms of knowing that give knowledge intellectual and creative power.

We must assess the kinds of learning that contribute to individual competence and to the use of human mentality. As changes occur in a student's knowledge, we must observe the accompanying development of the cognitive processes involved in problem solving and thinking with what they know (Anderson, 1985). We must move beyond the testing of simple memory and assess the power of interconnected structures of knowledge that represent meaning, and that account for complex levels of explanation and understanding. Concepts of coherent and connected structures of information, and the nature of knowledge organized for reasoning, must drive today's procedures of assessing learning outcomes. Our older conceptions have all too often resulted in the testing of fragmented information using a multiple-choice format that does not encourage inference and thinking with what we know. Educators now face the challenge of developing and assessing connected and usable knowledge.

#### THE DEVELOPMENT OF COMPETENCE AS A BASIS FOR ASSESSMENT

We can look at studies that compare the performance of beginners and experts in various fields to understand what is meant by the development of organized, connected knowledge. These studies describe how knowledge and skill change as individuals learn and become more competent in various subject domains as they gain experience in their field of work (Chi, Glaser & Farr, 1988; Glaser, 1996). The changing characteristics of achievement, as individuals move from the status of novice to expert performer, provide a framework for assessing levels of achievement (and a focus for instruction) (Glaser & Chi, 1988; Lane & Glaser, 1994). Let me give five examples of these characteristics of developing proficiency:

##### *Integrated, coherent knowledge*

Effective learning and expertise develop meaningful connections in a student's mind. This integration and connectedness is highly related to a learned individual's ability to comprehend and solve problem situations, as compared to a beginner's more



fragmented memory for isolated pieces of information. Integrated knowledge enables proficient students to make inferences, to think by making analogies to what they already know, and to link meaningful explanations to their performance. Good instruction emphasizes organized knowledge.

### *The proceduralization and usability of knowledge*

Competent performance is related to procedures for the use of knowledge. As learning occurs, the declarative information and facts that we know are not just stored in memory, but are related to ways in which this information can be used. Effective assessment involves determining not only what is known, but how it is used. Although experts and novices may be equally competent at recalling a principle or rule, novices recognize less frequently where such knowledge applies or how to implement it. In typical assessments, there is often a disassociated nature to what we measure that does not include the sense of actively applying one's knowledge.

### *Problem recognition and representation*

As students become more competent in a subject, the way they look at problems and situations changes, and this change can be assessed. An often-cited example comes from a study of students learning elementary physics. When presented with problems of motion (mechanics), beginning students will classify a problem in terms of obvious surface features such as an inclined plane problem or a problem using a pulley. A more advanced student would report the basics of the underlying principles and say, 'Oh, that's a problem involving the Conservation of Energy or Newton's Laws.' Based on the way the problem is represented, by surface features or by underlying principles, the solution is carried out in more or less expert fashion (Chi, Feltovich & Glaser, 1981).

### *Managing information*

Assessing the way people manage information will increasingly be a major problem of our time. The amount of available and changing information will force our assessments of achievement to emphasize the utility of current learning that enables students to use their knowledge for future learning. Students need to attain knowledge and skills for handling large volumes of information; for example, they must learn to take multiple perspectives and to generate organizing concepts so that they can use what they know in new situations. In essence, a good part of a student's education must focus on generative abilities to update his or her knowledge. This is a difficult assessment problem, but will need to be tackled in the future.

### *Integration of basic skills with advanced performance*

Assessments should measure whether fundamental skills have attained an efficiency

that enables a higher level of co-ordinated performance to proceed. Assessments of this kind are very important in relating early reading skills to meaningful word and story comprehension, and in integrating basic quantitative operations with mathematical problem solving procedures. This form of assessment can also be carried out at later stages of education with basic procedural and conceptual aspects of advanced algebra and calculus that need to be efficiently involved in solving complex problems. (This may be less of a problem when the use of computers for fundamental operations are available; students may require less practice in basic skills; they will, however, need to be taught to understand them.)

General components of proficient performance, like the five examples I have just described, differentiate and contrast the learning outcomes of successful and unsuccessful learners in many content domains. Along with the content of learning, these components give significance to the ways in which knowledge is used for higher-level cognitive activity that determines the excellence of our students, the competence of our communities, and the expectations of students and parents.

#### STANDARDS FOR LEARNING AND ASSESSMENT

With this general aspiration for high-level achievement in mind, much work is proceeding on the concrete and explicit statement of standards for the practical work of teachers and students. In many countries, this is a government-sponsored enterprise assisted by various curriculum groups and teacher organizations concerned with performance levels in reading, language arts, mathematics, science and learning applied to work. These groups are publishing specific products that can be useful to educational reform.

In the United States, a New Standards Assessment System established by a consortium of professional organizations and schools (National Center on Education and the Economy, 1997) has considered three interrelated components for an assessment system. In various school subject matters, these components are: (a) Performance Content and Process Descriptions: these descriptions specify what students should know and give examples of ways they should demonstrate the knowledge and skills they have acquired; (b) Work Samples and Commentary: these are the samples of student work that illustrate standard-setting performances; they are accompanied by comments that show the relationship between standards and the sample of student performance; (c) Examinations and Student Portfolios: performance standards are backed up with examinations and by a portfolio system that complements the examination. Portfolios contain accumulating evidence of student performance and pieces of work over a period of time. The examinations and selected student work are discussed with reference to performance standard descriptions.

At the present time, many countries have excellent examples of standards and suggestions of assessment practices which can be used as a basis for educational improvement. The trick is in their implementation, which employs communities and State systems to pull together with students, parents, teachers, curriculum societies

and education ministries to develop enterprises that are carefully evaluated for the achievement of stated goals. System-wide co-operation, shared high-level expectations and evaluation are the keys to their success.

## **Schooling and communities for learning and instruction**

If we accept goals for education that encourage thinking and the active use of knowledge, then we need to consider innovations in school environments and classrooms that can help produce these outcomes. It is necessary to extend teaching and the nature of school environments in directions that attain high standards for most children and youth, and that assist many teachers in using professional skills for accomplishing these goals. We will need to try different forms of pedagogy to accomplish this task, and build on the wisdom of our best teachers with modern knowledge of learning and cognition. School learning, and educational experiences, both in and out of school, will need to be designed to achieve the characteristics of competent performance that enable higher-order abilities in the use of acquired knowledge (Bruer, 1993).

Changes in the nature of teaching should be considered and carefully evaluated. In particular, pedagogical research suggests that for many instructional situations there needs to be a shift in focus from the teacher as a purveyor and teller of knowledge to the learner as a constructor and designer of his or her own learning under curriculum guidance. We are encouraged to consider changes from passive learning to active student performance for acquiring competence.

For this purpose, educational settings have been designed in elementary schools and elsewhere to help students learn as a community. They work together to understand and organize information in a domain of knowledge as part of their curriculum. In the course of this activity, they develop the capability for joint intellectual endeavour as they learn with the help of the teacher and by teaching each other. Students learn in these groups more through asking and investigating than from being told. They are encouraged to investigate a scientific topic, the nature of a geometric proof, the properties of numbers, and the characteristics of different genres of writing and literature. As they investigate, they ask questions, test what they are learning, compare it with the teacher's or other expert's knowledge, and question the details of this knowledge.

As they work together, the learning is very intentional and focused. Students are intent on meeting standards; they acquire performances that they can display, and use acquired knowledge and skill to learn the next lesson. This intentional learning can be seen in student classroom exchanges where students will comment on what another student has said, and then test out their thinking by saying, 'Well, I believe this about the matter,' and come to a position of expressing what they know. In this type of educational setting, what becomes apparent is how competent people think in various disciplines. For example, students not only learn the facts of history or biology, but they also learn how a historian or biologist thinks and gathers evidence to come to conclusions and propose explanations.

In these settings of learning, facts and information become part of the data of a learning and thinking enterprise. Students can be connected to events in the world and in their local community, and use this information from outside events for what they are learning. For example, they can use their mathematics and science knowledge to study ecosystems, and can send this information (via the Internet) to a nearby community for further information. They can use their writing ability to explain to other students and their parents how they interpret what they are learning. In general, students, with teacher guidance, can consider their information and interpretations, or ask for evidence from elsewhere to support their ideas.

These educational settings have been called 'communities for knowledge building' (Brown & Campione, 1990, 1994; Scardamalia & Bereiter, 1991, 1996). Students participate in the transmission of knowledge by seeking, sharing and acquiring knowledge among themselves with gradually decreasing guidance. These knowledge-building communities are distinguished by efforts to turn over to students processes that are usually under the teacher's control. Students are helped to formulate goals, to direct their own inquiry, to monitor their understanding and to use the resources available to them to design their own settings for acquiring knowledge. In this participatory environment for learning, teachers and students share the expertise they have, or take responsibility for finding out about needed knowledge that they can bring back to the group. There is a community of discourse in which learning through constructive discussion, conjecture, questioning, criticism and presenting evidence is the normal thing to do rather than the exception (Brown & Campione, 1994; Campione, Brown & Jay, 1992).<sup>2</sup> An important instructional concept is that, as students progress to higher levels of knowledge and performance, they also become increasingly skilled as active agents in supporting their own learning and the learning of others. Knowledge-building communities as environments for learning are currently undergoing careful evaluation and study (Cognition and Technology Group at Vanderbilt, 1992, 1994).

## **The integration of assessment and instruction**

In the future, testing and teaching, and assessment and learning should become more integrally related events (Frederiksen & Collins, 1989). As learning occurs, evidence is available to assess accomplishment and to decide upon the next instructional steps. Characteristics of assessment that are integral to instruction will include at least six aspects:

### *Access to educational opportunity*

As I have emphasized, in order to lessen the exclusionary aspects of education systems, assessment should be designed to survey possibilities for student growth. Programmes of assessment and instruction should enable teachers to recognize and support learners' strengths so that they can achieve in more powerful curricula.

*The display of competence*

Knowledge and skills should be measured so that the processes and products of learning are openly displayed. The criteria of performance by which students are judged will be apparent and clear so that these criteria can motivate and direct learning and teaching.

*Self-assessment*

Assessment will involve the teaching of self-assessment. Because assessment and instruction are integrally related, instructional situations will provide coaching and practice in ways that help students reflect on their performance. Such occasions for assessment enable students to set incremental standards by which they can judge their own achievement and develop self-direction for attaining higher performance levels.

*Socially-situated assessment*

The conditions of assessment can require performance in a social setting in which students contribute to a task and assist others. This has the advantage of encouraging students to develop and question their definitions of competence; students can observe how others reason and receive feedback on their own efforts. In this context, not only can performance be assessed, but also the facility with which a student receives and adapts to help and guidance.

*Instructional effectiveness*

Assessments are judged in terms of their effectiveness in informing teachers about how much time and emphasis should be given to certain concepts and content, and to cognitive skills in the curriculum. Outcomes of assessment can be interpreted in terms of how they influence teaching and instruction, and in terms of the information they provide for the development of classroom activities that relate to learning goals.

*Cognitive significance*

Assessments provide content coverage but do not neglect significant processes of performance, such as raising questions, representing and planning a problem prior to solution, and offering conceptual explanations for solution procedures. Constructing assessment procedures necessitates analysing the cognitive aspects of an activity and designing related scoring procedures that offer evidence of the cognitive performance involved.

## Final comment

Assessments of the learning abilities and achievements of students must be designed and used in ways that take account of the goals of modern society and of present knowledge of human learning. Reform and continued improvement will be impossible if we continue to carry the weight of practices that were designed for times gone by. New perspectives are now offered on the nature of knowledge and abilities that are brought to learning, and on the nature of competent achievement that results from instruction. Innovative systems that integrate access to learning, instruction and assessment can now drive the design of educational environments that support and respect human cognitive ability, and prepare people for dignified lives, competent work and social growth.

## Notes

1. Psychometric theory grew up primarily in this context of use.
2. To facilitate a community of learning, computer systems have been designed through which children can interact for the purposes of building, exploring and sharing knowledge. Their activity centres on a student-generated community database in which students provide comments on one another's entry notes by raising questions, suggesting information sources and providing counter-arguments or encouragement (Scardamalia & Bereiter, 1991).

## References

- Anderson, J.R. 1985. *Cognitive psychology and its implications*. San Francisco, Freeman.
- Brown, A.L. 1978. Knowing when, where, and how to remember: a problem of metacognition. In: Glaser, R., ed. *Advances in instructional psychology*, vol. 1, p. 77-165. Hillsdale, NJ, Erlbaum.
- Brown, A.L.; Campione, J.C. 1990. Communities of learning and thinking, or a context by any other name. In: Kuhn, D., ed. *Contributions to human development: developmental perspective on teaching and learning thinking skills*, vol. 21, p. 108-126. Basel, Switzerland, Karger.
- ; —. 1994. Guided discovery in a community of learners. In: McGilly, K., ed. *Classroom lessons: integrating cognitive theory and classroom practice*, p. 229-270. Cambridge, MA, MIT Press/Bradford Books.
- Bruer, J. 1993. *Schools for thought*. Cambridge, MA, MIT Press.
- Campione, J.C.; Brown, A.L.; Jay, M. 1992. Computers in a community of learners. In: DeCorte, E.; Linn, M. C.; Mandl, H.; Verschaffel, L., eds. *Computer-based learning environments and problem solving*, vol. 84, p. 163-192. Berlin, Springer-Verlag.
- Carraher, T.N.; Carraher, D.W.; Schliemann, A.D. 1985. Mathematics in the street and in school. *British journal of developmental psychology* (Leicester, U.K.), vol. 3, p. 21-29.
- Chi, M.T.H.; Glaser, R.; Farr, M., eds. 1988. *The nature of expertise*. Hillsdale, NJ, Erlbaum.
- Chi, M.T.H.; Feltovich, P.; Glaser, R. 1981. Categorization and representation of physics problems by experts and novices. *Cognitive science* (Norwood, NJ), vol. 5, p. 121-152.
- Cognition and Technology Group at Vanderbilt. 1992. The Jasper experiment: an exploration

- of issues in learning and instructional design. *Educational technology research and development* (Englewood Cliffs, NJ), vol. 40, p. 65-80.
- Cognition and Technology Group at Vanderbilt. 1994. From visual word problems to learning communities: changing conceptions of cognitive research. In: McGilly, K., ed. *Classroom lessons: integrating cognitive theory and classroom practice*, p. 157-200. Cambridge, MA, MIT Press/Bradford Books.
- Dweck, C.W. 1988. Motivation. In: Lesgold, A.; Glaser, R., eds. *Foundations for a psychology of education*, p. 187-239. Hillsdale, NJ, Erlbaum.
- Dweck, C.S.; Leggett, E.L. 1988. A social-cognitive approach to motivation and personality. *Psychological review* (Washington, DC), vol. 95, p. 256-273.
- Frederiksen, J.R.; Collins, A. 1989. A system's approach to educational testing. *Educational researcher* (Washington, DC), vol. 18, no. 9, p. 27-32.
- Glaser, R. 1994. Criterion-referenced tests: part I origins; part II unfinished business. *Educational measurement: issues and practice* (Washington, DC), vol. 13, no. 4, p. 9-11; p. 27-30.
- . 1996. Changing the agency for learning: acquiring expert performance. In: Ericsson, A., ed. *The road to excellence: the acquisition of expert performance in the arts and sciences, sports and games*, p. 303-311. Mahwah, NJ, Erlbaum.
- Glaser, R.; Chi, M.T.H. 1988. Overview. In: Chi, M.T.H.; Glaser, R.; Farr, M.J., eds. *The nature of expertise*, p. xv-xxviii. Hillsdale, NJ, Erlbaum.
- Glaser, R.; Silver, E. 1994. Assessment, testing, and instruction. In: Darling-Hammond, L., ed. *Review of research in education*, vol. 20, p. 393-419. Washington, DC, American Educational Research Association.
- Greeno, J. 1991. Number sense as situated knowing in a conceptual domain. *Journal for research in mathematics education*, vol. 22, p. 1-49.
- Holloway, S. 1988. Concepts of ability and effort in Japan and the U.S. *Review of educational research* (Washington, DC), vol. 58, p. 327-345.
- Lane, S.; Glaser, R. 1994. Assessment in the service of learning. In: Husen, T.; Postlethwaite, T.N., eds. *The international encyclopedia of education*. 2nd ed., p. 370-374. Oxford/Leuven, Pergamon Press.
- Lave, J. 1988. *Cognition in practice: mind, mathematics and culture in everyday life*. Cambridge, U.K., Cambridge University Press.
- National Center on Education and the Economy & University of Pittsburgh. 1997. *Performance standards*, vols. 1-3, Pittsburgh, PA, New Standards.
- Oakes, J. 1985. *Keeping track: how schools structure inequality*. New Haven, CT, Yale University Press.
- . 1990. *Multiplying inequalities: the effects of race, social class, and tracking on opportunities to learn mathematics and science*. Santa Monica, CA, Rand.
- Oakes, J.; Lipton, M. 1990. Tracking and ability grouping: a structural barrier to access and achievement. In: Goodlad, J.; Keating, P., eds. *Access to knowledge: an agenda for our nation's schools*, p. 187-204. New York, College Entrance Examination Board.
- Peak, L. 1993. Academic effort in international perspective. In: Tomlinson, T., ed. *Motivating students to learn: overcoming barriers to high achievement*, p. 41-59. Berkeley, CA, McCutchan Publishing.
- Resnick, L.B. 1995. From aptitude to effort: a new foundation for our schools. *Daedalus* (Cambridge, MA), vol. 124, p. 55-62.
- Resnick, L.B.; Nelson-LeGall, S. (in press). Socializing intelligence. In: Smith, L.; Tomlinson, P., eds. *Piaget, Vygotsky and beyond*. London, Routledge.

- Rogoff, B. 1990. *Apprenticeship in thinking: children's guided participation in culture*. New York, Oxford University Press.
- Rogoff, B.; Lave, J., eds. 1984. *Everyday cognition: its development in social context*. Cambridge, MA, Harvard University Press.
- Saxe, G.B. 1990. *Culture and cognitive development: studies in mathematical understanding*. Hillsdale, NJ, Erlbaum.
- Scardamalia, M.; Bereiter, C. 1991. Higher levels of agency for children in knowledge building: a challenge for the design of new knowledge media. *Journal of the learning sciences* (Mahwah, NJ), vol. 1, p. 37-68.
- .; —. 1996. Adaptation and understanding: a case for new cultures of schooling. In: Vosniadou, S.; et al., eds. *International perspectives on the design of technology-supported learning environments*, p. 149-163. Mahwah, NJ, Erlbaum.
- Shepard, L.A. 1991. Negative policies for dealing with diversity: when does assessment and diagnosis turn into sorting and segregation? In: Hiebert, E.H., ed. *Literacy for a diverse society: perspectives, practices, and policies*, p. 279-298. New York, Teachers College Press.
- Sternberg, R.J., et al. 1995. Testing common sense. *American psychologist* (Washington, DC), vol. 50, no. 11, p. 912-926.
- Stevenson, H.; Lee, S. 1990. Contexts of achievement: a study of American, Chinese, and Japanese children. *Monographs of the society for research in child development*, vol. 55, nos. 1 & 2, Serial No. 221.
- Thorndike, E.L. 1975. Mr. Binet's test 70 years later. *Educational researcher* (Washington, DC), vol. 4, p. 3-7.



---

**OPEN FILE**

**THE EVALUATION OF  
EDUCATION SYSTEMS:  
THE VIEW AT THE END  
OF THE 1990s**

---

# INTRODUCTION

---

## TO THE OPEN FILE

---

*Alejandro Tiana Ferrer*

---

At the end of the 1990s, education systems are facing a decisive challenge, namely that of improving the quality of education offered by institutions of learning. The scientific, technological and economic challenges experienced by our societies, the powerful social demand for a fuller and more broad-ranging education, the need to make the best possible use of scarce resources and the pressure in favour of sustainable and equitable development are forcing public authorities to respond to this inescapable requirement. Improving the quality of education has thus become a fundamental objective, sufficient to motivate a substantial proportion of current educational policies. As a number of international fora have stressed, the main challenge for education systems nowadays is not just providing education for all citizens, but making sure that the education provided is of a high-enough quality (OECD, 1992).

Among the factors that contribute to the qualitative improvement of education, the evaluation of the education system and its components undoubtedly stands out. In effect, any measures taken to improve educational activity have to be founded on accurate, comparative and balanced information. By offering a means of both obtaining and disseminating such information, evaluation provides a powerful tool of knowledge, measurement, democratic debate and, ultimately, change (Delors,

---

*Original language: Spanish*

*Alejandro Tiana Ferrer (Spain)*

Professor of the history of education and of comparative education at the National Distance Education University, Madrid. From 1989 to 1996, served as Director of the *Centro de Investigación y Documentación Educativas* (CIDE) and the *Instituto Nacional de Calidad y Evaluación* (INCE). He has been President of the Consortium of Institutions for Development and Research in Education in Europe (CIDREE) and of the governing board of the Centre for Educational Research and Innovation (CERI-OECD). Author or joint author of twelve books and over seventy studies on different subjects related to the history of education, comparative education and educational evaluation.

1996). It is hardly unexpected that, in this context, it has even been said that evaluation constitutes a basic tool in the service of a new style of policy-making and education system management (De Landsheere, 1994; House, 1993; Michel, 1996).

This being the case, many countries have started recently to develop a variety of educational evaluation plans and programmes. The initiatives undertaken so far have succeeded in opening up some promising approaches, creating new organizations and gaining valuable experience with regard to the possibilities and limitations of the use of evaluation in education.

But, despite the advances achieved in this area, the rapid development which evaluation activity is undergoing is obliging the educational authorities to explore new lines of action, to reflect about the experiments they have tried and to extract their own conclusions from the results. One of the most fruitful ways of conducting this reflection consists precisely in comparing home-grown ideas and initiatives with those of other countries. International debate provides a unique opportunity of improving our capacity to resolve the problems we come across in our domestic environment and for improving our practice.

These were precisely the motives which led to the holding in Rio de Janeiro of an International Seminar on Educational Evaluation in December 1997, organized by the Brazilian Ministry of Education and Sport and UNESCO's International Bureau of Education. The Seminar was attended by some 150 evaluation specialists, researchers and teachers at all educational levels, coming from fifteen countries of America, Europe, Asia and Oceania, as well as from the Brazilian federal and state ministries. This *open file* contains a selection of the papers presented at that seminar.

## **Consideration of new trends in educational evaluation**

In the light of the experience accumulated in the area of evaluating education systems, this open file aims to contribute to the discussion about what it means to evaluate an education system and what are the main challenges this task has to face at present. Rather than repeating familiar ideas and approaches, it tries to analyse and discuss the new trends that are taking shape. Discussing new trends entails exploring the recent past, as well as analysing the present situation, while still looking towards the future. This exercise is particularly necessary when we are dealing with an area undergoing such radical change.

Although the actual process of evaluation has a long tradition behind it, going back to the historic origins of psychometrics and experimental pedagogy, in the last three decades it has followed a course of marked, rapid expansion, with significant changes occurring in the way it is conceived, in the methods it uses and in the way it is organized institutionally. Evaluation has emerged from this series of changes completely renovated, to such an extent that present-day educational evaluation hardly bears any resemblance to the activity referred to as such as recently as the early 1970s.

One of the aspects where the new approach to evaluation is particularly noticeable is the practice of using it systematically to obtain evaluative information

about all the areas of education, and not just students' learning achievements. To take a common suggestive image, it is as if it had expanded by gradually colonizing neighbouring territories. After concentrating initially on students' learning, attention has gradually shifted to teachers, curricula, intervention programmes, schools and the actual educational administration. It could be said, in other words, that nowadays the whole of the education system is open to evaluation.

As might be expected, this expansion of the focus of attention and its application to new areas of activity have led to a careful reconsideration of the concepts and theoretical categories used, a review of the methods and procedures normally applied, a search for institutional and organizational models that are better suited to the development of the new functions and to the search for more effective ways of disseminating and making use of the information obtained. In this area, as in so many others, there are no universal standards handed down by tradition, but only trends offering greater or lesser promise, which have to be analysed in order to extract criteria applicable to individual situations.

The open file begins precisely by referring to the conceptual changes brought about by the current diversity of functions of evaluation as a consequence of its expansion. Following this line of thought, Caroline Gipps, in her paper, considers the connections and cross-influences between evaluation and learning, a topic of particular relevance at a time when rapid economic and social changes are increasingly obliging States to encourage their citizens to develop higher cognitive abilities, oriented to strategic learning. From this point of view, the existing relation between established evaluation mechanisms and the style of learning which they effectively promote is carefully analysed, with a view to uncovering any undesired side-effects.

The gradual expansion in the traditional functions of educational evaluation has given rise to a need not only to elaborate new conceptual frameworks, but also to develop new methods and tools for evaluating education systems. Two of these have been selected here, because of the interest they have aroused and their current spread.

Tjeerd Plomp's paper concentrates on the international comparison of educational results, a task which has involved several international organizations and associations, and which has given rise to a growing number of studies in recent years. He uses the work done by the International Association for the Evaluation of Educational Achievement (IEA) over the last thirty years, and especially its latest expression, the Third International Mathematics and Science Study (TIMSS). As a starting point, TIMSS looked at the different purposes which these types of study can fulfil (description of the level of results achieved; construction of criterial performance scales, monitoring of the changes occurring in education systems; analysis of differences between countries; transnational educational research) and at their main difficulties and limitations.

Norberto Bottani, in his contribution, scrutinizes recent attempts at devising and using educational indicators, understood as a form of tool which is able to provide significant, coherent information and can be a means of assessing the state and situation of education systems. On the basis of the experience gained with the

CERI-OECD project on International Indicators of Education Systems (INES), he analyses the functions of indicators in current education systems, the contributions they can make to decision-making in education and the conditions they have to meet in order to be effective. In his analysis, he combines theory with an operational and properly political approach, highlighting the interrelations between the three aspects.

But evaluating education systems does not mean taking only a macroscopic view. Recent research into factors contributing towards an improvement in the quality of education all agree in attributing a central role to the school itself. While the effect produced by national and local policies should not be disregarded, any qualitative improvement will depend in the end on individual educational establishments. Hence the growing importance given to evaluating schools as a means of improving education. In other words, discussing educational evaluation also implies considering what specific purposes are served by assessments of schools, what role they play in the evaluation of the education system as a whole, how internal and external evaluation initiatives can and should complement each other, and what role educational authorities and the schools themselves may be expected to play in institutional evaluation.

David Nevo, in his article, suggests that there is a need to devote a special area to the evaluation of schools, since they are really the place where the educational task takes place and where its effectiveness has to be assessed. To achieve this, he considers it fundamental to combine the self-evaluation performed by the schools themselves with external evaluation, based on the view that the dialogue between them is the most powerful tool for achieving institutional improvement. Applying these principles in practice requires giving careful consideration to the possibility of developing evaluation centred on the school, but also changing the approach to evaluation in order to make it less coercive.

One of the problems which is of most concern to policy makers and evaluators is how to establish a fair comparison between schools. The practice of establishing comparisons between schools, as a means of acquiring knowledge, of rendering accounts or of emulation, is spreading. Nevertheless, it does raise serious problems of fairness in the evaluation, since the background conditions of students will considerably affect the performance of a school. To offset this, various procedures have been developed to monitor this effect and to introduce it in the analysis.

In her paper, Sally Thomas surveys the new statistical methods which are being used in the United Kingdom to calculate 'value-added' measures, understood in the sense of the contribution made by schools to their students' learning, taking into account the latter's socio-economic characteristics and their starting level of knowledge. These new approaches raise the issue of the effectiveness of teaching, distinguishing between the different results which a particular school promotes in its students, the ages at which the best results are achieved and the characteristics of the school population which most benefits. One of the main contributions of these new methods is to suggest that educational effectiveness has to be considered not so much in general terms as in the light of individual circumstances and of an analysis of the areas of greater or lesser effectiveness of individual schools.

One of the questions which causes most concern whenever attempts are made to relate educational evaluation to education policy is the type of institutional mechanism involved. In more practical terms, the type of dependence maintained by the bodies responsible for evaluation with respect to the political authorities and educational administrations is often the subject of debate due to its implications. This is because the institutional model chosen will, to a great extent, determine the degree of independence enjoyed by evaluators in their task.

This is the topic dealt with by Ernest House, who highlights the concept of the 'credibility' of evaluation, which turns out to be more important than the concept of its independence. The basic problem faced by national evaluation systems is achieving credibility within their own context, as a means of exerting real influence. In order to achieve it, evaluation has to meet three basic requirements: it must be democratic, dialogic and deliberative. Even though it may serve to criticize policy decisions, it still provides a fundamental tool for the construction and management of policies in advanced capitalistic societies.

Lastly, education policy now faces the challenge of discovering how to process and use the information supplied by evaluation. Even though the strictly utilitarian concept, whereby the results of evaluation provide an immediate tool for decision making, has given way to a more enlightened approach, which accepts the possibility of indirect forms of influence, the existing relation between evaluation and policy-making continues to attract extensive interest. The proper dissemination of evaluation results is now more of a concern than ever to the supporters of evaluation, and to the evaluators themselves. The search for new methods of disseminating information is attracting increasing attention, both on account of the underlying democratic need and for the sake of simple effectiveness. The role of information intermediaries, who provide a link between evaluators and evaluation promoters and their audiences, is being given greater importance by those who believe that they perform a useful task in transforming the results of evaluation into usable knowledge. In other words, aspects such as the information flows associated with the process of evaluation, the uses made of it and the impact it can produce on education policy and practice reappear regularly in the current debate on evaluation.

Barry McGaw's paper tackles the subject of the use of the results of evaluation, analysing the shift in the focus of evaluation which is occurring at present, away from simple learning results towards what are referred to as benchmarks, or scales of performance criteria. This expression, which has been borrowed from the world of industry, has provided new conceptual and methodological tools for the evaluation, for the interpretation and for the use of educational results. Furthermore, evaluation can be put to different uses, according to the level of analysis adopted (macroscopic, for the education system as a whole; microscopic in the case of individual educational establishments), each one of which entails different requirements. There is no doubt that these new trends will help turn evaluation into a more useful tool for the interpretation of educational phenomena and for decision-making in the years ahead.

## A work agenda open to the future

As will be appreciated, the themes selected are only a small sample of the challenges and trends that are currently facing educational evaluation. In this as in other areas, there is a need to continue concentrating on a process of trial, reflection and debate as the only way of finding answers to the many questions raised.

To achieve this, the initiatives undertaken by national authorities can and should be closely complementary with those developed by international organizations. The work agenda open to the future is a broad one and it would not make sense to tackle it in diverging ways. Every effort must be made to take advantage of the possibilities offered by exchanges of experience and co-operation, which are invariably more productive than work in isolation. In the last resort, all the countries that have decided to give impetus to educational evaluation face similar problems. The guidelines of the agenda, which should gradually be falling into place, should highlight three types of action that are of particular relevance in the present circumstances.

Firstly, there is a need to promote joint research programmes in the area of educational evaluation. Even though some major projects have already been undertaken in some countries, crossing national frontiers always opens up new perspectives and suggests different solutions. The experiment conducted by the European Union in recent years shows how building up networks of evaluation researchers can make a powerful contribution to the generation and dissemination of knowledge in this area. From another point of view, international evaluation projects, such as those sponsored by the IEA, OECD or UNESCO's Santiago Office, provide a priceless opportunity for exploring new approaches and methods applicable to different national realities and contexts.

Secondly, there is a need to continue promoting exchange and co-operation in this area. The introduction of national systems of educational evaluation is still only a recent experience for many countries, about which much still remains to be learned. Because of the characteristics of this type of activity, the search for solutions must be taken beyond national frontiers, through juxtaposition, comparison and a critical and controlled transfer of knowledge. International meetings are an ideal occasion for bringing together the leaders and specialists involved in developing these new systems, so that they can discuss the problems they face together and begin to work out a few possible solutions. In some cases, holding regional meetings can help to support experiments that are still only at the beginning stage, through the adoption of shared programmes of action.

Thirdly, there is a need to continue improving the training of professionals working in this field. It appears that there are at present limited sectors of highly qualified specialists, but that there are still not enough well-trained experts to meet the growing demands of evaluation. It would also be advisable to review the evaluation component in initial and continuous training programmes for teachers. When so much is being said about the need to promote an *evaluation culture*, it may be worth paying particular attention to the training contents which would be most appropriate

in that respect. Considering that evaluation experience and knowledge are shared internationally, it is all the more appropriate to seek international co-operation. For this reason, the work agenda of the future should set aside a section for training tasks.

Lastly, a special place should be kept on the future agenda to meet the specific needs of developing countries. In all three areas of the generation and dissemination of knowledge, international exchange and co-operation and training tasks, these countries have to play a significant role if they are not to be left aside by the new trends which are now emerging in the developed countries. Without it presupposing purely mimetic assimilation or the creation of a new dependence, there is a need to contribute towards the build-up in those countries of material for reflection, professional skills and organizational methods which can extract the greatest wealth possible from evaluation, through intelligent adaptation of their priorities, agendas and necessities. The decisions taken by the international organizations to continue working in this area in the years ahead will need to take account of this specific requirement and to integrate it within their plans of action.

## References

- De Landsheere, G. 1994. *Le pilotage des systèmes d'éducation* [Steering education systems]. Brussels, De Boeck.
- Delors, J., et al. 1996. *Education: the treasure within*. Paris, UNESCO.
- House, E. 1993. *Processional evaluation. Social impact and political consequences*. Newbury Park, CA; London; New Delhi, Sage.
- Michel, A. 1996. La conducción de un sistema complejo: la Educación Nacional [The management of a complex system: national education]. *Revista Iberoamericana de Educación* (Madrid), no. 10, p. 13–36.
- Organization for Economic Co-operation and Development. 1992. *High-quality education and training for all*. Paris, OECD.



---

# STUDENT ASSESSMENT

---

## AND LEARNING

---

### FOR A CHANGING SOCIETY

---

*Caroline Gipps*

---

This article examines the traditional roles of assessment and the current demands for assessment to 'gear up' education systems in response to global economic changes. It argues that in the next millennium, and with the information revolution, it will be important to foster higher order skills and good learning strategies in a high proportion of our students. The article goes on to consider, set against the critical relationship among high-stakes testing, teaching and learning, the nature of high-quality assessment programmes. It will present evidence of impact of the national assessment programme on teaching in England and Wales in the United Kingdom.

#### **Introduction**

A number of authors have argued that a paradigm shift is taking place in assessment: a move from a testing and examination-as-hurdle model (where you make the exam as difficult as possible and give the candidate little guidance, the proof of quality being in the numbers that fail) to an assessment model where we try to give all

---

*Original language: English*

*Caroline Gipps (United Kingdom)*

Ph.D. (London). Professor of Education and Dean of Research at the University of London Institute of Education. She trained as a psychologist and worked as a primary school-teacher before moving into a career in research. Current research is on teaching, assessment and feedback strategies of teachers for schoolchildren in years 2 and 6. She has published widely on policy and practice in assessment. Recent publications include: *A fair test? Assessment, achievement and equity* (with P. Murphy) (1994); *Assessment: a teacher's guide to the issues* (with G. Stobart) (1996); and *Equity in the classroom: towards effective pedagogy for girls and boys* (with P. Murphy) (1996).

candidates a real opportunity to show what they know, understand and can do (by giving more guidance, by sharing criteria with the student, and making the tasks match real life or classroom tasks). A paradigm shift or 'scientific revolution' occurs when the old paradigm is unable to deal with an outstanding problem.

Assessment is now required to achieve a wide range of purposes: it has to support teaching and learning; provide information about pupils, teachers and schools; act as a selection and certification device; serve as an accountability procedure; and drive curriculum and teaching. The new forms and range of purposes for assessment mean that the traditional model underpinning assessment theory is no longer adequate.

The shift is caused in part by this dilemma: there are increased demands for testing at the national level (for a broad range of reasons which I will address) that must offer comparability, while, at the same time, our understanding of cognition and learning tells us that we need assessment to map more directly onto the processes we wish to develop, including higher order skills, which makes achieving such comparability more difficult. However, it is not just that we wish to move beyond traditional testing and its associated technology, but that the change involves a much deeper set of transformations, hence the paradigm shift: our underlying conceptions of learning, of evaluation and of what counts as achievement are now radically different from those which underpin the traditional model of assessment (Gipps, 1994).

The message of this paper is that assessment is an important part of education, and that whenever possible it must be of a type suitable to, and used for, the enhancement of good quality learning. This is not to say that traditional standardized tests and examinations have no role to play in assessment policy, but they are not the only approaches. We need to design assessment programmes that will have a positive impact on teaching and learning.

## Functions of assessment

High-status examinations have traditionally been used to select, sort and classify. For example, civil service exams were instituted in the United Kingdom in 1855 to identify potential members of the civil service through means other than patronage. These early exams were used to identify competence and to qualify for entry to particular professions or universities. They had two 'side effects'—one was the development of the notion of a syllabus or curriculum, and the other was that the use of a written, theoretical test for entry into high-status professions or institutions invested the assessment technique itself with a similar high status—a status it still retains (Broadfoot, 1996).

My sociologist colleagues, notably Patricia Broadfoot, describe assessment as a device used within developed societies with mass educational provision as a mechanism of social control; as a rational and justifiable basis for the allocation of 'unequally desirable social roles' as individuals are required to demonstrate competence through a certification procedure based on competition. The use of a justifiable procedure and the notion of competition in which individuals compete on an apparently equal basis, allows those who are not successful to accept their own

failure. Intelligence quotient (IQ) testing was a means of social control 'unsurpassed in teaching the doomed majority that their failure was the result of their own in-built inadequacy' (Broadfoot, 1979, p. 44).

National systems will always need to monitor levels of performance for planning and accountability purposes. But a new role for assessment has emerged recently. Assessment is being used around the developed world to 'gear up' education systems in response to global changes. Economics drive the rhetoric: technological developments demand better educated, more thoughtful and flexible workers across the labour market to strengthen the country's technological base and to foster a spirit of enterprise and initiative. The apparent mismatch in the 1980s between the output of schools and the needs of the labour market in the United Kingdom, as indicated by the number of unqualified school leavers and by the number of young unemployed, suggested that education had departed from the 'real world' of work and the result has been to seek to reconnect education with the economy (Neave, 1988). But the model of schooling which is developing in England and Wales as a result is one of traditional subject boundaries, of a 'transmission' model of teaching within formal classrooms, of tests and exams which are being pushed to a formal paper and pencil, one-off examination model.

In fact, what we need in order to prepare young people for the next millennium is to teach *all* of them not only the basic skills, but also the higher order skills of problem solving, critical thinking and evaluation. We must teach them how to become effective learners aware of, and in control of, their own learning, as the information revolution, together with the regular job and career changes anticipated in the next century, require us all to keep learning beyond school.

## Assessment and learning

Assessment theorists have not traditionally talked about learning. But there is a critical relationship among what is tested and how it is tested, with what is taught and how pupils learn. In order to encourage the teaching and development of higher order skills, thinking processes and problem solving, we need to use assessment which directly reflects these processes.

Standardized achievement tests examine students' abilities to recall and apply facts learned routinely. Even items which are designed to assess higher level activities often require no more than the ability to recall the appropriate formula and to make substitutions to get the correct answer. Students who conceive of knowledge as collections of facts will use surface learning strategies that are aimed at successful memorization. Deep learning in 'good' learners, on the other hand, involves thinking about the meaning of what is being learned (Entwistle, 1992). (I am reminded though that Bertrand Russell said 'most people would die sooner than think, and most people do.') This notion of purposeful or deep learning clearly has implications for curriculum and pedagogy, as well as for assessment.

Traditional teaching for tests and examinations often encourages rote learning. Rote learning is essentially passive and can lead to shallow or surface learning: the

learner masters facts which s/he can then recall in a test situation. Pupils may become very efficient at this form of memorization, but facts learned in this way may be quickly lost or forgotten. Information to be retained must either be understood, ‘interacted with’ and logged into conceptual maps in the brain, or used repeatedly after learning. Shallow learning will allow pupils to manipulate formulae or work through detailed exercises even though they do not understand fundamental principles. This means that pupils will find it hard to use these facts/concepts in other circumstances or contexts, thus limiting the value of their learning.

On the other hand, deep learning involves not only understanding, but an intention to understand the material, and requires an active approach to learning. By active learning I do not mean that learners have to be seen to be physically busy, or to be ‘discovering’ things but that they must be actively thinking for themselves and interacting with the material. The teacher is expected to teach, but learning can only take place within the pupils (see Table 1).

TABLE 1. Deep learning and shallow learning

Deep learning approach	Surface or shallow learning approach
An intention to develop personal understanding	An intention to be able to reproduce content as required
Linking ideas together using integrating principles	Lack of recognition of guiding principles or patterns
Relating evidence to conclusions	Focusing learning on assessment requirements
Active interaction with the content, particularly in relating new ideas to previous knowledge and experience	Passive acceptance of ideas and information

Source: Entwistle, 1992; Marton & Saljo, 1984.

Of course, rote learning is important for some purposes: there are some things that are probably most efficiently learned by rote, such as number bonds, spelling and multiplication tables. These are skills that are most useful to us when they have been practised sufficiently to become automatic. They are also frequently needed in everyday life, so they are not easily forgotten. Similarly, there is not time for pupils to approach all learning in a deep way—hence the notion of ‘strategic learning’ which entails a judicious mix of surface and deep learning.

## Assessment to support learning

In order to encourage and support a mix of learning approaches we need to employ a mix of assessment approaches.

- Multiple-choice or short answer tests are a quick and efficient way of testing recall of simple facts, e.g. number bonds, foreign language vocabulary and basic skills such as multiplication and spelling. Unless the items are very carefully constructed, they are ineffective in assessing the understanding of concepts.
- Essay-type answers, practical tasks and projects (as part of a timed test, an examination or part of course-work, or as part of everyday classroom assessment) are better able to assess understanding and to encourage a deeper level of learning. Most importantly, if understanding is to be assessed, methods are required that involve the learner in applying their knowledge and linking it to real contexts. It cannot be assessed by asking for the recall of isolated, decontextualized pieces of information which will shift teaching and learning away from understanding and towards the memorization of the information necessary to succeed in the assessment.
- For effective progress in learning, tasks have to be pitched at an appropriate level of difficulty. For the teacher to know the pupil's level of progress, informal observation, questioning and providing assessment tasks on a regular basis are important; this so-called 'teacher assessment' has the key features of informality and regularity so assessments are made across contexts and time to build up a more complete picture than in a one-off test or exam.

Higher order skills such as analysis, interpretation, critique, synthesis, applying knowledge and skills to new tasks, constructing a convincing argument, etc., are complex activities which cannot generally be assessed in simple ways that involve ticking right answers or machine-scoring. Part of our challenge for the new millennium is to teach such higher order skills to a much broader range of students than in the past. Assessment of meaning and understanding is therefore crucial in developing such skills.

Thus in order to encourage and support strategic learning, and to develop effective learners, we need a full range of assessment approaches matched appropriately to particular learning objectives:

- formal and informal tests to assess recall of facts and acquisition of basic skills;
- more detailed and complex tasks and exams which can assess understanding and application; and
- informal, continuous teacher assessment (TA) based on observation, questioning and routine classroom learning assessment tasks.

To summarize, we must test for meaning and assess understanding; we must not shy from assessing complex skills in context or authentic settings. Good assessments model the real learning activities that we wish students to engage with, and teachers to teach.

## Assessment in a changing society

So, what do we know about good practice in assessment that will help us design assessments for the future?

There are some general principles: instead of focusing on ranking and comparison

in relation to some notion of 'average' performance—a profoundly limiting concept—we need to support an approach in which clear standards are set for performance, against which pupils are assessed; these standards and assessment processes are shared with pupils (progressively so as they get older). Assessment which ranks and classifies in relation to others should be left until as late as possible in the education system; this should not be the model used at earlier stages. Pupils must be encouraged to monitor and reflect on their own work/performance (with the positive and constructive help of teachers). Feedback to pupils, which is a key factor in the assessment process, must emphasize mastery and progress.

The publication of test data at the class and the school levels distorts the educational process and encourages 'cheating' of various kinds. Teachers and pupils cannot avoid this: they are caught in a trap. And, in this situation, who takes responsibility when teachers do not teach the things that are not tested? Where high-stakes external tests are unavoidable, using performance-based assessment together with TA will mitigate the worst effects on teaching but even this, if stakes are too high, will over-promote the assessed activities and run the risk of narrowing the curriculum. When high-stakes tests and exams allow a high proportion of pupils to fail, there are self-esteem and motivation problems for low-scoring pupils, with the concomitant risk of drop-out. Assessment against clear standards, in a low-stakes programme, with constructive feedback and a focus on the individual's own performance, rather than comparison with others, is more likely to maintain the engagement of pupils and retain them in the system.

Good quality assessment is time consuming and therefore needs to use quality tasks so that pupils are not wasting their time: the tasks need to be anchored in important and relevant subject matter and the nature and mode of the tasks need to be based on what we know about equitable and engaging assessment.

High stakes are unlikely to be reduced with assessment for certification and selection purposes and therefore the style and content of assessment tasks is particularly crucial: we need to design accountability assessment which will provide quality information about pupils' performance without distorting good teaching practice. We also need to remind policy makers of their responsibility for the evaluation of the educational and social consequences of test use at every level.

These are some of the things we know about assessment which should help us to guide assessment in the near future. But what about assessment in the long term?

We cannot describe assessment for the millennium until we can describe education for the millennium. We are only beginning to have a view of education in a post-modern world. There is agreement that economic globalization, cultural commodification and the compression of time and space which comes with the information revolution imply radical transformations in education (Kenway, 1992; Green, 1994) but there is less agreement about what this might mean in, for example, curricular terms. We have to consider the nature of education for the millennium in relation to the knowledge and information revolution: knowledge is one of the most important competitive factors in the future of any society. We live in an 'instantaneous, twenty-four hour information world' (Dalin & Rust, 1996, p. 33) or put in another way: all information will only be six seconds away.

This means that:

- students will be bombarded by information and teachers will need to help them organize, conceptualize and select from this huge array;
- learning how to learn will become a fundamental aim of the curriculum;
- the notion of the textbook has to be rethought (giving 'students basic knowledge and overviews and help them to explore the many facets of a field', *ibid.*, p. 146);
- the classroom of the future will expand beyond its four walls as students move out to access information and experience while information from outside is brought into the classroom; and
- the teacher's role will inevitably change because of these other changes. Different organizational structures with larger groupings for some purposes, accompanied by use of teacher aides, students working together and using networked information retrieval and computing systems, all seem likely.

That great servant of education, Jerome Bruner, writing about the future, argues that we need to move on from an impoverished conception of teaching '... in which a single, presumably omniscient teacher explicitly tells or shows presumably unknowing learners something they presumably know nothing about' (Bruner, 1996, p. 20). Bruner talks instead of developing the classroom as a community '... of mutual learners, with the teacher orchestrating the proceedings. *Note that, contrary to traditional critics, such sub-communities do not reduce the teacher's role nor his or her "authority"*. Rather, the teacher takes on the additional function of encouraging others to share it. Just as the omniscient narrator has disappeared from modern fiction, so will the omniscient teacher disappear from the classroom of the future' (p. 21–2, my emphasis).

How will we conceive of assessment in the classroom of the future? Not, one hopes, of children sitting in front of rows of computers completing endless series of worksheets. Bad pedagogical practice does not become good pedagogical practice by being translated from pencil and paper to disk and screen, nor by the adoption of multiple-choice tests which are machine-marked. The process is quick and efficient, but it means that students do not have to be able to write anything. In the next millennium we will still want youngsters to be able to write, even if they do it on a keyboard.

A number of issues keep recurring in discussions of the future: emphasizing the local rather than the global (Kenway, 1992; Harrison, Bailey & Dewar, 1996) and the subjective rather than the objective (Lincoln, 1992); accepting a range of methodologies as valid (Lather, 1992; Griffiths, 1992); and the role of technology.

How do these relate to assessment?

1. the local rather than the global: We must emphasize the role of the teacher so that assessment can be grounded in a 'local' context supporting the professionalism of the teacher, and flexibility of approach (whilst maintaining rigour by reporting against national standards and being subject to external moderation).
2. the subjective rather than the objective: This means putting the learner at the core in assessment. It also involves re-defining the power relations in assessment

and involving the learner more as a partner. This does not mean teachers giving up control—it means learners taking responsibility for their performance and monitoring their learning—what we call self-regulated learning and metacognition. We know that both of these are vital to effective learning and it is never too young to start.

3. accepting a range of methodologies as valid: Good practice requires that we use a range of assessment strategies so that all learners have a chance to perform well, and a range of pedagogic strategies (which include a range of materials, content, teaching styles and classroom arrangements) for different groups of pupils, and for different subject areas (Murphy & Gipps, 1996). The diversity among pupils as individuals and as learners supports the proposal for pedagogic strategies and a range of assessments. What we know about differences among learners in any case challenges traditional education systems that assume that everyone can and does learn in the same way (Gardner, 1991).

## **Findings from national assessment in England and Wales**

I want now to illustrate the impact of different types of assessment programmes on teaching. Assessment of every child aged 7 and 11 is carried out within primary schools in England and Wales. Pupils are assessed through standardized tests and TA in the areas of reading, writing, mathematics and science. This assessment programme has been in existence since 1991 and we have been fortunate enough to be able to carry out an evaluation of it for six years, focusing first on the assessment programme for 7 year olds (in Year 2, the end of Key Stage 1), and then for 11 year olds (in Year 6, the end of Key Stage 2).

### **THE NATIONAL ASSESSMENT PROGRAMME**

The 1988 Education Act legislated for a national curriculum and introduced national assessment. The national assessment programme is a crucial accompaniment to the national curriculum for it was through the assessment programme that standards were to be raised. The first stage of the development of the national curriculum and assessment programme was the setting up of the Task Group on Assessment and Testing (TGAT). The report of this group (DES, 1988) put forward a blueprint for the structure of the curriculum to which all subjects had to adhere. Subjects are divided up into a number of attainment targets and these are articulated at a series of eight levels. The series of levels is designed to enable progression: most 7-year-old pupils would be at level two in the system while most 11-year-old pupils would be at level four and so on. The attainment targets were articulated at each level by a series of criteria or statements of attainment (SoAs) which formed the basic structure of a criterion-referenced assessment system. (The numerous SoAs have now been changed to fewer, broad-level descriptions). The Secretary of State in 1988 accepted the recommendations of TGAT that pupils should be assessed by their teachers for



formative and diagnostic purposes over the ages 5–16, with standard assessment tasks (SATs) being used nationally to moderate TA at ages 7, 11 and 14, when they would be formally reported for summative purposes. The SATs at ages 7 and 11 would be performance assessments, related to extended themes and would resemble good classroom practice, but might incorporate formal tests in specific areas.

However, a large number of changes of ministers, policies and personnel in curriculum and assessment agencies has meant that the model of national assessment now implemented departs in a number of ways from that recommended by TGAT (Black, 1993; Gipps, 1992; Daugherty, 1995). Standard assessment tasks have become formal national tests, confined at age 11 to mathematics, English and science. Assessment at Key Stage 1 is now limited to English and mathematics, although science was included until 1993. The tests have been required to give reliable assessment information for each individual pupil, and thus have become more comprehensive, rather than sampling the curriculum thinly to moderate TA at the class level. Starting in 1996, the tests at age 11 were used as the basis for primary school 'league tables' although there are no such tables for the scores of 7 year olds. Little emphasis or resources have been devoted to TA, which therefore remains unclarified in regard to method. TA is reported to parents alongside national test results, but teachers may postpone making their assessments until after receiving the results of the national tests, and there is no requirement for moderation of standards.

#### THE RESEARCH PROJECT

The project 'National Assessment in Primary Schools: An Evaluation' (NAPS) has been funded in two phases at London University (the Institute of Education and King's College) by the Economic and Social Research Council. The first phase (1990–93) concerned the introduction of national assessment at Key Stage 1 (age 7) (Gipps et al., 1995); the second phase moved on to Key Stage 2 (age 11) but with a small element of continued monitoring of Key Stage 1 (Brown et al., 1996).

The project has focused on schools in four Local Education Authorities (LEAs). The LEAs were selected to be as nationally representative as possible. Eight schools were selected randomly by stratified sampling within each LEA, giving a total of thirty-two schools. Methods of gathering data included classroom observation, interviews and questionnaires.

Other researchers have carried out survey-type studies with larger samples; we designed our study to look in more depth at teachers' practice and their views.

## Findings

#### KEY STAGE 1 (AGE 7 ASSESSMENT)

Key Stage 1 assessment began in 1991. At this stage SATs were performance-based tasks conducted and scored by the teacher. In addition, TAs had to be made across the core curriculum *before* giving the SATs. Increased levels of discussions and

collegiality were observed in schools in the early stages of the implementation: from teachers of other ages to support second-year teachers who, it was felt, were faced with an awesome task; and among second-year teachers throughout the study to negotiate meanings for SoAs and to standardize judgements. Head teachers' involvement supporting their staff meant that there was a feeling of 'being in it together' which also helped to develop collegial ways of working. Helping each other with SATs and moderating SATs and TA brought teachers out of their classrooms and into working contact with each other.

One theme which we first came across in 1991 (Gipps, 1992) was the raised professionalism of second-year teachers; this was reported in one-third of our schools and involved second-year teachers leading assessment training and policy development. By 1994, five of the eleven primary head teachers who were left from the original study (having continued through to the second stage of our research looking at Key Stage 2) stressed the contribution their second-year teachers were making to assessment at Key Stage 2.

In only the second year of national assessment, the Government was able to show that standards had risen: the then Secretary of State for Education, John Patten, held a press conference at the end of 1992 proclaiming raised standards, since the percentage of the 7-year-old population reaching higher levels had risen in reading, spelling and mathematics ('Seven-year-olds' results show improving standards', Department for Education and Employment press release, 21 December 1992). Patten took this as evidence that the national curriculum was working. From our study we would say that this rise in levels of performance was due to teachers teaching more of what was required in the SATs: punctuation, spelling, handwriting and mental arithmetic. There was more attention to 'the basics' in 1992 and 1993, and this showed up in the children's improved levels of performance. The experience of doing the SATs also helped teachers to understand the SoAs so they could teach in a more focused way in later years. Of course, given the changes in the curriculum and to the assessment tasks in each of the three years, together with the lack of dependability of results (Shorrocks et al., 1992) such changes in patterns of performance should be treated with caution.

What our NAPS teachers were doing was not teaching directly to specific test items (although children were doing 'quick' mental arithmetic in order to be able to do sums in the test within five seconds), but teaching areas of the curriculum so that their children could do assessments on those topics. The difference between the situation which we have been observing and that in most other countries, where teaching to the test is observed, is that in England and Wales we have an imposed curriculum as well as imposed testing. There is, therefore, something other than tested items for teachers to teach towards. In addition, the assessment tasks have changed each year and there has been a rolling programme of attainment targets included in the testing, so teachers cannot easily become too narrow in preparing pupils for the tests.

What our study shows clearly is the shift in TA practice (notwithstanding individual variation) from an intuitive approach to one based on evidence and written

records. Also, it is clear that the bulk of our teachers became more knowledgeable in assessment rather than being technicians operating an imposed system. But more significant were the changes in classroom practice.

Greater care in planning, close observation of children and a more detailed understanding of individual progress impacting on teaching were reported by over half of the head teachers. The SATs acted as a training device, and group moderation broke down barriers. This came about, however, not because of having to give an external standardized test (as with sixth-year teachers and the Key Stage 2 testing in 1994), but because of the demands of TA and the performance-type activities in the SATs. These requirements, together with the (albeit too detailed) specification of the curriculum had, in the head teachers' views, led to an improvement in planning, teaching and assessing in second-year classes. We do not intend to downplay the problems of the assessment programme, for there were many (overload, too many and yet inadequate assessment criteria, low levels of training—particularly for TA, undependable results, etc.) but to make the point that if we wish to raise real standards of teaching and formative assessment (which in turn supports teaching and learning) then we need more than imposed external traditional tests.

#### KEY STAGE 2 (AGE 11 ASSESSMENT)

Key Stage 2 testing is at an earlier stage of introduction: it began in earnest in 1995. The assessments are, for the most part, standardized tests not assessment tasks. League tables are planned for this year, hence one might predict a fairly routine teaching-to-the-test impact.

This impact was directly observed in the following ways. In 1995, schools which had been involved in piloting Key Stage 2 tests since 1993 reported that they had changed their forms of organization as an outcome:

- fourteen had changed from mixed-ability teaching to some form of setting;
- eight had moved away from cross-curricular topic work towards more subject-based teaching;
- fourteen heads had introduced regular, formal testing throughout the whole Key Stage;
- the tests had also made an impact on teaching: e.g. a greater focus on reading and spelling and presenting mathematics in a way that children were more likely to meet in the test papers; and
- all of the teachers had done something to prepare their pupils for tests: revision in science, past papers in mathematics or timed tests in English.

We also found, however, that teachers had changed their teaching style in many cases. We interviewed twenty-nine Key Stage 2 teachers in depth about their teaching approach in 1995. Fifteen out of the twenty-nine teachers reported different teaching modes for different circumstances. They described when they would become chiefly 'transmitters' of knowledge, when they would encourage children to discover for themselves and when they felt they were 'constructing learning together'. These would change according to:

- teaching different subject areas;
- teaching different children;
- teaching different age groups;
- using different forms of pupil organization; and
- school ethos.

None of the teachers endorsed the 'transmission' approach to teaching on its own, yet few rejected it outright; they limited this approach to certain purposes.

That half of our sample explicitly espoused the mixed-mode approach to teaching is interesting in the light of our teachers' more general comments on the effect of the national curriculum on their teaching practice. Although not a single teacher believed that children learn solely from the transmission of facts, one-quarter felt pushed into doing more of this because of the amount of work to be covered by the national curriculum, or in some cases by the tests. They also felt unable to use children's ideas as starting points for work and regretted this. Examples of their comments are:

Considering the model 'the teacher conveys information to children and that is how learning takes place', I think that is the way I almost felt pushed to go by the weight of the curriculum.

'the teacher conveying information to children'. I think that has come much more to the fore with the national curriculum simply because of the volume of content that we have to deliver.

What I think and what the Government thinks are two completely different things. If I think how I was teaching five years ago, and how I'm teaching now. There always was input from me, but probably more and more now because there is so much to be got through.

## Conclusions

The effects on Key Stage 2 TA skills has been less marked so far than for Key Stage 1 teachers; this is due, we propose, to the limited requirements for TA and the standardized nature of the tests. However, because of the high stakes there has been more effect on their organization and teaching practise in order to achieve high results.

Through both Key Stages we can discern quite clearly a teaching to the test effect; since the curriculum is broad and assessment of the 'process' skills is required through TA, this effect is not, so far, particularly negative or narrowing. Although many teachers regret the loss of autonomy, the national curriculum and assessment programme has served to bring practice across schools closer, which may well be desirable in a national system.

The differences in the assessment programme between the two Key Stages offers an interesting comparison of impact. At Key Stage 1, the combination of moderated TA and performance assessment had a broad effect on teacher's practice—including more teaching of the 'basics'. At Key Stage 2, the combination of standardized tests, league tables and a downgrading of TA led to preparation for the tests, revision and more 'didactic' teaching.

The findings concerning teachers' views of assessment and changing practice in the primary phase are important, not only in England and Wales but also for other countries that are contemplating introducing a system of national testing. In 1993 and 1994, teachers in England chose not to carry out the national tests. This boycott should demonstrate the pitfalls of introducing a system without the support of the teaching profession. Their action led to the assessment programme being simplified. It seems that the current model of blanket external tests and separate TA is now perceived by teachers in England as a broadly acceptable one. Most teachers would prefer, rather than two separate results, some combination of TA with national test results to give a single, more valid result for each subject. This seems worth aiming towards as having the most positive impact on teaching and learning whilst keeping the curriculum broad.

## Bibliography

- Black, P. 1993. The shifting scenery of the national curriculum. In: Chitty, C.; Simon, B., eds. *Education answers back*. London, Lawrence & Wishart.
- Broadfoot, P. 1979. *Assessment, schools and society*. London, Methuen.
- . 1996. *Education, assessment and society*. Buckingham, U.K., Open University Press.
- Brown, M., et al. 1996. The validity of national testing at age 11: the teacher's view. [Submitted to *Assessment in education*.]
- Bruner, J. 1996. *The culture of education*. Cambridge, MA, Harvard University Press.
- Dalin, P.; Rust, V.D. 1996. *Towards schooling for the twenty-first century*. London, Cassell.
- Daugherty, R. 1995. *National curriculum assessment: a review of policy 1987–1994*. London, Falmer Press.
- Department of Education and Science. 1987. *The National Curriculum 5–16: a consultation document*. London.
- Department of Education and Science. Welsh Office. 1988. *National Curriculum: Task Group on Assessment and Testing: a report*. London.
- Entwistle, N. 1992. *The impact of teaching in learning outcomes in higher education*. Sheffield, U.K., CVCP Staff Development Unit.
- Gardner, H. 1991. *The unschooled mind. How children think and how schools should teach*. London, Fontana Press.
- Gipps, C., ed. 1992. *Developing assessment for the national curriculum*. London, ULIE/Kogan Page. (Bedford Way Series.)
- Gipps, C., et al. 1995. *Intuition or evidence? Teachers and national assessment of seven year olds*. Buckingham, U.K., Open University Press.
- Gipps, C. 1994. *Beyond testing: towards a theory of educational assessment*. London, The Falmer Press.
- Gipps, C.; Murphy, P. 1994. *A fair test? Assessment, achievement and equity*. Milton Keynes, U.K., Open University Press.
- Green, A. 1994. Post-modernism and state education. *Journal of education policy* (Basingstoke, U.K.), vol. 9, no. 1, p. 67–83.
- Griffiths, M. 1992. *Making a difference: feminism, post-modernism and the methodology of educational research*. Paper given to Economic and Social Research Council Conference, Liverpool.

- Harrison, C.; Bailey, M.; Dewar, A. 1996. Responsive reading assessments: is post-modern assessment of reading possible? *In: Paradigms in reading assessments*. London, Routledge.
- Kenway, J. 1992. *Marketing education in the post-modern age*. Paper presented to the American Educational Research Association Conference, San Francisco.
- Lather, P. 1992. Critical frames in education research: feminist and post-structural perspectives. *Theory into practice*, vol. 31, no. 2, p. 2–13.
- Lincoln, Y.S. 1992. *Emergent paradigms and policy conundra: problems in exercise and use*. Paper given to Economic and Social Research Council Conference, Liverpool.
- Marton, F.; Saljo, R. 1984. Approaches to learning. *In: Marton, F.; Hounsell, D.; Entwistle, N., eds. The experience of learning*. Edinburgh, Scottish Academic Press.
- Murphy, P.; Gipps, C., eds. 1996. *Equity in the classroom: towards effective pedagogy for girls and boys*. London, Falmer Press.
- Neave, G. 1988. Education and social policy: demise of an ethic or change of values? *Oxford review of education* (Oxford), vol. 14, no. 3, p. 273-83.
- Nuttall, D. 1987. The validity of assessments. *European journal of psychology of education* (Lisbon), vol. 11, no. 2, p. 108–18.
- School Examination Assessment Council. 1991. *National curriculum assessment at key stage 3: a review of the 1991 pilots with implications for 1992*. London, EMU:SEAC.
- Shorrocks, D., et al. 1992. *ENCA 1 project report*. London, School Examination Assessment Council.
- Wood, R. 1986. The agenda for educational measurement. *In: Nuttall, D.L., ed. Assessing educational achievement*. London, Falmer Press.

---

# THE POTENTIAL OF INTERNATIONAL COMPARATIVE STUDIES TO MONITOR THE QUALITY OF EDUCATION

---

*Tjeerd Plomp*

---

## Introduction

This paper will discuss the relevance of international comparative studies for evaluating and monitoring the quality of education systems. It upholds the premise that an international overview of local educational experiences brings forth new perspectives beneficial to those seeking to improve their education systems. The paper will focus mainly on first results of the Third International Mathematics and Science Study (TIMSS), which at present is being conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). The results of TIMSS will be used to illustrate the premise of this paper and the relevance of the IEA's mission: to contribute through international comparative studies to enhancing the quality of education. It will show by example how IEA-types of study may serve educational policy makers and practitioners. Concluding remarks will point to the importance of participating in international comparative assessment studies.

---

*Original language: English*

*Tjeerd Plomp (Netherlands)*

Chair of the International Association for the Assessment of Educational Achievement. Staff member of the Department of Curriculum, Faculty of Educational Science and Technology, University of Twente. He is also chair of the Committee on Multimedia in Teacher Training (COMMITT), an advisory committee to the Dutch Minister of Education, Culture and Science. E-mail: [plomp@edte.utwente.nl](mailto:plomp@edte.utwente.nl)

## Third International Mathematics and Science Study (TIMSS)

TIMSS is the largest and most ambitious study of comparative educational achievement ever undertaken. The TIMSS achievement test in mathematics and science included:

- Forty-five countries;
- five grade levels (3, 4, 7, 8, and the final year of secondary school);
- more than half a million students;
- testing in more than thirty languages;
- more than 15,000 schools;
- nearly 1,000 open-ended questions, generating millions of student responses;
- performance assessment;
- questionnaires from students, teachers and school principals containing about 1,500 questions;
- thousands of individuals to administer the tests and process the data.

TIMSS was conducted with attention to quality at every step. Rigorous procedures were applied to translate the tests and numerous regional training sessions were held on data collection and scoring procedures. Quality control observers monitored testing sessions. The samples of students selected for testing were scrutinized according to rigorous standards designed to prevent bias and ensure comparability. This monitoring of the quality of the study resulted in countries being indicated that did not meet all the quality criteria in the tables with their results.

TIMSS achievement results have been published by the International Study Center at Boston College in the United States.<sup>1</sup> Some of these results will be summarized and discussed to illustrate the potential richness of international comparative assessment studies.

Table 1 contains achievement test results for mathematics in Grades 7 and 8 (Beaton et al., 1996a). Figure 1 presents multiple comparisons for Grade 8 science achievement (Beaton et al., 1996b).

Table 1 and Figure 1 illustrate one of the purposes of international comparative achievement studies, namely to provide policy makers and educational practitioners with information (indicators) about the quality of their education system in relation to relevant reference groups of similar nations. This is called the 'mirror' function: it allows countries to determine whether or not they like their educational profile as compared to other countries.

Table 1 gives just 'horse race' data. For example, for Grade 8 mathematics, Spain is placed thirty-first and Portugal thirty-seventh out of forty-one countries. For Grade 7, Spain is placed thirty-second and Portugal thirty-sixth. Figure 1, however, provides more information. Figure 1 tells, for science, which countries have mean achievement scores that are significantly lower or higher than those of the comparison country, as well as the countries which do not have significantly different scores. Applying the approach of Figure 1 to Spain, for example, results in



TABLE 1: Average achievement in mathematics

Eighth grade*		Seventh grade*	
Singapore	643	Singapore	601
Korea, Rep. of	607	Korea, Rep. of	577
Japan	605	Japan	571
Hong Kong	588	Hong Kong	564
Belgium (Fl)	565	Belgium (Fl)	558
Czech Republic	564	Czech Republic	523
Slovak Republic	547	<i>Netherlands</i>	516
Switzerland	545	<i>Bulgaria</i>	514
<i>Netherlands</i>	541	<i>Austria</i>	509
<i>Slovenia</i>	541	Slovak Republic	508
<i>Bulgaria</i>	540	Belgium (Fr)	507
<i>Austria</i>	539	Switzerland	506
France	538	Hungary	502
Hungary	537	Russian Federation	501
Russian Federation	535	Ireland	500
<i>Australia</i>	530	<i>Slovenia</i>	498
Ireland	527	<i>Australia</i>	498
Canada	527	<i>Thailand</i>	495
<i>Belgium (Fr)</i>	526	Canada	494
<i>Thailand</i>	522	France	492
<i>Israel</i>	522	<i>Germany</i>	484
Sweden	519	Sweden	477
<i>Germany</i>	519	UK–England	476
New Zealand	509	United States	476
UK–England	508	New Zealand	472
Norway	506	<i>Denmark</i>	465
<i>Denmark</i>	503	UK–Scotland	463
United States	502	Latvia (LSS)	462
<i>UK–Scotland</i>	500	Norway	461
Latvia (LSS)	498	Iceland	459
Spain	493	<i>Romania</i>	454
Iceland	487	Spain	448
<i>Greece</i>	484	Cyprus	446
<i>Romania</i>	482	Greece	440
Lithuania	477	Lithuania	428
Cyprus	474	Portugal	423
Portugal	454	Iran, Islamic Rep. of	401
Iran, Islamic Rep. of	428	<i>Colombia</i>	369
<i>Kuwait</i>	392	<i>South Africa</i>	348
<i>Colombia</i>	385		
<i>South Africa</i>	354		

\* Grades 7 and 8 in most countries.

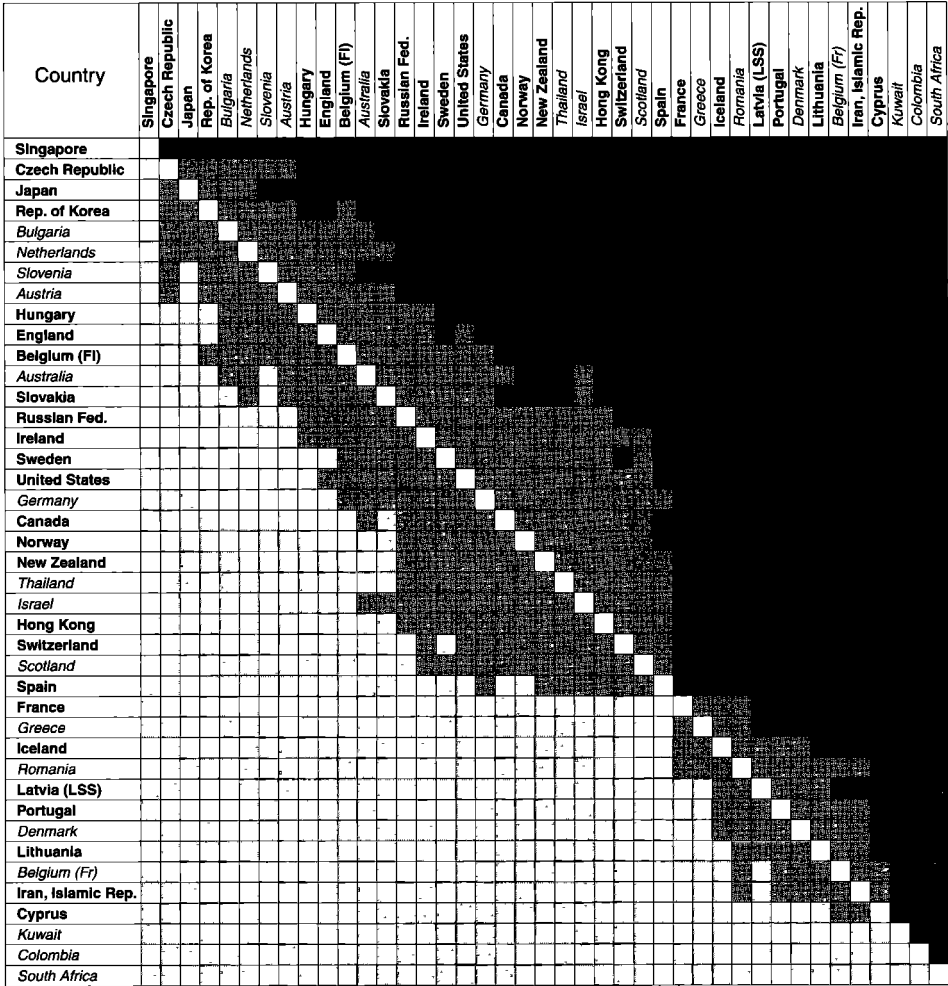
Latvia is annotated LSS for Latvian-speaking schools only.

Countries shown in italics did not satisfy one or more guidelines for sample participation rates, age/grade, specifications, or classroom sampling procedures. The report presents standard errors for all surveys estimated.

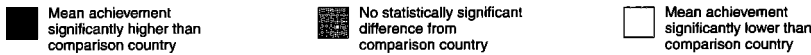
Source: IEA Third International Mathematics and Science Study (TIMSS), 1994–95.

FIGURE 1: Comparisons of achievement in the sciences—upper grade (Grade 8\*)

Instructions: Read across the row for a country performance with countries listed in the heading of the chart. The symbols indicate whether the mean achievement of the country is significantly lower than that of the comparison country, significantly higher than that of the comparison country, or if there is no statistical significance between the two countries.\*\*



Countries are ordered by mean achievement across the heading and down the rows.



\* Eighth grade in most countries; see Table 2 for information about the grades tested in each country.  
\*\* Statistically significant at .05 level, adjusted for multiple comparisons.  
Because coverage falls below 65%, Latvia is annotated LSS for Latvian-speaking schools only.  
Countries shown in italics did not satisfy one or more guidelines for sample participation rates, age/grade specifications, or classroom sampling procedures.

Source: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

an overview of countries which are performing significantly better/poorer or statistically not different from Spain in science (see Table 2).

TABLE 2: TIMSS – Science: Spain vs. other countries

<b>SIGNIFICANTLY HIGHER ACHIEVEMENT:</b>		
Singapore	Austria	Australia
Korea, Rep. of	Hungary	Germany
Czech Republic	Netherlands	Canada
Japan	UK-England	Ireland
Bulgaria	Slovak Republic	Thailand
Slovenia	United States	Sweden
Belgium (Fl)		
<b>NO SIGNIFICANT DIFFERENCE:</b>		
Hong Kong	Norway	New Zealand
Russian Federation	Switzerland	UK-Scotland
<b>SIGNIFICANTLY LOWER ACHIEVEMENT:</b>		
Iceland	Denmark	Cyprus
Romania	Iran, Islamic Rep. of	Lithuania
France	Latvia (LSS)	Colombia
Greece	Portugal	South Africa
Belgium (Fr)		

This type of information tells policy makers in Spain how well their education system is doing in comparison with important reference countries. It shows also that league tables, like Table 1, provide limited information and may result in misleading interpretations, as they do not reflect any statistical information. Figure 1, on the other hand, which does reflect this type of information, shows that not only is Spain not really performing better than Scotland or worse than Switzerland in science, but that European Union (EU) partners like Ireland, Belgium, the Netherlands, Germany and Sweden are performing significantly better. However, this information (in tables and figures like the ones given) does not help policy makers, curriculum developers and educational practitioners to understand why their education system is performing as it does; it can't answer the question of why Spain's performance is poorer than many of its EU partners.

This leads to the second purpose of international comparative achievement studies, namely to attain an understanding of the reasons for observed differences in achievement between education systems. To meet this second purpose, information about learning and teaching processes is needed, and in-depth analysis of achievement results in the context of these background data. IEA international comparative studies collect these different kinds of background data. However, in TIMSS, much of the in-depth analysis still needs to be done. IEA considers this an important task of the

participating countries themselves, as they can state best the research and analysis questions that are relevant to their education system.

## **IEA: What it is and does: mission, history**

TIMSS has been and is still being conducted under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). IEA's mission, as stated before, is to contribute through comparative assessment studies to enhancing the quality of education. IEA achieves this aim by conducting international comparative studies in which educational achievement is assessed in the context of process and input variables.

Over the past thirty-eight years IEA has developed from a co-operative of research institutes to an extensive and valuable network now representing fifty-five education systems.<sup>2</sup> Many of these countries are represented by policy makers in the IEA General Assembly. National Research Co-ordinators and centres for IEA studies are often the most prominent in a country; some are part of various ministries of education, others are linked to universities or are independent research centres. By its nature, IEA provides a network of institutes and individuals, which altogether represent much experience and intellectual capacity. In this way, IEA is a meeting place for policy makers, educators, scientists and researchers.

Over the years IEA has conducted many survey studies on basic school subjects. Most of these studies were curriculum driven, that is the test grid for measuring educational outcomes was based on an analysis of the curriculum of the participating countries. All of these studies included instruments to measure school and classroom process variables, as well as teacher and student background variables. Examples are the studies conducted in mathematics, science, reading literacy, civics education, and English and French as foreign languages.

Next to this, IEA also conducted studies that were not curriculum based, such as the Pre-Primary Project (a study of policies and practices in child care), and the Computers in Education Study, which has been succeeded by the Second Information Technology in Education Study (SITES). SITES started in the last part of 1997 with an indicators module; a limited school survey is planned for November 1998; and two other modules are planned, namely a module of international comparative case studies on innovative practices in the use of information and communication technology and (for the year 2001) a survey of schools, teachers and students.

Even though IEA is conducting several on-going studies, the largest comparative study ever undertaken by the organization has been TIMSS. Some of the results of TIMSS will be discussed in this paper, however, the results of population 3 (end of compulsory education) will only be released in February 1998.

As mentioned above, IEA recognizes two purposes for international comparative achievement studies: (i) to provide policy makers and educational practitioners with information about the quality of their education system in relation to relevant reference groups; and (ii) to assist in understanding the reasons for observed differences

between education systems (which serves policy makers' needs, but is clearly a researchers' interest).

In line with these two purposes, IEA strives in its studies for two kinds of comparisons. The first consists of a straight international comparison of effects of education in terms of scores (or sub-scores) on international tests. This is illustrated for TIMSS in Table 1 and Figure 1. The second concerns how well a country's intended curriculum ('what should be taught in a particular grade') is implemented in the schools and achieved by students. This kind of comparison focuses mainly on national analyses of a country's results.

Since IEA was founded initially as a research co-operative, it was primarily interested in international comparative studies from a research perspective. However, in the second half of the 1980s, IEA started to recognize the increased interests of policy makers in educational indicators. Thus, IEA took it as a challenge to serve the interests of policy makers as well as researchers. The inclusion of IEA achievement indicators in publications of the Organization for Economic Co-operation and Development (OECD) illustrates how IEA has started to become successful in this way. The last edition of the OECD's, *Education at a glance* (1996), presents a number of indicators based on TIMSS results.<sup>3</sup>

Now, not every study should have a size and design as comprehensive as TIMSS, but the conceptualization (which will be summarized in the next section) of this study, and many other IEA studies, allows for designing studies that meet the needs of both policy makers and educational practitioners.

## Functions of IEA studies

As already stated above, IEA studies do more than just make straight comparisons in the form of league tables. The following functions illustrate the importance of international comparative achievement studies (and of educational indicators).

(i) *'Description'*: to provide policy makers and the educational community with information about the status of 'their' education system in an international comparative context: this is considered by many already interesting in itself. Many policy makers have now recognized that this kind of information is a good starting point for generating questions for in-depth analysis ('mirror' function). Our discussion of the results presented in Table 1 and Figure 1 illustrates this. Broad interest worldwide in TIMSS results illustrates the relevance of this function.

(ii) *Benchmarking*: This function can be best illustrated with an example. Within TIMSS, some Asian and European countries (such as Belgium and the Czech Republic) have received the highest test scores for mathematics. If a particular country was interested in improving its education in mathematics, it could analyse its 'own' case against the Asian and/or European countries on many variables related to curricular aspects of mathematics and science education. These variables would include: curricular materials, pedagogical approaches and instructional processes, school

variables, teacher background, teacher training and in-service training, etc. Such analyses may result in proposals for change, although no easy answers could be expected. For such a country, an important question in a next IEA study would be whether it is performing closer to the reference countries chosen.

(iii) *'Monitoring' of quality of education*: One step further than benchmarking is monitoring: the regular assessment of educational processes on different levels in the education system with the purpose of bringing about change when and where needed ('informed decision-making'). This function is an example of assessment-led monitoring of the curriculum (but in the case of IEA studies on the basis of curriculum-based assessment). For this use, trend data are needed, i.e. a cycle of regular assessments in the subject areas which are being monitored (like the IEA and OECD cycle of studies in mathematics, science and reading literacy). For this reason IEA has been asked to repeat TIMSS for the Grade 8 population. Testing will be done in 1998 for the southern hemisphere and in 1999 for the northern hemisphere. Countries that have not participated in TIMSS so far, are welcome to join the study.

(iv) *'Understanding' of reasons for observed differences*: Policy makers may want to understand the differences between or within education systems from the perspective of national policy-making. This function goes one step further than just collecting data for monitoring purposes: ultimately it serves policy makers' needs, but it is also clearly a researchers' interest. An example would be the analysis conducted in the United States of data from the IEA Second International Mathematics Study (SIMS), which resulted in the monograph, *The underachieving curriculum* (McKnight et al., 1989). Again, no easy answers can be expected on what measures should be taken to improve education in a country. But this kind of research may lead to more 'informed decision-making,' and ultimately policy decisions about changes in education, or to initiatives such as the National Council for the Teaching of Mathematics (NCTM), which has developed well-known standards for the teaching of mathematics within the United States.

(v) *'Cross-national research'*: This function refers to exploratory and/or in-depth research of the IEA databases. Many examples of cross-national research can be found in IEA volumes. Here, we are only going to mention two examples: Postlethwaite & Ross (1994), who conducted exploratory research of the IEA Reading Literacy Database (data collection in 1990-91) in an effort to find indicators discriminating between effective and ineffective reading programmes; and the Keeves' (1996) monograph, *The world of school learning: selected key findings from 35 years of IEA research*, in which the author draws ten key findings out of all IEA studies conducted until 1994 and discusses their implications for educational planning.

## What data to collect: some practical and theoretical considerations

The question of what kind of data should be collected in an international comparative assessment study cannot be answered unambiguously. The question is not a trivial one when one realizes that more than twenty countries participate in most IEA studies, and in TIMSS more than forty countries are currently participating. Many participants may differ in what goals they want to achieve from the study. Some may want to emphasize just the description on a small number of indicators, while others strive for a large number of variables (next to achievement) to be able to analyse their country's data properly. Besides, according to its mission, IEA wants to create opportunities for conducting cross-national analysis to enhance the understanding of the functioning of education systems at all levels. On top of this, there is the dilemma between desirability and feasibility: researchers may desire to collect as much data as possible to be able to do in-depth secondary analyses of a rich database, while the usually restricted possibilities to collect data in schools, as well as limited budgets, put severe limitations on the size of the data collections. So, in these type of studies compromises have to be found among the interests of all participating countries. IEA is therefore striving for a design and for instruments that are as 'equally unfair' as possible to all participating countries. Next to this, a well-thought-out conceptual framework addressing the issues is necessary for an effective and efficient study. Almost all of the functions mentioned above need the measuring of educational achievement and other outcomes of education on three levels of the education system, as follows:

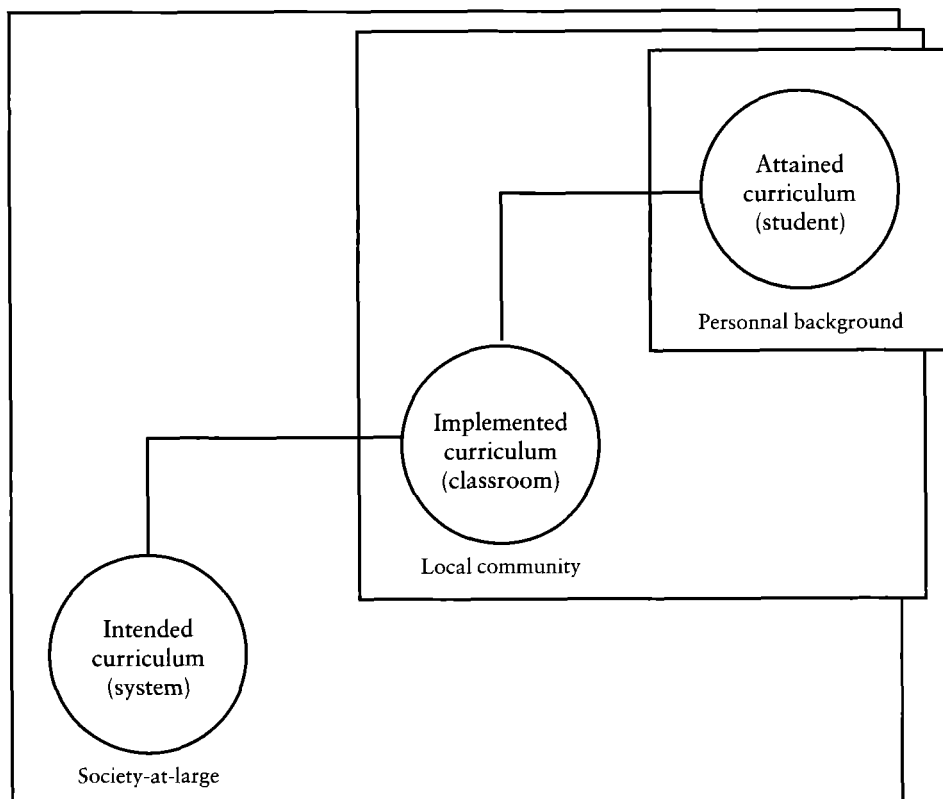
Assessment of:	System level:
• what students learn	micro
• what and how schools and teachers teach	meso
• what the community values (what students should learn)	macro

IEA studies address all three levels by distinguishing among three appearances of the curriculum:

- intended curriculum: what should be taught and learned (measured by analysing documents such as official syllabi, course outlines, text books);
- implemented curriculum: what actually is being taught or taking place in schools and classroom: content, time allocations, instructional strategies, etc. (measured through questionnaires or observations);
- attained curriculum: what students attain or learn in terms of cognitive skills, attitudes, etc. (measured through tests).

In the conceptual model for TIMSS, the variables influencing education are seen as 'situated in a series of embedded contexts starting from the most global and moving to the most personal one', as is illustrated in Figure 2 (Robitaille, 1993; p. 26–27).

FIGURE 2: The conceptual framework for TIMSS (Robitaille, 1993)



For more information about the conceptual approach of IEA, see Robitaille & Garden (1992) or Plomp (1992).

In a typical IEA study, many activities have to be completed to provide data and indicators of good quality, such as: curriculum analysis; instrument development (including pilot testing, translation, etc.); sampling; production of instruments; data collection, cleaning and file building; quality control in participating countries of each component; data analysis; and report writing.

## What data to collect: some examples

Within the practical and theoretical considerations discussed before, the questions of what data should be collected in national and international assessment studies can still be answered in various ways. Again, the answers depend on the functions and research questions the study intends to address. Participating countries may want to use an international comparative study to find answers to some national questions as well. Therefore the 'what data' question has to be answered separately for each study. Here we will present some typical examples of IEA studies.



## DATA FROM WHAT TARGET POPULATIONS?

The choice of the target population(s) is clearly a reflection of the policy or research questions one is interested in. For example, in its cycle of achievement data collections, the OECD will collect data from 15 to 16 year olds, in an attempt to provide policy makers with a baseline profile of the achievement of students at (or close to) the end of compulsory schooling.

On the other hand, TIMSS data have been collected in Grades 3 and 4 (population 1), Grades 7 and 8 (population 2) and the final year of secondary school (population 3), which allows for several comparisons. First, the growth between two adjacent grades can be measured. Second, by including common items on the tests for both populations, growth in mathematics and science from Grade 4 (elementary school) to Grade 8 (junior secondary school) can also be measured. In addition, comparisons between populations 2 and 3 can be made. Moreover, these target populations allow for monitoring the quality of education during compulsory schooling.

## MULTIPLE ASSESSMENT MEASURES

In TIMSS achievement, data have been collected in two ways. First, the achievement tests taken by all students consisted of both open-ended and multiple-choice questions. Second, a sub-sample of students in populations 1 and 2 also completed a series of performance assessment tasks. The performance assessment, which was the same for both populations, was administered in a 'circus' format in which a student completed three to five tasks. The results are reported in Harmon et al. (1997). Table 3 presents some of the results for students in Grade 8 from the achievement testing and performance assessment in combination with achievement results taken from Figure 1 and Table 1.

Table 3 illustrates the 'mirror' function of this descriptive data, which may lead to important questions for policy makers and educational practitioners in many countries. It shows that a number of countries have similar scores for all assessment measures. For example, Singapore is consistently on top and Spain, Portugal and Colombia are consistently below the international average.

Interesting questions can be asked in, for example, the cases of the Netherlands and the Czech Republic. Both countries score well above the international average on the mathematics and science achievement tests, but only close to the international average on performance tasks. If one values the capability of pupils to solve problems and perform tasks, then the satisfaction in these two countries about their high scores on the achievements tests should not overshadow the concerns they may have with their average results on the performance test.

Some countries however, do have one result that deviates from a pattern. Take Switzerland, for example. Switzerland is scoring very well for performance tasks and mathematical achievement, but only average for science achievement. These examples illustrate that analysing the descriptive results on multiple assessment measures allows countries to state questions that may lead to further, in-depth analyses and/or to discussions about the emphasis and focus of the curriculum.

TABLE 3: TIMSS grade 8: Achievement and performance scores for mathematics and science

Mathematics				Science			
Achievement test (scale pts)		Performance tasks (av. %)		Achievement test (scale pts)		Performance tasks (av. %)	
Singapore	643	Singapore	70	Singapore	607	Singapore	72
Czech Republic	564	Switzerland	66	Czech Republic	574	UK-England	71
Switzerland	545	Australia	66	Netherlands	560	Switzerland	65
Netherlands	541	Romania	66	Slovenia	560	UK-Scotland	64
Slovenia	541	Sweden	65	UK-England	552	Sweden	63
Australia	530	Norway	65	Australia	545	Australia	63
Canada	527	UK-England	64	Sweden	535	Czech Republic	60
Sweden	519	Slovenia	64	USA	534	Canada	59
New Zealand	508	Czech Republic	62	Canada	531	Norway	58
UK-England	506	Canada	62	Norway	527	New Zealand	58
Norway	503	New Zealand	62	New Zealand	525	Netherlands	58
USA	502	Netherlands	62	Switzerland	522	Slovenia	58
UK-Scotland	498	UK-Scotland	61	UK-Scotland	517	Romania	57
Spain	487	Iran, I.R.	54	Spain	517	USA	55
Romania	482	USA	54	Romania	486	Spain	56
Cyprus	474	Spain	52	Portugal	480	Iran	50
Portugal	454	Portugal	48	Iran	470	Cyprus	49
Iran	428	Cyprus	44	Cyprus	463	Portugal	47
Colombia	385	Colombia	37	Colombia	411	Colombia	42
Intl. average	513	Intl. average	59	Intl. average	516	Intl. average	58

Sources: Beaton et al. (1996a, b), Harmon et al. (1997)

## BACKGROUND DATA

Background data are always collected in IEA studies (see Figure 2). Such data allow one to address research questions, such as what factors are contributing to good quality education. Another reason to collect such data is that it allows countries to search for determinants of national results in an international context. In the IEA Reading Literacy Study, Postlethwaite & Ross (1994) concluded that a large number of background variables were influencing reading achievement. These were divided into several categories, namely indicators of:

- student activities at home;
- school context;
- school characteristics;
- school resources;
- school initiatives;

- school management and development;
- teacher characteristics;
- classroom conditions, of teacher activities;
- teaching methods.

Postlethwaite & Ross (1994) analysed cross nationally these indicators in light of the question of what makes a school effective in teaching reading. They found that, to increase student reading performance, voluntary out-of-school reading should be fostered among students (particularly during the primary school years); schools should have classroom and/or school libraries; and teachers should emphasize reading for comprehension.

In general, the accumulated experiences gained in IEA studies in combination with the questions to be addressed in a study determine to a large extent what background data should be collected from schools, teachers and pupils.

#### NEED FOR NATIONAL ASSESSMENT

International comparative studies can be used by a country to study its own educational practice in an international comparative context. In the case of Switzerland, Moser (1997) analysed in TIMSS the extent to which instructional practices (child-oriented *versus* subject-oriented instruction) and instructional variables (autonomy of students in child-oriented classes *versus* on-task behaviour in subject-oriented classes) were influencing learning outcomes, not only in mathematics achievement but also in internal activity, self-activity and interest in mathematics. He concluded that instructional practices and instructional variables do not have a significant effect on mathematics achievement, but many effects on other learning outcomes. In light of the much better results in Japan (a country with a high emphasis on subject-matter instructional practices and on on-task behaviour), he concludes that instructional practices in Switzerland can improve on these aspects.

Another example of a national analysis from Switzerland is related to our earlier conclusion that in TIMSS Switzerland is doing quite well on the performance tasks and on mathematical achievement, but only average on science achievement. Ramseier (1997) analysed possible causes for this and concluded that this can be explained from a discrepancy between the Swiss science curriculum (teaching priorities) and the science part of the international achievement test. Most international comparative studies allow for a limited number of national questions ('national option'). The example of Switzerland illustrates how important it is that countries participating in international comparative studies think beforehand about the national (policy and/or research) questions they want to address through such a study; they also need to identify typical characteristics of their national system for inclusion in the background questionnaires, in order to allow for relevant national analyses.

### Concluding remarks

First of all, the relevance of participating in an international comparative study increases for a country if important reference countries also participate. For that

reason, a study like TIMSS has great relevance for the European Union, North America and a number of Asian countries. To increase the relevance of IEA studies for the Latin American region, more countries should be involved. It is a promising sign that Brazil and Chile recently joined the IEA. The relevance of the participation of these countries in the repeat of TIMSS in 1998 and 1999 will increase if more Latin American countries join the study.

IEA-types of studies are logistically and methodologically complex. An important feature of IEA studies is the training of National Research Co-ordinators (NRCs). Training is an essential component of the study, as many researchers appear to be unfamiliar with the methodology, especially the specifics of international comparative studies.

A not-so-tangible benefit of participating in such studies is the development of a network of researchers and specialists (in sampling, psychometrics, test development, data analysis, etc.), which can later be tapped when countries begin developing their own evaluation and national assessment studies.

An aspect that is often overlooked is the possibility of attuning national assessments to international assessments. Proper linking of the two will not only increase the benefits a country can get from investments in assessment studies, but is also cost efficient.

Another cost aspect is related to the question of what data should be collected. As was illustrated earlier, policy and research questions should determine primarily what data should be collected. On the other hand, when cost factors come in and influence what data will be (or will not be) collected, one runs the risk of limited usability of the data collected. If IEA had collected only achievement data in TIMSS, but no data about schools, teachers and students, a country like Switzerland would never have been able to conduct national analyses in an international context and would have missed a unique opportunity to address some important national questions. It is often only a small increase in cost which makes the difference between collecting just achievement data or getting a rich data-set which allows for in-depth analyses of important issues.

## Notes

1. See references for a full list of publications from this study.
2. See Husén & Postlethwaite, 1996, for a concise description of the history of the IEA.
3. Examples of IEA publications which are addressing relevant policy questions are Postlethwaite & Ross (1994) and Keeves (1996); another relevant source is Kellaghan (1996).

## References

- Beaton, A.E., et al. 1996a. *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA, Center for the Study of Testing, Evaluation and Educational Policy, Boston College.
- , et al. 1996b. *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA, Center for the Study of Testing, Evaluation and Educational Policy, Boston College.

- Harmon, M., et al. 1997. *Performance assessment IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA, Center for the Study of Testing, Evaluation and Educational Policy, Boston College.
- Husén, T.; Postlethwaite, T.N. 1996. A brief history of the International Association for the Evaluation of Educational Achievement (IEA). *Assessment in education* (Abingdon, UK), vol. 3, no. 2, p. 129–141.
- Keeves, J. 1996. *The world of school learning: selected findings from 35 years of IEA research*. Amsterdam, IEA.
- Kellaghan, T. 1996. IEA studies and educational policy. *Assessment in education* (Abingdon, UK), vol. 3, no. 2, p. 143–160.
- Martin, M.O., et al. 1997. *Science achievement in the primary school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA, Center for the Study of Testing, Evaluation and Educational Policy, Boston College.
- McKnight, C.C., et al. 1989. *The underachieving curriculum: assessing U.S. school mathematics from an international perspective*. Champaign, IL, Stipes Publishing Company.
- Moser, U.P. 1997. *Annual conference of the American Educational Research Association, Swiss analysis of the TIMSS data*. Chicago, IL.
- Mullis, I.V.S., et al. 1997. *Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA, Center for the Study of Testing, Evaluation and Educational Policy, Boston College.
- OECD. 1996. *Education at a glance: OECD indicators*. 4<sup>th</sup> ed. Paris, OECD/CERI.
- Plomp, T. 1992. Conceptualizing a comparative educational research framework. *Prospects* (UNESCO/IBE, Paris), vol. 22, no. 3, p. 278–288.
- Postlethwaite, T.N.; Ross, K. 1994. *Effective schools in reading: implications for educational planners*. Amsterdam, IEA.
- Ramseier, E. 1997. *European Conference on Educational Research: task characteristics and task difficulty: analysis of typical features in the Swiss performance in TIMSS*. Frankfurt.
- Robitaille, D.F., ed. 1993. *Curriculum frameworks for mathematics and science*. Vancouver, Pacific Educational Press. (TIMSS monograph, no. 1.)
- Robitaille, D.F.; Garden, R., eds. 1989. *The IEA study of mathematics II: contexts and outcomes of school mathematics*. Oxford, Pergamon Press.

---

# THE OECD EDUCATIONAL INDICATORS: PURPOSES, LIMITS AND PRODUCTION PROCESSES<sup>1</sup>

---

*Norberto Bottani*

---

## **Purposes of the OECD educational indicators set**

The methodology of educational indicators is now widely used for evaluating national education systems. Hundred of indicators can be generated, but it is meaningless to produce a large amount of indicators. The abundance of information and measures do not generate better knowledge. Moreover, a large set of indicators is very difficult to manage. Organization for Economic Co-operation and Development (OECD) countries have been trying since 1991 to generate a small set of educational indicators and to identify core indicators that provide essential information about education systems. Despite these efforts, the OECD's set of educational indicators has not evolved very much. The number of indicators included in the set has remained stable, around fifty. It is difficult to identify criteria for selecting indicators. This operation requires a conceptual framework reflecting the aims of the evaluation and of the education system.

---

*Original language: English*

*Norberto Bottani (Switzerland)*

Director of the Unit for Educational Research of the Canton of Geneva. From 1976 to 1997, Principal Administrator of the Centre for Educational Research and Innovation at the Organization for Economic Co-operation and Development. Editor of OECD reports and books on early childhood education, multicultural education and educational indicators. Author of many articles published in American, British, French, German and Italian educational journals. E-mail: Bottani@uni2a.unige.ch

## **The conceptual framework of the OECD educational indicators set**

The scheme by which the OECD educational indicators set is organized was developed over several stages. Dissatisfaction with the prevailing organizing models of the 1980s, with their emphasis on causality and lack of interpretative power, invited the adoption of a flexible organizing framework. The International Indicators of Education Systems (INES) conceptual framework is organized around a combined approach using both conceptual and pragmatic bases and incorporating policy concerns.

The conceptual and theoretical parameters explaining the choices that have influenced the construction of the indicators set have been described in two publications: *The OECD international education indicators: a framework for analysis* (1992b) and *Making education count. Developing and using international indicators* (1994).

The ambition of the OECD in proposing a set of international educational indicators was to cover a wide spectrum of educational services, to broaden the horizons beyond the world of formal school settings, and hence to include data on private education (including non-profit and commercial forms as well), on training activities run in firms, on household purchases of educational goods and services, on R&D in education, and on educational staff—not only on teachers.

The conceptual framework leading the selection of the indicators has a crucial function in the implementation of an information strategy designed to provide the foundation for policy decisions and sound management about which data collection and reporting initiatives should be pursued. As François Orivel (see note 1) rightly indicated in his analysis of the first three OECD sets of educational indicators, the factors which influence students' attainments are very numerous and their impact varies depending on whether certain other factors are present or not. Furthermore, some factors are malleable (i.e. they can be changed by outside intervention), while others cannot be changed or society refuses to change them. For example, class size or the duration of teacher training can be modified; natural aptitudes, family or geographical environment or peer acceptance cannot be modified. It is virtually impossible to measure all of the factors which might have an impact. For this reason, despite the relevance of the research on education, there is no agreement on the most influential factors concerning students' performance. Results which have a high degree of consistency are relatively rare. Therefore, when selecting indicators able to correctly and quickly inform users of education systems about the performance and the state of education, certain choices must be made. These choices are neither simple nor neutral.

The OECD uses macro-data for building its educational indicators. The nature of the data used helps to select among the different factors which shall be retained. Nevertheless, the OECD's aggregate indicators are usually too general to allow detailed analysis of the relationship between resources invested and results obtained.

However, their usefulness lies elsewhere: they help to raise certain relevant questions at the macro-economic level. Orivel demonstrated this with one of the most familiar composite indicators, educational expenditure relative to Gross Domestic Product (GDP). Since 1975, there has been a certain convergence among decision makers to allocate a similar average of GDP to education, around 6%. 'The range of this indicator has gradually shrunk. The countries which allocated over 6% of the GDP have tended to stabilize or cut back this figure, while those that allocated less than 4% of their GDP to education have sought to reach the median figure of 5 or 5.5%.' Orivel said that 'this trend towards a certain convergence is probably not a coincidence, although we do not have explicit statements by decision makers on this point. It is tempting however to hypothesize that the lack of solid formal criteria on which the public authorities can base a specific level of commitment to education, encourages decision makers not to allocate significantly more or less than the average.' Orivel added that 'this is the kind of argument often used in making public decisions on the financing of R&D and national defence, based on the logic that when partners are doing more, countries must try to catch up with them so as not to fall behind or to reduce the likelihood of steering the wrong course.' Orivel concluded that 'in many areas of educational policy, the factors on which policy choices are based are ill-defined or contradictory, and it can be helpful to decision makers to look at what other comparable countries are doing, whether because the vast majority of countries have chosen a given solution, or because a country which has chosen an original approach seems to be obtaining favourable results.' International competition and cross-national comparisons focused attention on indicators and were paramount for the development of an international set of educational indicators.

For these reasons, the OECD decided to adopt a very simple conceptual framework which was a relatively sparse variation of an input/output model. This model was and is largely adequate to produce information useful to a policy audience. Virtually every significant effort to conceive educational indicators has used an input/output or processed model. In order to produce policy-relevant variables that can be clearly and succinctly summarized, presented and considered, it was not necessary to adopt a complex and research-oriented model. The OECD therefore used a three-part framework to organize the indicators providing information on educational outcomes, on the characteristics or features of the school system in each country, and on the social and economic background within which the education system operates.

The indicators were deliberately not organized according to a 'model'. A model would imply a sophisticated, casual connection among indicators; it is not reasonable to assume that the variables represented by these indicators function in these ways. From the very beginning of the INES project there was consensus about the fact that the organization and the selection of the data was not compatible with a causal model. This concept was felt as hardly productive, given the lack of a generally accepted representation of relationships within education.

But, in order to be useful and practical, the organizing framework must be



simple, global and parsimonious. The framework for an international set of educational indicators cannot pretend to represent all of the relationships between elements in the system and still be workable. Yet the set must be faithful, to some extent, to the complexity of education systems.

The aim of the framework was not to offer another comprehensive model of education but rather to provide a way to contextualize the indicators and the links that exist between them. Though the framework does not aim to build links of a causal nature, it helps to put into perspective some of the interactions that take place in the functioning of education systems.

There are various ways for organizing indicators. They can be organized conceptually, reflecting a working model of the system they are intended to describe. They can be organized pragmatically, like the gauges in a locomotive, to inform a set of specific, practical planning decisions, such as hiring staff where they are needed, allocating space or addressing specific achievement problems. They can be organized around policy issues, like equity, acquisition of learning, contribution to economic productivity and so on. The OECD adopted a pragmatic-conceptual approach for organizing the indicators. The OECD set is a combined approach using both conceptual and pragmatic bases and incorporating policy concerns. An approach of this nature seemed to be the most appropriate way to organize comparative international educational data for a number of reasons. First, these indicators would be used less for narrow, immediate management purposes and more for global analysis of general trends and differences among countries. Second, the issues important to various countries will vary somewhat, defeating an issues orientation. Third, it encompasses the different intellectual traditions of OECD member countries and draws strength from the synergy of the combined approaches. Finally, a systematic, conceptual view of education has emerged that seems to have general acceptance and validity.

This mixed nature had important implications for the kinds of indicators that have been suggested. Had the approach been strictly conceptual, the indicators would have been conceived for providing information on the key elements designated by the logic or by the research as basic components of a full, working model of the education system. In a purely pragmatic approach, those indicators that experience and political priorities suggest are of interest might have been listed eclectically. Instead, and following from the mixed basis, some of the indicators have been selected because they were logical and empirical in nature while others were practical and policy-sensitive in their orientation.

Using a combined approach should provide some in-built balance between stability and flexibility in a set of indicators that should evolve over time. Utilizing a conceptual model to design an indicator set would have more than likely generated a relatively rigid product that risked not being attuned to present concerns. Moreover, a lot of time would have been spent in reaching agreement on the model without producing any useful information and without any improvement in the data gathering strategy. A policy-issue approach would have had similar consequences from the operational point of view (it is necessary to remember that there was a large consensus

about the lack of validity of the data so far gathered) and would have produced highly volatile sets of indicators as priorities come and go.

The domains included in the OECD indicator set encompass the important information that can be displayed about the nations' education systems: the status of their accomplishments, their features and characteristics, and the conditions under which they operate. For analytical purposes, this framework represents education as a multi-component process. It is important to note that many of the educational elements in the framework interact in a fairly complex fashion. In statistical terms, the organizing framework is not intended to be a simple regression model. Many of the elements in it are parallel to one another, are recursive, and operate at several levels: at the level of the country as a whole, at the school level, and at the level of organizing units in between.

In the framework, general demographic and economic conditions are seen as background factors that determine the job schools must do in each country and that determine the resources available to fund school programmes. More affluent systems can support more elaborate programmes. At the same time, lower levels of education in the past, or greater levels of poverty now, compound the task of providing quality education in many countries.

Educational programmes and processes are seen as the input variables determining the success of the school programme. They are the features by which educational programmes in various countries may vary and which may explain, or at least be associated with, educational outcomes.

Outcomes are seen in part as reflecting general social, economic and historical conditions affecting education in each country, and partly as reflecting the combined effects of the policies, programmes, practices and educational decisions constituting schooling in each country.

It is imperative to bear in mind the recursiveness and multi-level nature of this framework. Among many of the indicators in this framework, relationships could operate in various directions at the same time. For example, student and staff attitudes might improve as outcomes improve. Public support could increase as a result of greater programme success. Simply improving the completion rate through secondary schooling will improve the educational background of the general population in a country. Resources available will be related partly, but not entirely, to resources expended. These are examples of some of the interactions and feedback loops that may operate in this framework. They also illustrate and explain why a general framework was adopted, as opposed to attempting to determine and work out all of these interrelationships.

Second, the framework has a multi-level nature. In each element, variability within the system may be as important as variability between countries. This will affect interpretation, as global comparisons between countries may become irrelevant if wide disparities exist among schools or provinces within a country. Yet disparities are significant in that they may reflect systematic problems, such as unequal services or distribution of resources, that should be addressed.

As stated before, the major purpose of educational indicators is not to be an

explication of education as a process or verification of cause-and-effect relationships in the system. If this is true, then how and to what purpose should indicators, and specifically international indicators, be used? The best answer is that they are partial and preliminary information about the functioning of an education system. They should lead to further investigation or consideration of why the system is functioning as it is, thereby requiring additional information and analysis.

The conceptual and organizational framework adopted in 1991 provided a frame for producing four sets of indicators, but it was not appropriate for stimulating the production of new complex or compound indicators. Member countries had requested that more synthetic and analytical use be made of the information and indicated that the selection of indicators should be driven less by the availability of data, than by the conceptual and policy framework.. In other terms, the practical-empirical approach has been more influential on the construction of the four first sets of indicators than on the conceptual-policy approach. The existing framework is outdated: it is not effective for piloting and controlling indicator development nor does it provide analytical support for the interpretation of the data. Changes adopted in the fourth set clearly demonstrate the need for a new framework and a revision of the original one.

## **Limits of the OECD educational indicators set**

### THE CONTEXT OF EDUCATION

Reviewing the sets so far published, it appears that the domain 'context of education' is the domain which has changed the most over the four editions of *Education at a glance*. It grew from five to thirteen indicators and went down to four indicators in the 1996 edition, the most significant change being the introduction in the third edition of results from a survey on public opinions and expectations regarding national education systems (OECD, 1995*b*; INES, 1995). Nevertheless, it is necessary to say that indicators of attitudes and expectations will not be generated every year, because the OECD does not have the resources to implement an annual public opinion poll. Therefore, one could expect for the next few years a high variability of indicators in this domain.

In the first edition of the OECD indicators set, there are two common sub-groups of indicators in this domain, i.e. demographic indicators and economic indicators. Both of these sections are quite poor. They provide only very general information, mostly focused on the educational attainment of the population, gender differences in education, labour force participation and education, and the national income per capita. This data shows the proportion of a country's wealth which is devoted to education and provides an interesting first look at the relation between the rank by the total educational investment and the rank by GDP per capita. Some countries like Finland and Canada spend considerably more than countries with a higher GDP per capita, such as the United States. In particular it is interesting to note that some of the less affluent countries (like Ireland and Spain) spend a greater

percentage of their per capita GDP to achieve the same level of spending on education as some of the most prosperous countries (like the Netherlands, Japan, Germany, Australia and France). If we consider that these countries must also spend a greater percentage on other basics, and that the least prosperous countries must educate the largest number of students because the size of the young population in these countries is higher than elsewhere (for example, 5–29 year olds represent one-half of Turkey's total population and less than one-third of Germany's), we have an interesting frame of reference for assessing educational policies.

The educational context indicators produced by the OECD never included an indicator on productivity—productivity being defined as the GDP per employed person. In other words, productivity is a measure of the average productive capacity of a country's employees. Countries with higher levels of productivity have a larger economic capacity from which to invest in socio-economic infrastructure, improve education and raise their standard of living. Between 1961 and 1991, the United States experienced the lowest average annual increase in productivity of the G-7 countries (1.05%). Japan achieved the highest productivity gains of the remaining countries, with an average annual increase of 4.8% (NCES, 1996).

The weakest element and unresolved problem of this section is the lack of data on health, housing, cultural and social statistics. This information is crucial for drawing a clearer picture of the educational context and the conditions of functioning.

#### COSTS, RESOURCES AND SCHOOL PROCESSES

The domain 'costs, resources and school processes' is the most stable and developed. The number of indicators in this domain has varied from twenty-one to twenty-six—a peak reached in 1995 when the OECD calculated two indicators on educational R&D. These disappeared in the 1996 edition and several indicators on human resources subsequently disappeared in the following edition.

The strong financial section was substantially reorganized between 1992 and 1996. Thanks to the work carried out to improve, in terms of comparability, the quality of the definitions and measurements of financial resources, the expenditure categories have been revised. A clear distinction has been made between educational expenditures or costs and sources of educational funds. The implementation of new questionnaires on finance, with their accompanying definitions and instructions, allowed for the implementation of explicit categories of transfer payments and subsidies. This change made it possible to trace the flow of educational funds from initial sources to ultimate users and to calculate both initial (before transfer) and final (after transfer) distributions of educational funding by source.

Moreover, the new instruments separate expenditures by or for educational institutions from subsidies to students. This distinction is especially relevant at the level of tertiary education. The total funds used for subsidizing the living costs of students is very impressive but it is not possible, for the moment, to obtain adequate figures or even acceptable estimations. There are multiple sources of funds, not only by levels of government, but by types of administration. In France, for example, a

1995 estimate was that the total amount of financial aid for student living expenses (subsidies for room, board and other living expenses) was around 24 billion French francs. Although excluding these funds from the indicators calculation helps to avoid major distortions, it is not an acceptable solution. Student scholarships and other subsidies are an important part of some countries' educational budgets and simply cannot be ignored. The actual educational expenditure figures are therefore underestimated and the rank of countries in terms of educational expenditures could change dramatically if one includes student subsidies in the total. Finally, in an effort to reduce the number of OECD indicators while identifying those that are 'a must', the 1996 edition of *Education at a glance* further simplified the presentation of indicators on costs and published only five financial indicators.

#### PARTICIPATION

The section on 'participation' comprised six indicators in the first edition, seven in the two following editions and five in the 1996 edition. In this last edition, two indicators disappeared: the indicator on transition characteristics from secondary to tertiary—but the data has been included in the indicators on participation in education—and the indicator on access to tertiary education.

Indicators in this section have been constantly refined and improved. One example is the indicator on participation in formal education relative to the population aged 5–29 which provides, for the first time, information about the percentage of the young population attending full-time formal education. In the third and fourth editions of *Education at a glance*, the indicator includes a table on the average schooling expectancy, providing information about the hypothetical duration of schooling for a 5-year-old child. The average expectancy is about 14–15 years. It is lowest—13 years or less—in the Eastern countries and in Turkey. It is highest, 15.5 years or more, in Belgium, Denmark, France, Germany and the Netherlands. It is interesting to realize that a 5-year-old student in one of these countries has an average of fifteen years of schooling ahead. The student will not leave the formal education system before the age of 20. School expectancy has increased in recent years in almost all countries for which data are available. The increase since 1985 exceeds more than one year in many countries, and in the Nordic countries it is more than eighteen months. The longer duration of schooling is another factor that contributed to the observed rise in enrolment rates over recent decades.

Two other new approaches developed in the 1995 edition are: the calculation of the ratio of enrolled students in second education programmes at the upper secondary level; and the calculation of the indicator on the transition from childhood education to primary education for pupils between 4–7 years of age. But the significant new indicator in this section is that on adult enrolments in job-related continuing education and training which announces future developments in the area of adult education and lifelong learning indicators.

## STAFF

The indicators on 'staff' have been enriched substantially between the second and third edition thanks to the implementation of a survey on teachers which has not been duplicated in the following years. As a result, this section—as the section on the context of education—is quite unstable over time. In 1996, it was not possible to calculate some of the indicators on teachers published in 1995. The OECD's indicators continue to be quite weak in this area. The most serious limitation of these indicators is the exclusion of data about the staff employed in tertiary education. Two indicators did not change in any of the editions: the indicator on staff employed in education, and the indicator on the ratio of students to teaching staff.

The most original and innovative section devoted to school environment and school/classroom processes was included for the first time in the 1995 edition and amplified in the 1996 edition, with an increase in indicators from seven to ten. Despite this significant evolution, this section has not yet been fully consolidated because data sources that the OECD can use for calculating indicators about the nature, effectiveness and quality of educational processes and schools are quite erratic. Because of the diverse and complex nature of the activities within schools, many of these questions are not as easily addressed as enrolments, expenditure or examination results. Nonetheless, a certain number of national, regional, school and classroom-level characteristics can be assessed, using data reported by those involved or data drawn from policy statements. The effectiveness of schooling is not only reflected in curricular variables but also in the instructional environment which schools provide and in the importance attached to education outside school. The 1995 attempt to use International Association for the Evaluation of Educational Achievement (IEA) data from the background questionnaires of the Reading Literacy Study (RLS) to calculate indicators on school practices was not very successful. The IEA data have been collected to serve as an aid in interpreting students' reading scores; they were not appropriate for constructing general indicators on school practices. Only the indicators on hours of instruction and on grouping within classes have been calculated using IEA data from the RLS. To gather adequate information on staff, teaching time, teachers' salaries, curriculum and school processes, the OECD has been obliged to conceive and implement a specific survey on teachers and schools. An international survey of schools administered during the school year 1995–96 and based upon a probability sample of approximately 400 primary schools per country (which were selected with probabilities proportional to their size) allowed the presentation in 1996 of seven process indicators related to primary education. The primary schools have been defined by the national institutional structure of the education system, irrespective of the ages and grades of the students enrolled. The rationale behind this choice was that having information which corresponds to actual school situations would offset any loss of international comparison due to differences in ages and grades. Results have been weighted by the number of students enrolled. These seven indicators examine:

- stability in educational staff (as a proxy indicator of consistent education);
- school leadership (with a special focus on the part played by educational versus administrative leadership);
- co-operation between staff, and among head teachers and staff;
- monitoring and evaluation procedures to ascertain the extent to which the goals and intentions have been achieved;
- the ways in which variation in student achievement is addressed by schools and how schools respond to different needs of students;
- the achievement orientation of school policies (relating to the setting and monitoring of standards both at the student and the school levels); and
- the extent to which parents are informed on various school matters, particularly on the performance of their children, as well as the extent to which parents are involved in instruction or contribute directly to decision-making in schools.

These indicators illustrate differences in educational practices which can be influenced by educational policy. They can act as both benchmarks and as examples of good practices, and are an important element in judging the performance of schools. They should be interpreted in the light of the national educational context.

## RESULTS OF EDUCATION

The section of indicators on the 'results of education' is both the section that has changed the least in terms of the number of indicators (from ten indicators to twelve), and the one that has changed the most in terms of content. In the four editions, there are three common sub-groups, i.e. student outcomes indicators; system outcomes indicators and labour market outcomes. The most unstable sub-group is that on student outcomes. These indicators have changed in every edition, trying to use data provided by two non-governmental sources: the IEA and the Educational Testing Service. The indicators are focused on achievement in mathematics, sciences and reading. Sometimes the calculations concern the population of 9 year olds; but are usually for 14 year olds. It is evident that the OECD did not find the appropriate data for producing a stable set of achievement indicators which provide the information expected by public opinion, users of education systems and policy makers. In the third edition of *Education at a glance*, the OECD was obliged to use, for the second time, data from the RLS carried out in 1991. In the second edition, RLS data was used for calculating the three standard indicators on student achievement: multiple comparisons, student scores distribution, and within and between school variances. In the third edition, the same set of data was used for calculating an indicator on the progress in reading between the ages of 9–14. The background questionnaires of the IEA survey were used for calculating a second indicator on the amount of reading, based upon the frequency of reading reported by 14 year olds. Despite the interest of these indicators and the rationale for calculating them, one has the impression that these indicators have been invented to avoid leaving cells empty in the sub-group on student achievement. The two other sub-groups of this domain—system outcomes indicators and educational outcomes at

the labour market level—are more stable because the data sources are regular. For the system outcomes indicators, the data come from the joint UOE questionnaires which are directly managed by the OECD; for the labour market outcomes, indicators the data sources are the *Labour force surveys*. If the OECD would like to produce policy-relevant and up-to-date indicators on student achievement, it will be necessary to revise the data strategy in this domain.

## Production process

The challenging task of building an OECD set of educational indicators forced member countries and the OECD Secretariat to innovate methods of work and to implement a new form of international co-operation. The main characteristics of the methodology implemented for building an international network of data providers and users, and for generating a consensual set of international indicators on education, have been the following:

- constant feedback between countries and the OECD Secretariat;
- periodical evaluation of the progress of work;
- collective engagement in identifying common concern and preparing common tools; and
- constant communication among all the actors (data producers, data providers and other people involved) about objectives, tasks and priorities to achieve.

Considering the hostility of many educationalists and key decision makers towards the development of educational indicators when the activity started in 1988, it was necessary to invent a new style of co-operation between member countries, able to generate mutual trust and consensus among all participants.

The development of a set of indicators at the international level is a complex, multi-stage operation. Many actors are involved, having various degrees of accountability. Moreover, data gathering in education is not easy: some information is rather simple, some consumes a lot of resources, some may be considered ethically indefensible. It is therefore essential to reach an agreement between the data producers and data providers about:

- the conceptual and policy relevance of the data to collect;
- the proper way to measure indicative events, features or performances in education;
- the data collection conventions;
- the publication and reporting procedures; and
- the financial resources available for covering the costs of data gathering and processing.

Four main reasons concur to explain the relevance of these components.

*There are many actors involved*, spread over a number of administrations and engaged in different roles: researchers in different disciplines, policy makers, administrators, statisticians, etc. All these interlocutors belong to various cultures, they perceive problems in different ways, practise different languages and have different scientific backgrounds. Effective co-operation among them cannot be



conceived without special organizational arrangements, without the development of a common language or without a set of common tools.

*The peculiarity of the educational issues.* Education is a very sensitive domain: it concerns the survival of a community, the well-being of a country and the capacity of a nation to generate knowledge, to understand its historical and political context, and to produce and transmit an image of its identity. Therefore, information about nature and procedures of education is peculiar, highly political and culturally protected. The construction of an international set of comparative data about education modifies traditional perspectives about the uniqueness of local, regional or national education systems. Member countries hardly agree to dismantle these beliefs and to openly co-operate to build a trans-national representation of education which invariably demonstrates the homogeneity of educational development and the relative weight of the national and local contexts. This process can be envisaged only on the basis of strong mutual trust among data providers, data producers and data analysts.

*The nature of the work* implies exchange of information and communication: in gathering data, the OECD manipulates statistical information—data involving thousand of people, their occupations, their institutions, their earnings, etc. It is not possible to do this work without legitimacy which can derive only from common concern and agreement about the rules of respect for confidentiality and privacy.

*The political implications of the work.* The production of a set of international educational indicators is not only statistical work or research, but essentially a political task which concerns not a limited number of specialists, but a large number of decision makers and users. Therefore, the international statistical programme in education cannot be managed by a restricted number of professionals. Fundamental decisions about the data gathering programme, data processing and data reporting have to be decided collectively by all the categories of actors which are concerned by the implementation of a democratic information strategy on education.

These four reasons constitute the rationale for implementing new forms of co-operation in the domain of educational statistics and indicators. In order to achieve this aim, the OECD implemented a system of networks. Considering that networking was instrumental for operationalizing mutual trust and for generating a consensus between different actors, the OECD member countries agreed in 1988 to implement a new process of work based upon networks.

In the INES project, a network is a flexible and relatively informal structure, recruiting members on a voluntary basis without strong linkages to power and decision structures. Networks have been created for developing and preparing clusters of indicators identified as relevant by the member countries. A leading country assumes the responsibility of organizing the activity. The network approach has proven to be very effective. It has been a powerful way of mobilizing specialists and for generating indicators on areas scarcely explored: such as student achievement, labour market destinations by educational level, the qualification of educational staff, the expectations of public opinion towards education, the loci of decision-making at various levels, etc.

The idea of networking has recently become fashionable in both the public and

the private sectors. But if the purposes of networks are ill-defined, they may lack direction, commitment from their members and impact. The INES networks, in contrast, have been assigned the tangible objective of producing useable indicators in their particular domains. Even though this objective has sometimes been elusive, they have developed strong working methods in an attempt to reach them.

The INES network is worth looking at as an innovative methodology that has achieved a number of things that were unlikely to be achieved through other methods. A number of distinctive aspects are particularly worth considering in relation to further developments at the national and international levels:

- The use of networks to develop standardized definitions for comparing nationally collected data. Networks have managed to find solutions to long-standing comparison problems, for example on educational spending and on pupil-teacher ratios.
- The use of networks to identify strategic areas where new data is needed. Although it has taken longer to come to grips with this task, networks have created international data collection instruments, whose agreed relevance among countries participating in the network is reflected in their commitment to collect the required data.
- The management of a large number of participants in the developmental process. The high level of participation is sustained partly by the semi-autonomous nature of the networks, and by the dialogue that it creates among their members independently of the international Secretariat. Networking is effective not just in terms of indicator production, but also in terms of international understanding among participants.
- The mechanics of networking, involving not only regular working meetings but also a constant feedback between the networks' members and between the networks and the international Secretariat.

Networks resemble clubs—a set of people who meet regularly and acquire a 'group identity' find it easier to work together than strangers. This makes them more effective, but also potentially leads to decisions taken in the interests of the club that take insufficient account of wider interests and priorities.

At the core of a network's activities are its regular meetings as a group. These meetings take place about twice a year, and usually last two or three days. An important aspect of network meetings is their difference in style to most international meetings. Rather than behaving as 'delegates' merely stating the view of each country, the members tend to work collaboratively towards common goals. The often elusive task of defining and collecting internationally comparable indicators in the educational field is seen to demand intensive brainstorming rather than diplomatic posturing. The informal style that has developed is reinforced by socializing and discussions that take place outside the sessions, among a relatively stable set of members. This pragmatic and productive method of working has not always been present in all networks. Much has depended on the character of the chairman, and on the progressive development of familiarity and trust among members as they get to know each other.

It is doubtful whether the networks would have developed this team ethos if they did nothing but meet twice a year. But the nature of their tasks, often requiring an iterative effort of exploring various possible ways of defining and producing indicators, demands continuous communication, as proposals are circulated and commented on. The main result of this intensive feedback was a unique sense of ownership. It is rare to achieve a result like this at the international level where usually groups are loosely coupled.

Networking constituted a major feature of the INES project and was paramount for the development of INES. Networking is the best procedure to :

- get frequent, if not immediate, information about basic goals to create constancy of purpose throughout the project;
- generate frequent feedback for achieving long-term goals where effort and sense of purpose might tend to flag;
- invest resources, available time and energies for the continuous improvement of the work;
- sustain collective efforts and build a sense of community, ownership and an high level of mutual trust among all of the actors engaged in the process;
- stimulate collective effort and co-operation;
- create an information-driven and purpose-driven organization;
- produce a 'learning organization'; and
- stimulate a strong critical control of the data, the results and the reports, and also a definite will to provide national data and improve their quality.

## Conclusion

In brief, there are five key principles which guided the OECD work in preparing its collection of educational indicators:

1. A priority to policy-making: the development of a system of indicators by statisticians must answer policy needs. Statisticians are not in a position to determine what the public must know. There should not be any conflict between the scientific community and policy makers about the determination and publication of the relevant indicators.
2. Indicators comparability: the OECD has made an effort to produce comparable information so that countries can understand the state of their education system in relation to that of other member countries. Achieving comparability was the winning card of the OECD operation.
3. Communication: indicators are used to communicate with policy makers and the public. Since the golden rule of communication is clarity, indicators must be both clear and simple without being superficial. This entails a limited number of indicators. A current debate within the OECD concerns reducing the core of indicators from around fifty today to twenty. Last year, two sets of indicators were produced: a technical set for statisticians and a reduced set of twenty indicators for policy makers and journalists.
4. Statistical credibility: statistical accuracy, reliability and relevancy must not

hinder the publication of indicators. It is better to publish approximations than nothing at all.

5. Accountability: those who work in education must be accountable and responsible for the sector outcomes as measured by indicators. In turn, they must be given a voice in assessing whether the resources they are provided with correspond to the demands placed upon them by policy makers.

These are key principles which oriented the OECD's work and thanks to which success was achieved, especially on the political front as exemplified by the great attention devoted by the media to the OECD's set of indicators. The OECD's findings are now debated in parliaments. This work shows that choosing the right indicators and targeting the right audiences can have an enormous impact on the education system and contribute to the improvement of its quality across the board.

### Note

1. While drafting this paper, I used parts of many OECD internal unpublished documents, produced within the OECD project on International Indicators of Education Systems (INES), which I led from 1988 to 1995. In particular, I am in debt to Donald Hirsch and François Orivel for the papers they prepared for the last General Assembly of the INES Project held in Lahti (Finland) in June 1995.

### References

- INES. 1995. *Public expectations of the final stage of compulsory education*. Edinburgh, The Scottish Office Education Department. (Technical Report Network D)
- NCES. 1996. *Education indicators. An international perspective*. Washington, DC, United States Department of Education. Office of Educational Research and Improvement.
- OECD. 1992a. *Education at a glance. OECD indicators*. Paris.
- . 1992b. *The OECD international education indicators: a framework for analysis*. Paris.
- . 1993. *Education at a glance. OECD indicators*. Paris.
- . 1994. *Making education count. Developing and using international indicators*. Paris.
- . 1995. *Education at a glance. OECD indicators*. Paris.
- . 1996a. *Education at a glance. OECD indicators*. Paris.
- . 1996b. *Education at a glance. Analysis*. Paris.

---

# DIALOGUE EVALUATION: A POSSIBLE CONTRIBUTION OF EVALUATION TO SCHOOL IMPROVEMENT

---

*David Nevo*

---

Extensive educational reforms or innovations, even those advocating bottom-up approaches, have usually been initiated at the national or State level by political parties, national educational administrations or researchers and scholars from academe. However, to be successful, the results of such initiatives must be seen at the school level. Having observed educational innovation throughout the world, it is clear to me that if something significant happens at all in education, it does not happen at the ministry of education or at the university; it happens at the school and in the classroom. Reforms can be conceived by policy makers. They can be planned in ministries of education by experts and administrators. But, to have an effect, they must be implemented in the school by teachers and school principals. To change education is to change the school, and evaluation used to improve an education system cannot be effective without bringing the evaluation process—in its full capacity and potential—into the school building. If evaluation is to be used to improve education it has to be dealt with at the school level.

Evaluation has always existed within the school, mainly in its classrooms, but in a very limited capacity. Evaluation was there even before administrators began

---

*Original language: English*

*David Nevo (Israel)*

Ph.D. Professor of Education and Head, Department of Curriculum and Instruction, School of Education, Tel Aviv University. Professional interest in evaluation theory, programme evaluation, school-based evaluation and student assessment. Current research is focused on combining internal and external evaluation, and working with schools to improve evaluation. Author of *School-based evaluation* (1995). Editor-in-Chief of *Studies in educational evaluation*. E-mail: nevo@ccsg.tau.ac.il

to use it for decision making or the monitoring of operations. But evaluation within the school was usually limited to the evaluation of students, executed by means of tests, grades and report cards. The notion of a systematic evaluation of instructors, programmes, projects or the school as a whole, has reached the school building only in recent years, and is still far from being common practice in even the most advanced education systems. Even schools that have adopted innovative approaches to student evaluation, such as performance assessment, portfolios or exhibitions, do not necessarily extend their perception of evaluation beyond the domain of student assessment.

But it is not only the school that has had a narrow perception of evaluation. For many years this was the general perception of evaluation in education (Nevo, 1983, 1995). Educational evaluation was generally equated with *student* evaluation and the words 'evaluation' and 'measurement' were synonymous. Until the mid-1960s most books carrying titles like *Educational evaluation, Measurement and evaluation in education*, etc., dealt mainly (in most cases, only) with student evaluation, focusing on test construction, reliability and validity; scoring; and comparing the pros and cons of various types of test items. The change came in the late 1960s, and mainly in the United States, when the concepts of *curriculum evaluation* and *programme evaluation* were suggested in the context of newly developed curricular materials and large-scale projects for socially disadvantaged students (Scriven, 1967; Stake, 1967; Stufflebeam et al., 1971).

New ways to perceive evaluation were developed, distinguishing between 'evaluation' and 'measurement,' and extending the scope of evaluation beyond student assessment with the introduction of programme and curriculum evaluation, followed later on by teacher evaluation (Millman, 1981; Millman & Darling-Hammond, 1990). Around this time it was suggested that evaluation could assist with decision-making, classification and accreditation. As a result, evaluation methodology was advanced by combining quantitative and qualitative research methods and was supplemented by other methods, such as judicial procedures (Wolf, 1979) and art criticism (Eisner, 1979). Educational evaluation also became an established profession. University programmes were developed to train professional evaluators at both the M.A. and Ph.D. levels, setting agreed-upon standards to control the quality of the evaluation process (Joint Committee on Standards for Educational Evaluation, 1981, 1988, 1994).

But all this happened—in spirit and in practice—far away from the school building. Evaluation experts, concerned with conceptual problems related to programme and project evaluation, rather than with issues of interest to teachers and school principals, practised their profession with large-scale projects at national and State levels. In the United States, the cradle of modern evaluation and the country with the world's largest testing industry, evaluation provided the conceptual basis and methodological solutions for the accountability movement. It supported the idea of posing evaluation requirements as a means of controlling schools, and it helped create the illusion that instruction could be test-driven. Educational evaluators seemed more interested in working with administrators

and serving the information needs of policy makers, than in helping teachers improve the evaluation methods they had been using for decades, or providing parents and students with the kind of information *they* need as 'evaluation clients'.

Over the last decade, fortunately, things have begun to change. Several countries have been reforming their education systems to empower local schools without necessarily eliminating national standards and requirements. Examples can be found in New Zealand, Norway, Spain and the United Kingdom. Even in the United States, where the school district is still the heart of the education system, interest in the school level has re-awakened. Concepts like *school-based management* and *accelerated schools* are catching on. President Bush's, and subsequently President Clinton's, America 2000 Program attaches great importance to the empowerment of local schools and communities, while setting national standards that all schools should meet. The time seems right to bring some of the conceptual and methodological assets that have been accumulated by programme evaluators over the past thirty years into the school building, to be used for the benefit of students, teachers and school principals.

To use evaluation at the school level calls for a school-based approach taking advantage of current conceptions of educational evaluation and providing practical guidelines for such an application. Thus, in the next part of this paper, I shall present a number of general evaluation principles applied in my approach to school evaluation. On the basis of my work with many schools to develop their evaluation capabilities, I shall then show how such an approach can be implemented realistically at the school level. The paper will conclude by suggesting that a change of discourse might be needed to move educational evaluation from a discourse of description and judgment to one of dialogue.

## Evaluation principles for school evaluation

On the basis of the current understanding of educational evaluation and its application, I would like to present the following evaluation principles applied in our approach to school-based evaluation.<sup>1</sup>

1. *Students and their achievements should not be the only objects of school evaluation.*

As mentioned before, there is a tendency to equate school evaluation with testing and student assessment. We suggest that any entity within the school is a potential evaluation object, provided there is some justified need to describe its nature and/or assess its merit. Typical objects of school evaluation are programmes, projects, instructional materials, teachers and other school personnel, students and the school as a whole.

2. *Outcomes or impacts should not be the only things looked at when evaluating a programme, project or any other evaluation object within the school.*

This does not mean that outcomes and impacts are unimportant, but that an intelligent description or fair judgement of merit cannot be limited to the narrow scope of outcome measures, e.g. test scores. To represent the full scope of an educational object, evaluation should focus on at least four types of information regarding the quality of that object. These four types of information are related to the goals of the object, its strategies and plans, its process of implementation, and its outcomes and impacts.

3. *School evaluation has to serve both the formative and summative functions of evaluation, and must provide information for planning and improvement, as well as for selection, certification and accountability.*

The concept of 'formative evaluation' stresses the importance of an evaluation being constructive and useful, but the suggestion that formative evaluation is an alternative to summative evaluation can be an excuse to avoid demands for accountability. There is no doubt that evaluation should be constructive to pedagogy and administration within the school, but it should also help the school demonstrate its merit to educational authorities, parents and the general public. The tendency to avoid summative evaluation is especially disturbing when it occurs in innovative or experimental schools, which use formative evaluation to improve their operation but seem to forget that they should also have summative evaluation to demonstrate their merit as a viable alternative to conventional schools.

4. *The internal evaluation needs of a school can best be served by a team of teachers and other educators, for whom evaluation is only part of their job definition, supported by appropriate training and external technical assistance.*

The school evaluation team is composed of teachers and other educators (e.g. school counsellors) whose professional training is not in evaluation and for whom evaluation represents only a small part of their job description. The strength of such an evaluation team is in its educational and pedagogical background, its familiarity with the educational and social context of the school, and its non-threatening relationship with the school faculty and administration. Its major weakness is its lack of evaluation skills and experience, but this can be remedied by appropriate training and technical assistance. I clearly prefer a team of teachers supported by technical assistance over a part-time or full-time internal evaluator, whose sole responsibility is to conduct an internal evaluation of the school.



5. *School-based evaluation should be built upon a combination of internal and external evaluation, but internal evaluation is needed before a useful external evaluation can be conducted.*

I conceive school-based evaluation neither as a synonym of internal evaluation nor as an antonym of external evaluation, but as a combination of both. The various functions of evaluation within the school can and should be served by both internal and external evaluations. However, the existence of an internal evaluation within the school is a condition needed prior to any constructive use of an external evaluation. Obviously, internal evaluators better serve the function of formative evaluation, and external evaluators the function of summative evaluation. But a school that does not have an internal mechanism for self-evaluation will have difficulty in developing a positive attitude toward the evaluation process, and will lack the self-confidence necessary for constructive dialogue between the school and the external evaluation team. In such cases, the evaluation becomes a source for accusations, instigating defensiveness instead of acting as a basis for dialogue between internal and external decision makers.

## **The process of developing a school-based evaluation system**

Various conditions have to exist to provide an organizational context in which school-based evaluation can flourish and grow. Such conditions must be developed gradually and in a co-ordinated way, combining available school resources with additional resources raised outside the school. School evaluation is not cost-free, nor is it an effortless endeavour, but our experience of working with schools in several countries indicates that it is feasible and can be done if we want it to happen.

Three elements are required to organize an evaluation-minded school. They are: *awareness*, *training*, and *organizational resources*. The development of school-based evaluation is a typical bottom-up activity, like many other school-based activities (e.g. school-based management), and it would be nice if it started as a school initiated project. This is indeed the case with some schools and could probably be so with others, especially in decentralized education systems. But unfortunately this may be an unrealistic expectation in many education systems, especially those with long traditions of centralization and bureaucratization. In such education systems, the school will have to be encouraged to initiate the development of a school-based evaluation and will also need to be provided with the resources to maintain it, possibly as part of a general trend of school empowerment.

The process of developing an organizational context that would yield the above-mentioned conditions and lead to the development of a school-based evaluation system consists of four major phases: (1) basic training, (2) establishment of school evaluation teams, (3) institutionalization of the evaluation, and (4) establishment of a dialogue between internal and external evaluators.

## PHASE ONE: BASIC TRAINING

The process of developing school-based evaluation starts with a regional in-service training workshop offered to principals and teachers from interested schools. The workshop provides fifty to eighty hours of basic training in programme evaluation, testing, data collection procedures and data analysis. It can be conducted on a weekly or bi-weekly basis, or in a one-week seminar supplemented by several monthly meetings.

The workshop has two major goals. The first goal is to help schools familiarize themselves with the nature of school-based evaluation, so that they can make an educated decision regarding the development of the evaluation process in their schools. For this purpose it is important that principals, or at least their assistants, participate in the workshop. It is also important that during the workshop, participants make formal and informal presentations in their schools, sharing issues discussed in the workshop, in an effort to expand the circle of those familiar with the idea of school-based evaluation, i.e. those who will participate in making the decision about introducing evaluation to the school.

The second goal of the workshop is to teach the 'language' of evaluation so that those who are going to take an active part in the evaluation understand its potential benefits. There should be no expectation that during the workshop, principals and teachers will learn 'how to do evaluation' and be able to go back to their schools and get going without any further assistance. It is very difficult to learn how to perform an evaluation without actually doing one. Therefore, they will learn how to conduct an evaluation only in the next phase, when they actually perform an evaluation in their schools with the assistance of external evaluation tutors.

## PHASE TWO: ESTABLISHING A SCHOOL EVALUATION TEAM

Schools that decide to go ahead and initiate an evaluation establish internal evaluation teams consisting of three to four teachers, at least some of whom have attended the basic training workshops. The decision to establish an internal evaluation team is usually made by a school toward the end of the first phase.

The team consists of teachers and other educators (e.g. the school counsellor), who are partially released from their duties to work on the evaluation. The rule of thumb is that team members should be released for about four to six hours a week and the team co-ordinator for about six to eight hours a week. Teachers should serve on the team for a period of two to three years, and a rotation system should be established so that each year one or two members are replaced. Over the years, all teachers should have an opportunity to participate in the school evaluation team.

Once an evaluation team has been established, the first evaluation object has to be chosen. It can be a school project, a programme of instruction, or some domain of activity within the school, like staff development or parent involvement. This is the object on which the school will practice its internal evaluation capability, and on which the evaluation team will learn how to perform an evaluation. An external

evaluation tutor will work closely with the evaluation team on problems encountered in designing the evaluation, finding or developing appropriate measurement instruments, collecting and analysing data, and reporting findings.

#### PHASE THREE: INSTITUTIONALIZING THE EVALUATION

After the school has experienced several evaluations and has considered the usefulness and relevancy of such evaluations to the school's administrative and pedagogical activities, it is time to institutionalize school evaluation. The internal evaluation team has to become a permanent component of the school's structure and needs to receive continuous funding beyond the experimental stage. Arrangements have to be made to provide the team with technical assistance (usually much less intensive than the assistance provided in the previous phase), supplied at the team's request. The team will continue to evaluate one or two projects or school activities each year according to the needs of the school. The results of all evaluations will be accumulated and organized in a school database, together with other information made available to the school, such as test scores, survey results, or the results of relevant external evaluations. The database must be organized so as to be useful to teachers and administrators in their daily work, while also serving as a basis for the development of a school portfolio, which contains data on the overall quality of the school. Such a portfolio will help the school demonstrate its quality when necessary, and serve as a basis for dialogue with external evaluators.

#### PHASE FOUR: ESTABLISHING A DIALOGUE WITH EXTERNAL EVALUATION

During the last phase of development, which may be implemented parallel to the third phase, the school should be ready for an external evaluation by a national or regional educational authority. External evaluation should be mandatory in all schools, regardless of whether they decide to establish an internal evaluation team. However, schools that decide to develop internal evaluation teams should be given a moratorium on external evaluation for a period of one or two years to allow the internal evaluation capacity to become operational.

But schools should not only learn how to cope with mandatory external evaluations, they should also commission external evaluations by professional evaluators, groups of parents, or teams of teachers from other schools. External evaluations initiated by the school could be directed to evaluation objects of special interest to the school, or to the school as a whole. This would extend the scope of the dialogue between internal and external evaluation, which is indeed the heart of our approach to school-based evaluation.

### **Changing the discourse of evaluation**

The traditional discourse of evaluation, is coercive in nature by virtue of the power vested in the evaluator as the carrier of knowledge or information and as the judge

of quality. Such a discourse creates a relationship between the evaluator and the evaluatee or evaluation client that might be similar to what Freire (1970) called 'banking'. According to this great Brazilian educator, banking is a two-part asymmetric activity, including a 'depositor' and a 'depository'. When applied to education, the student is a depository and the teacher is the depositor. The depositor possesses all knowledge and wisdom; the depository is empty and ignorant. The role of the teacher is to deposit knowledge and wisdom into the mind of the student. The role of the student is to absorb and digest all deposits. Such an approach, according to Freire, is the source of *oppression* in education. To stop this oppression, education has to be based on a *dialogue* rather than on a process of banking.

A process similar to Freire's banking might be observed in evaluation. When the evaluator provides information to the decision maker or client about the objects that he has evaluated, he is the provider or depositor, and the client the depository. The evaluators have many 'good answers' to questions that they hope are relevant to their clients and will therefore be useful to them. The relationship between the evaluators and their audience is even more asymmetric when the evaluator is not only telling the client what is going on, but also how good or bad it is, and what is best, for things in general and specifically for the client.

Like students who do not learn very much in a Freirian banking process where teachers simply try to deposit knowledge into their heads, educators won't learn very much from evaluators whose aim is to deposit ready-made descriptions of reality and judgements into their minds. Like students, who are not motivated to use what they learn from such teaching, educators are not motivated to use evaluation that is a one-way coercive discourse of description and/or judgement. Thus, dialogue in evaluation is needed for two purposes. First, to provide a better learning process to understand reality and, second, to increase motivation to use what has been learned. Dialogue can make evaluation more insightful and increase the use of evaluation results.

Education is a very complex enterprise and evaluation is a very young profession. To understand the complexities of education with the limited means of evaluation, educational evaluators must be more modest in the promises they make regarding their ability to shed light on the 'real problems' of education, and their ability to determine the quality of its practice. They must also be more flexible in their readiness to collaborate with evaluatees and other evaluators, in a joint effort to understand the important issues of educational practice and its improvement. Evaluation must be a basis for dialogue rather than a source of assertive one-sided descriptions and authoritative judgements, and evaluators should aspire to become constructive participants in dialogue, rather than providers of indisputable information or authoritative referees.

Dialogue is also needed to increase the chances of the evaluation being used at all. Research shows that evaluation has a better chance of being used by its clients if they are involved in initiating it, and if they are kept informed about its progress through interaction with evaluators during the evaluation process. An on-going dialogue between evaluators and their audiences might be very constructive in developing positive attitudes toward evaluation and more receptiveness for its findings.

Evaluatees might be less threatened by external evaluation findings, and thus less defensive about them, if they have their own findings to support, complement, contradict or dispute the external findings. If evaluatees have nothing of their own to show when presented with negative external evaluation findings, all they can do is defend themselves by discrediting the findings or by presenting excuses for their failure. Neither of these activities—discrediting the evaluation or defending failure—are very constructive behaviours for improvement. A dialogue that incorporates different perspectives and is based upon multiple sets of findings, can be a more helpful way to understand problems in education and possible ways of coping with them.

To make evaluation a more insightful and useful tool, we need to move from a descriptive or judgemental discourse to a more dialogic one. To do this, some specific principles need to be followed to create solid ground for dialogue. This would change the perspective on evaluation and would eventually be expressed in various school evaluation activities, such as student evaluation, teacher evaluation, programme evaluation and evaluation of the school as a whole.

First, the relationship between the evaluator and the audience has to be a *two-way relationship*. This means the relationship should have a two-way flow of information—a two-way process of mutual learning. Both parties involved in the dialogue are not necessarily equal, but there is symmetry in the assumption that each has something to learn from the other, and something to teach the other. At the beginning of the process nobody knows *everything*, but both parties know *something*, and as they engage in dialogue they come to know more and more.

Regarding the evaluation of students, the assumption would be that not only the teacher knows how good the student is—the students themselves also have something to say about the quality of their work. To find out how good a student really is, those two sources of information have to be combined. To evaluate the student, the teacher and the student have to ‘compare notes’ and exchange information and mutual interpretations of available information. An example of such a two-way process can be found in some of the new *alternative assessment* methods, such as the *portfolio* method or the *exhibition* method where the student’s input and reflections are an important component of the evaluation. Moreover, the two-way process can also be observed when, parallel to student evaluation, the school collects student ratings for teacher evaluation. This reflects the assumption that students can learn something from teachers about their learning, and that teachers can learn something from the students about their teaching.

When we combine external and internal evaluation to evaluate a school as a whole for the purpose of accountability or school improvement, we assume that there is no single objective way to assess the quality of a school. The so-called truth about a school’s quality is not vested with the external evaluator or the school itself. Even if there is one single truth about a school, it rests somewhere in-between and can thus only be revealed through dialogue between the two parties, with *both collecting data and interpreting it*.

Second, there has to be *mutual respect and trust* between the parties. Each has

to believe that the other has a genuine interest in understanding what is at stake and in making a significant contribution to such an understanding. An education system that follows a bureaucratic conception of teaching, defining the role of its teachers as implementers of a curriculum prescribed by administrators and experts, does not trust teachers to understand teaching, nor does it trust them to be qualified to assess its quality. Such a system will probably seek an accountability programme based on external evaluation or employ national supervisors to evaluate and supervise each teacher's work. An education system with a more professional conception of teaching, will expect its teachers to plan, design and evaluate their work, and will strive to achieve identified goals and meet defined standards of excellence. Such a system will see teachers as professionals, who are interested in understanding the problems of their profession, maintaining its standards and assuring its quality. Such teachers are perceived as trustworthy partners for teacher evaluation and as respectable professionals whose evaluation will be better served by teacher portfolios than by the standardized test scores of their students. In an education system like that, teacher evaluation can be based on a dialogue between external evaluation and internal self-evaluation of teachers.

Third, evaluators have to be modest, *acknowledging their limitations*. Some evaluators tend to make unwarranted promises of objective descriptions and unbiased assessments, which they hope to obtain by virtue of their powerful evaluation methodology and instrumentation. Such promises are far beyond the capabilities of current evaluation methods and are usually an expression of exaggerated 'professional pride,' being on the verge of arrogance and deceit. At times, this arrogance is rationalized by an ideology of isolation urging evaluators to preserve their objectivity by detaching themselves from the clients and evaluatees to avoid bias and co-option.

Fourth, evaluation has to be *perceived as a process* and not as a one-shot activity. Unless you want to deal with trivia, evaluation is a complex process of trying to understand complex issues. Sometimes you use figures to understand things. Sometimes you create thick descriptions and portrayals to get the real picture. Evaluation is a process of presenting, analyzing and discussing findings with pertinent audiences, comparing them with other findings, collecting additional information, yielding more findings, coping with added complexities, and so on and so forth. If this is the way you perceive evaluation, then dialogue is an integral part of the process of confronting various perspectives and various findings in an on-going attempt to increase your understanding. Maybe this is what the whole process is all about. But, if you deal with simple problems that can be handled by simple one-shot evaluations, why bother with a dialogue? But then, why bother at all?

Fifth, evaluation has to be *fair play*, and it has to be fair to both parties involved in the dialogue. To be fair, it must first be conducted legally, ethically and with due regard to the welfare of those who are associated with the evaluation or may be affected by its consequences. Teachers who believe that they are being evaluated by the principal or by parents in an unfair way cannot be expected to participate in a sincere dialogue regarding such an evaluation.

For evaluation to be fair, both sides need to be clear about its purpose and

expected benefits, about what will be its price and who will have to shoulder the costs—not only the financial costs. It is unrealistic to expect a school to move into a dialogue with an external evaluator (even if the school has joined voluntarily), if at a certain point it comes to the conclusion that it is being exploited for purposes that do not serve its direct needs. This can happen when a school believes that a certain evaluation is serving mainly the research interests of the evaluator, or the information needs of the education system.

Sixth, evaluation has to be *relevant* to the issues at stake. A dialogue—any dialogue—is a demanding undertaking, requiring openness, self-confidence and energy from its participants. People will tend to avoid dialogue if it is not related to issues that are really important to them. If an evaluation deals mainly with trivial matters or is preoccupied with methodological sophistication—rather than issues of substance—school people will lose interest in the evaluation and will refrain from participating in any dialogue. The tendency of some evaluators to baffle their clients with methodological sophistication when they cannot dazzle them with findings, is a naive attempt to make the evaluation look good when it does not. Such attempts only increase the alienation of school people from evaluation and spoil its reputation.

Seventh, both parties have to be *responsible for the consequences* of the evaluation. If a dialogue develops between external and internal evaluation parties, a joint sense of responsibility for the consequences may also develop. If the external evaluation of a school is conducted by a national or regional authority, and we want to move into a dialogue between the school and that authority regarding the quality of the school, the school cannot be left alone to cope with the findings of this evaluation. The evaluating authority must commit itself to a full partnership in developing solutions to the problems that will be revealed. There is little motivation for a school to move into a serious dialogue for school improvement if the burden of improvement rests only on its shoulders. Leaving the school alone to solve its problems not only decreases the chance of finding appropriate solutions, it implies that the school itself is to be blamed for these problems. And accusations, more than stimulating problem solving and constructive action, arouse defensiveness and counter-accusations. Unfortunately, demands for accountability and the use of parental choice as a remedy for school problems are, in fact, examples of the tendency of education systems to *not* share with their schools the responsibility of bearing the consequences of school evaluation.

## Conclusion

Evaluation can play an important role in improving education within a school, but to make this possible it must change its nature—and not merely in a technical way. So far the relationship between an evaluator and a client has been conceived essentially as an asymmetric one. Such asymmetry goes against the principles of constructive dialogue and presupposes a distinction between evaluator and client that has to be reconsidered. The distinction, suggested by prominent evaluators with the good

intention of developing client-oriented approaches to evaluation, stresses the importance of serving the information needs of identified clients rather than the research interests of individual evaluators.

But this distinction between evaluators and clients also stresses the distinction between those who know and those who have to be told; between those who collect systematic information and those who think intuitively; between those who find out what is wrong and those who have to fix it; between those who have the authority of knowledge and those who have the authority of action.

Maybe what we need to do is diminish this distinction to create a basis for dialogue. Maybe the rule should be that nobody may have the authority to evaluate if he or she is not ready to share the responsibility for the consequences of the evaluation. And nobody should be expected to use an evaluation, if he or she is not *involved in* the evaluation. Evaluators must be ready to learn from audiences and clients, and not only set out to educate them. And clients should not only be recipients of information but also collectors and providers of information.

Schools will probably never be good clients of the evaluation process unless they start *doing* evaluations, thereby becoming an equal partner in the dialogue for school improvement. Schools have to be evaluation minded if evaluation is to play a significant role in improving education. But schools are made up of students, teachers and administrators, and so only they can make it evaluation-minded. Students, teachers and school administrators should become active participants in the dialogue of evaluation, rather than passive recipients of evaluative descriptions and judgements. They should stop complaining about how bad evaluation is and start doing it better.

## Notes

1. For a more detailed presentation of this approach and its underlying principles see Nevo (1995).

## References

- Eisner, E.W. 1979. *The educational imagination*. New York, Macmillan.
- Freire, P. 1970. *Pedagogy of the oppressed*. New York, Continuum.
- Guba, E.G.; Lincoln, Y.S. 1981. *Effective evaluation*. San Francisco, CA, Jossey-Bass.
- ; Lincoln, Y.S. 1989. *Fourth generation evaluation*. Newbury Park, CA, Sage.
- Joint Committee on Standards for Educational Evaluation. 1981. *Standards for evaluations of educational programs, projects, and materials*. New York, McGraw-Hill.
- . 1988. *The personnel evaluation standards*. Newbury Park, CA, Sage.
- . 1994. *The program evaluation standards*, 2nd ed. Thousand Oaks, CA, Sage.
- Millman, J., ed. 1981. *Handbook of teacher evaluation*. Beverly Hills, CA, Sage.
- Millman, J.; Darling-Hammond, L., eds. 1990. *The new handbook of teacher evaluation: assessing elementary and secondary school teachers*. Newbury Park, CA, Sage.
- Nevo, D. 1983. The conceptualization of educational evaluation: an analytical review of the literature. *Review of educational research* (Washington, DC), vol. 53, no. 1, p. 117–128.



- . 1995. *School-based evaluation: a dialogue for school improvement*. Oxford, U.K., Pergamon.
- Scriven, M. 1967. The methodology of evaluation. In: Stake, R.E. ed. *AERA Monograph Series on Curriculum Evaluation*, no. 1. Chicago, Rand McNally.
- Stake, R.E. 1967. The countenance of educational evaluation. *Teachers college record* (New York), vol. 68, p. 523–540.
- Stufflebeam, D.L., et al. 1971. *Educational evaluation and decision making*. Itasca, IL, Peacock.
- Wolf, R.L. 1979. The use of judicial evaluation methods in the formation of educational policy. *Educational evaluation and policy analysis* (Washington, DC), vol. 1, p. 19–28.

---

# VALUE-ADDED MEASURES OF SCHOOL EFFECTIVENESS IN THE UNITED KINGDOM<sup>1</sup>

---

*Sally Thomas*

---

## Introduction

Over the last thirty years much research has been undertaken on the influence of the school. Early work by Coleman, Jencks and others (Coleman et al., 1966; Jencks et al., 1972) concluded that family and neighbourhood characteristics have a greater impact on pupil performance than individual school characteristics. However, subsequent research (Rutter et al., 1979; Mortimore et al., 1988; Goldstein et al., 1993; Sammons, Thomas & Mortimore, 1997) has demonstrated both that schools typically receive variable intakes of pupils and that the outcomes of schooling are not totally determined by their intakes. For example, Thomas and Mortimore (1996) have reported that up to 70–75% of the school variation in 16-year-olds' performance in the General Certification of Secondary Education (GCSE) is explained by pupil intake measures. However, importantly, the remaining unexplained variation between schools—25–30%—incorporates the impact of individual schools (and other external factors) on the performance of pupils.

The issue of appropriate and valid ways of reporting schools' performance is of vital importance. Few would disagree that there are a wide variety of important

---

*Original language: English*

*Sally Thomas (United Kingdom)*

Ph.D. Senior Research Lecturer and Associate Director of the International School Effectiveness and Improvement Centre (ISEIC), University of London Institute of Education. She has contributed to a wide variety of research projects covering education, psychology and statistics, and has acted as a consultant on assessment in Australia, Brazil, China, Netherlands and the United States. Her publications include articles in a variety of academic journals and a recent book (with P. Sammons and P. Mortimore) on *Forging links: effective schools and effective departments*. She is on the editorial board of the *British educational research journal*. E-mail: s.thomas@ioe.ac.uk

goals of schooling, such as enhancing positive pupil attitudes towards learning and behaviour, encouraging community values and vocational aspirations, as well as promoting academic achievement. However, the scope of this paper is mostly limited to the aspect of schools' performance measured by examination and test results, although examples of other types of outcomes are also described. Recent trends in Government policy in the United Kingdom have focused attention on schools' raw examination results, taking no account of the context of the school or the characteristics of the pupil intake. Nevertheless, it has been argued strongly by many educational practitioners as well as academic researchers that, taken on its own, information about schools' raw exam results—such as those published by the Department for Education and Employment (DfEE)—will always be a very inadequate measure of performance and, without any knowledge of the context of that school, can be misleading (McPherson, 1992; National Commission on Education, 1993). For example, the late Desmond Nuttall (1990, p. 25) argued that 'natural justice demands that schools are held accountable only for those things that they can influence (for good or ill) and not for the pre-existing differences between their intakes.'

Raw results are the appropriate answer to the question 'How has my son or daughter performed?' but not to questions about the school's performance. They do not answer questions about how effective the learning and teaching is in the individual classroom or school. They do not indicate if a school is performing as well as it might be.

This paper aims to provide an overview and examples of what are commonly called 'value-added measures of school effectiveness' (VAM). Illustrations of how VAMs have been incorporated into school self-evaluation frameworks in the United Kingdom are also provided. In contrast to raw examination results, VAMs provide the starting point for a way of evaluating school performance that takes account of intake factors outside the control of the school but which have a considerable impact on pupil attainment. In addition, I will show how more detailed VAMs can also be employed to examine the *internal* variations in a school's effectiveness at the classroom, subject and departmental levels.

## The difficulty with league tables

Policy makers and academic researchers (Nuttall et al., 1989; Dearing, 1993) have highlighted the danger of using league tables solely based on raw results. Schools with high achieving intakes will tend to do well for that reason alone. Neither the initially high achieving nor the initially low achieving school is assisted by the publication of raw league tables. In the former, the need for improvement may not be appreciated; in the latter, serious demoralization of staff may occur through no fault of their own. Crucially, VAMs show whether some schools are performing markedly better or worse than other schools having taken into account intake factors. This point was powerfully illustrated in *The Guardian's* publication of schools' raw and value added 'A'-level examination results in 1992 and 1993 (Thomas, Nuttall & Goldstein, 1992) (see Table 1).

TABLE 1. *The Guardian* analysis, 1992

Value added	Raw	Number of post-16 institutions (n = 402)
A	C	1
A	B	48
C	A	6
C	B	39

Notes: A = scores in top 25%; B = scores in middle 50%; C = scores in bottom 25%

In Table 1, out of a total of 402 institutions for post-obligatory education (age 16), six were in the top 25% (Group A) of raw 'A'-level scores, but when contextual factors and prior attainment at the GCSE level were taken into account, their value-added scores were in the bottom 25% (Group C). Overall, almost one in four institutions changed their categorization from the raw to the value-added results.

Therefore, to make judgements about an individual school's performance, a range of information is needed in addition to the raw test or examination results. Information about the value added by a school to individual pupil's progress is one important source.

Measuring the educational value added contributed by a school to individual pupil achievement has a number of purposes:

- it offers a fairer and more meaningful way of presenting school examination results;
- it is a tool which can provide both detailed and summary data that a school can analyse as part of its self-evaluation;
- it can be used to examine trends in value-added performance over time, in relation to school improvement initiatives;
- it provides performance measures that can be contrasted against other types of data available in schools such as measures of pupils' affective or vocational outcomes or information about the views of key groups obtained using teacher, parent and pupil questionnaires; and
- it can provide additional guidance in monitoring and target-setting for individual pupils and specific groups of pupils (such as boys or girls, or certain ethnic groups).

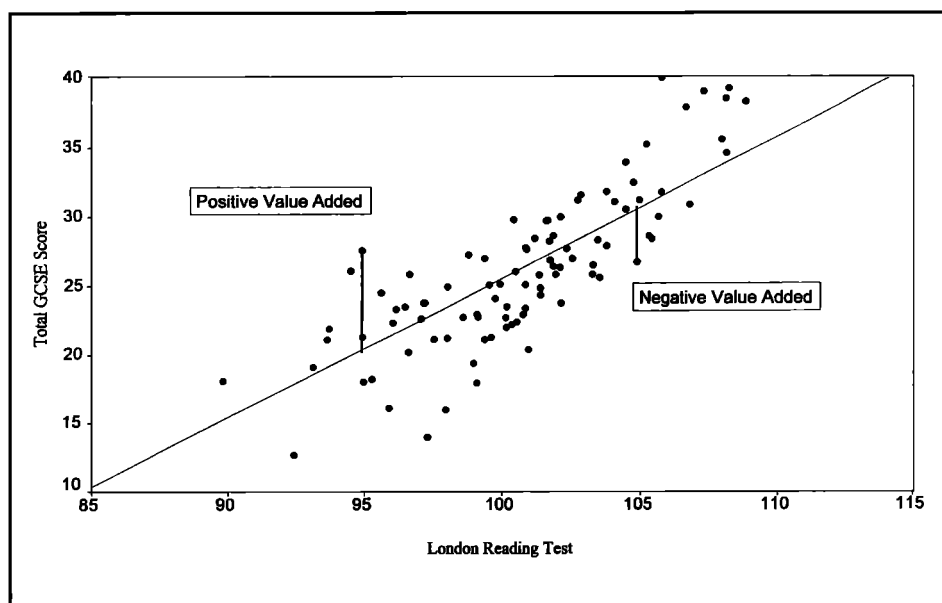
The concept of value added is, therefore, both an indicator of a school's effectiveness and a tool for head teachers and their staff to analyse the extent to which they have effectively raised pupil achievement. As argue below, however, it is not a magic wand. It has real limitations which need to be well understood.

## What is meant by value added

The term value added is a measure of the relative progress made by pupils in a school over a particular period of time (usually from entry to the school until public examinations in the case of secondary schools, or over particular years in primary schools) in comparison to pupils in other schools in the same sample. It compares outcomes after adjusting for varying intake achievement and reflects the relative boost a school gives to a pupil's previous level of attainment in comparison to similar pupils in other schools. Accurate baseline information about pupils' prior attainment is therefore necessary to calculate the value-added component (Mortimore, Sammons & Thomas, 1994; Thomas, Sammons & Mortimore, 1995). A simple example of aggregated school-level data is shown in Figure 1 to illustrate the value-added component. Each point represents one school. A positive value-added score (residual) indicates that a school may be performing above expectation. A negative value-added score indicates a school may be performing below expectation. However, information about the 95% confidence interval is required to evaluate whether an individual school's value-added performance is likely to have occurred by chance. In other words, the confidence interval is vital to judge whether a school's performance—above or below expectation—is *statistically significant* (see example in Figure 2).

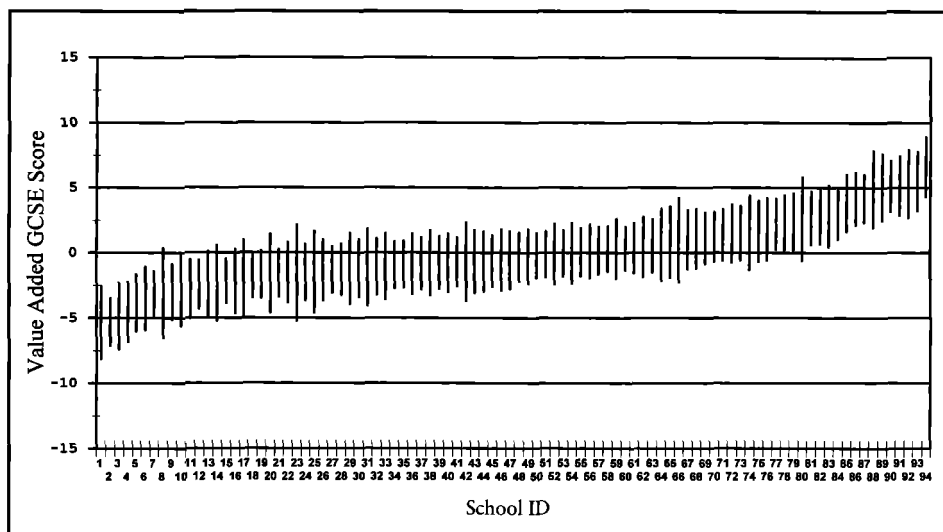
Information is also needed about the statistical uncertainty of performance measures when different schools are compared. The example in Figure 2 shows that the 95% confidence intervals for most schools overlap, preventing fine distinctions between schools. The issue of statistical uncertainty provides a very important caveat in interpretation of any comparative data. Essentially, confidence intervals reflect

FIGURE 1. Value added example



the inherent lack of precision in the calculation of any numerical measures, such as value-added, or indeed, raw scores. Moreover, when making comparisons between schools, the size of the confidence interval changes according to the number of schools being compared (Goldstein & Healy, 1995).

FIGURE 2. 95% confidence intervals



## A national value-added framework

The Government has now accepted the case for VAMs and the School Curriculum and Assessment Authority (SCAA) has published recommendations for a national value-added framework (SCAA, 1994, 1997; Fitz-Gibbon, 1995). However, the major difficulty of introducing a national framework for VAMs is the lack of reliable standardized assessments to measure the prior attainment of pupils entering school. There are no national assessments of pupils entering primary school and the national curriculum assessments at Key Stages 1–3 may not differentiate sufficiently between pupils or be sufficiently reliable for the purpose of measuring value added. Finely differentiated and reliable attainment measures are necessary to describe accurately pupils' starting points. However, if there were to be any change to the national curriculum assessments, the benefits of teacher and standard task assessments in enhancing the quality of teaching and learning would need to be maintained and, at the same time, complemented with assessments that can be used for the purpose of measuring value added. Current developments include the requirement of Local Education Authorities (LEAs) to implement a recognized system of baseline testing for 5 year olds (SCAA, 1996). Some LEAs, such as Surrey and Hampshire, have already employed baseline assessment to evaluate school effectiveness for the infant or junior phase (Sammons & Smees, 1997).

## The development of VAMs

The development of VAMs as indicators of school effectiveness has arisen from a variety of sources.

### THE NEEDS OF ACADEMIC RESEARCH

Many school effectiveness studies, in particular those carried out prior to the mid-1980s, were hampered by the limited statistical techniques available (Scheerens, 1992) and did not have access to the recently developed, sophisticated and now widely preferred method of analysis—multi-level modelling (Goldstein, 1987, 1995; Paterson & Goldstein, 1991).

### THE NEEDS OF LEAS

The requirements of the 1980 Education Act and the 1991 (Schools) Bill (section 16) for schools to publish their 'raw' public examination results placed a much greater emphasis on the search for fairer and more accurate measures of school performance and this has led to the increasingly widespread and systematic collection of pupil data by local education authorities including information about pupil examination and assessment outcomes and other pupil and school characteristics (Hill, 1994, 1997).

### THE NEEDS OF INDIVIDUAL SCHOOLS

Individual schools have also addressed the issues of school performance and effectiveness as an aspect of their own internal monitoring and evaluation and external inspections, such as those carried out by the local education authority and, at the national level, by the Office for Standards in Education (OFSTED) and DfEE (DfEE, 1996). Various projects such as those set up by the London Institute of Education, the University of Durham and the National Foundation for Educational Research have assisted schools by providing VAMs (Spours & Hodgson, 1996; Hill, 1997).

Academic researchers, LEAs and schools have employed a wide variety of different procedures for measuring school effects using either pupils' background factors (such as socio-economic status) or pupils' prior attainment data (or both), as well as different levels of sophistication in the analysis (employing individual pupil-level data or cruder aggregated school-level data) (Gray, Jesson & Sime, 1990; Thomas, Nuttall & Goldstein, 1993*a*, 1993*b*; Thomas, Pan & Goldstein, 1994; Hill, 1994, 1997). However, recent research has shown that pupils' prior attainment is the most important and reliable indicator and predictor of their subsequent attainment—explaining up to 59% of the total variance in pupils' academic outcomes and consequently up to 76% of the variance between schools' performances (Thomas & Smees, 1997).

An important issue revolves around what value-added methodology should be employed when prior attainment data are lacking (whether at the national or local level). In the absence of prior attainment data, socio-economic information and other pupil characteristics (e.g. fluency in English, gender, age, education of parents, entitlement to free school meals) have sometimes been used as approximate measures of previous attainment.

An example of this approach is described in an OFSTED report (Sammons et al., 1994) which proposes an interim procedure for evaluating school performance in the absence of nationally available attainment data on entry to secondary schools. However, it should be emphasized that research shows that when intake attainment data are available, the inclusion of socio-economic information in the calculation of VAMs adds little in accounting for differences between schools (Thomas & Mortimore, 1996). Nonetheless it is useful as a method of 'fine-tuning' the VAMs and will make a difference to the estimates obtained for a few individual schools. This is particularly important for schools which serve skewed intakes with high numbers of certain groups (e.g. boys, or those from low-income families). It must also be emphasized that the presence of good intake attainment data on individual pupils is crucial as without this, valid value-added analysis is not possible. In terms of the value-added concept—*pupils' relative progress in attainment*—this aspect of performance cannot be evaluated without a longitudinal methodology incorporating both baseline and outcome measures. Thus, merely having socio-economic data or aggregate school level data is, to a large extent, inadequate.

## **Statistical techniques and methodology for measuring value added**

It is important to use the appropriate statistical techniques and methodology for measuring value added as an indicator of school effectiveness. The following section describes different approaches and issues related to this methodology. The points raised are not exclusive but aim to provide a stimulus for further investigation and development of VAMs.

### **DETAILED PUPIL LEVEL DATA AND MULTI-LEVEL MODELLING**

Calculating the effect that a school has on an individual pupil's progress is complex. The more information it is possible to have about individual pupils, sub-groups of pupils, and all pupils in a school as well as across schools in a LEA, the more reliable and informative any subsequent analysis is likely to be. Multiple regression analysis is the statistical technique that enables this analysis to happen, whilst multi-level modelling is a recent development of this technique.

A key challenge for researchers has been to develop models which allow the statistical analysis to separate out the effect of the school experience on individual pupil outcomes (what pupils achieve) and the extent to which pupil intake



characteristics (those things the pupils arrive at school with such as the level of attainment they have already reached, their socio-economic background) affect pupil outcomes. Multi-level modelling is now widely recognized as the most accurate and flexible tool for examining the hierarchical nature of pupil attainment data, especially when the aim is to disentangle the complexity of schools' effectiveness (Goldstein, 1987, 1995).

#### DIFFERENT MODELS FOR SEPARATING AND MEASURING EFFECTS

Thomas and Mortimore (1996) compared five alternative models to establish the best value-added approach. The best model for calculating VAMs involved controlling for a range of individual pupil intake factors:

- pupils' prior attainments in verbal, quantitative and non-verbal cognitive ability tests;
- gender;
- age;
- ethnicity;
- mobility; and
- entitlement to free school meals.

However, as noted previously, the level of attainment an individual pupil has when s/he begins at a school is the key component in valid value-added analyses.

#### CONTINUITY OF PRIMARY SCHOOL EFFECTS

Other value-added research has focused on the issue of the continuity of primary school effects at the secondary level (Sammons et al., 1996; Goldstein & Sammons, 1997). Initial results indicate a lasting impact of primary school effectiveness on pupils' progress in secondary school. In other words, pupils from primary schools where the learning and teaching was effective appear to continue to make better progress at secondary school than pupils from less effective primary schools. Therefore new developments in the methodology of calculating VAMs may need explicitly to take into account previous schools attended by individual pupils.

#### PUPILS CHANGING SCHOOLS

This last point is also directly relevant in dealing with the sometimes frequent occurrence of pupils changing schools *within* an educational stage. Hill and Goldstein have recently argued that this issue has important implications for the accuracy of a school's effectiveness measures which relate to a period of time when many pupils may have left and other new pupils have started (Hill & Goldstein, forthcoming). Further work is required to examine how this issue can be addressed in a national value-added framework.

## STABILITY OVER TIME AND INTERNAL VARIATIONS IN SCHOOLS' EFFECTIVENESS

Overall statistics of pupil performance cannot give an accurate picture of how effective a school is at raising and maintaining the achievement of all its pupils, or how capable it is of sustaining its standards over time. The availability and analysis of individual pupil-level data is essential to examine different aspects of a school's effectiveness. Some schools that may appear to be effective in terms of the overall VAM may not be so effective in terms of individual departments, for different groups of pupils or over different periods of time.

As an example, I will use the results of a study of ninety-four inner-city London secondary schools (Thomas et al., 1997a, 1997b). This project analysed three aspects of schools' GCSE performance:

- stability of results over three years (1990–92);
- consistency across subject departments; and
- differential effects for different groups of pupils (such as high and low attainers).

A value-added approach was employed controlling for selected pupil background measures of prior attainment at secondary transfer, gender, age, ethnicity and income. Seven different school-outcome measures were investigated: total GCSE score, and scores in English, English literature, mathematics, science, French and history.

*Stability over time*

The evidence over the three years showed that schools' effects on the total GCSE performance score were more stable (with correlations year to year ranging from 0.82 to 0.85—a perfect relationship being represented by 1.00) than those for specific subjects. At the subject level, results for French were the least stable and those for history the most stable. These correlations are slightly lower than similar research by Gray and colleagues that looks at total GCSE score only (Gray et al., 1995; Gray, Goldstein & Jesson, 1996), but this may be due to the special influence of teacher and pupil mobility in inner-city schools. Despite evidence of broad stability in some areas, the correlations indicate that there is also a substantial degree of change over time in subject results, and highlight the necessity of looking at results in more than one year and trends in results over a minimum of three years. It is important to emphasize that 'real' improvement (or decline) in performance resulting perhaps from a shift in educational policy or practice can only be identified by examining long-term changes in results over time.

*Consistency across departments*

The evidence concerning consistency between departmental and overall school results (taking into account three years of data) ranges from fairly strong in some cases to fairly weak in others. For a substantial proportion of schools (e.g. approximately one-third in 1991), there were significant departmental differences in terms of

effectiveness. These differences would not be apparent if only total GCSE scores for a school were used. These findings are in line with previous research which looks at a more limited range of outcomes (Goldstein, 1993) and suggests strongly the need to look at school performance in detail—not just at total GCSE performance but also at the subject level—to tease out effective departments. Similar work at the primary level has been carried out by Tymms and colleagues as part of the SCAA national value-added project (Tymms & Henderson, 1995). However, one limitation of looking at individual subjects is that sometimes there are only a few pupils taking a particular examination (particularly at ‘A’-levels) and therefore the statistical uncertainty of the value-added score may be relatively large. Nevertheless, the impact of this limitation will naturally reduce as data are built up over time and sample sizes increase.

### *Differential effects for different groups of pupils*

The study also addressed the important question, ‘Are some schools more effective for particular pupil groups?’ The value-added analysis showed that some schools obtained differing value-added results for different groups of pupils. The findings were particularly strong for pupils categorized by prior attainment measures and by ethnicity. Significant differential effects were also found for girls versus boys and for pupils from low-income families versus other pupils.

However, the results also suggested that all pupils in schools and departments that were effective for the average pupil were likely to perform relatively well at the GCSE level but particular groups (such as those not entitled to free school meals) were likely to perform especially well. In contrast, all pupils in schools and departments that were less effective for the average pupil were likely to perform poorly but particular groups (such as Asian pupils) were likely to perform not quite so poorly. These findings highlight again the usefulness of value-added techniques in allowing a more sensitive analysis of data, and ensuring that under-achievement of particular groups of pupils within each school is recognized.

Similar work has been carried out with primary schools (Sammons, Nuttall & Cuttance, 1993; Thomas, 1995) and post-16 institutions (Goldstein & Thomas, 1996; O’Donoghue et al., 1997), indicating that internal variations in effectiveness need to be monitored at all stages of statutory education. Overall this evidence strongly suggests that schools need to monitor and evaluate their performance using a range of different VAMs for particular groups of pupils, for individual departments, for different pupil cohorts and points in time (and where possible also for different year groups within the school).

## **Value-added results as a tool for school self-evaluation**

I have emphasized the need for schools to analyse data in a more sensitive and detailed way, at a range of levels: individual pupils, various pupil groups, sub-groups,

subject level, whole school and LEA-wide. Importantly, schools need to collaborate with other schools at the local, regional and national levels in order to provide comparative data for value-added analyses. Numerous LEA projects are currently in progress and further details can be found in a recent report by the United Kingdom's National Foundation for Educational Research (Hill, 1994, 1997).

However, the collection and analysis of *both* quantitative and qualitative data is essential if schools are to be 'self-evaluating' institutions. 'Indicators do not provide the diagnosis or prescribe the remedy—they are simply suggestive of the need for action' (Nuttall, 1990).

Value-added data are helpful for school self-evaluation by raising questions about changes and/or consistency in results over time, highlighting differences between individual departments in a school compared to the whole school value added, and allowing schools to compare themselves with other schools (within limits of statistical uncertainty).

There are numerous different issues and approaches to be considered when interpreting value-added results:

- Consider the importance of confidence limits when making comparisons between schools—if the confidence intervals of two particular schools overlap then there is no *significant* difference between their performances (see for example Figure 1).
- Bear in mind limitations of the methodology for your school. How relevant to your school's results are issues of: measurement error, missing data, data accuracy and the retrospective nature of the data?
- Track changes in results over time to examine real improvements and/or random fluctuations in performance
- examine departmental and/or teacher effectiveness versus summary measures of school effectiveness (e.g. total GCSE performance for the average pupil) and their implications for whole school policies
- examine differential effectiveness for different groups of pupils (e.g. boys/girls, high/low attainers) and implications for equal opportunities
- employ a wider range of VAMs to reflect more fully the aims of schooling (e.g. using pupil attitudes and vocational as well as academic outcomes)

## **An example of an LEA value-added analysis**

In 1992 a project was initiated by Lancashire LEA aimed at developing, as far as possible, the most accurate, appropriate and fair measures of secondary school effectiveness. Results were to be fed back to schools, in confidence, to assist individual self-evaluation. The project is ongoing, and since 1993 ninety-eight secondary schools in Lancashire have been involved in collecting detailed information about individual pupils. The method of analysis uses multi-level modelling techniques to calculate a wide range of different VAMs (Thomas & Mortimore, 1996).

Pupils' attainment on entry to the school is employed for the GCSE analysis, including three different measures of prior attainment (National Foundation for

Educational Research Cognitive Abilities Test Verbal, Quantitative and Non-Verbal scores). A wide variety of pupils' background factors—such as gender, ethnicity and mobility—outside the control of the school, are also included in the multi-level analysis of pupil outcomes. Key findings from the study include that in 1993, one-quarter of the seventy-nine schools moved up or down twenty or more places when their value-added results were compared with average raw scores. More than half moved ten or more places. As can be seen in Table 2, the project has been extended each year. In 1996, forty different VAMs were provided for each school, incorporating five different GCSE outcomes (total, best five, English, mathematics and science), relating either to the current year (e.g. 1996) or the combined results of three years (e.g. 1993–96). For the purpose of examining trends over time, the approach of employing separate measures which reflect the results of three consecutive GCSE cohorts provides a 'rolling average' measure of school and departmental performance. Additional scores for each outcome relating to pupils of above, below and average attainment on entering the school were also provided. Schools were then able to use this information to focus their attention on individual pupils, groups of pupils or subject departments. Lancashire LEA advisors have also provided additional information to schools as well as giving support and development in using the value-added results for school self-evaluation (Scanlon, 1996).

New developments of the Lancashire value-added project have included collecting pupil attitude data for all Year 9 (14-year-old) and Year 11 (16-year-old) pupils in the spring terms of 1996 and 1997, using instruments designed originally for the Improving School Effectiveness Project (ISEP) (MacBeath & Mortimore, 1994; Thomas, Smees & McCall, 1997). The 1996 findings from seventy schools summarized in Table 3 suggest that in some cases pupil attitudes (represented by five attitude scales: pupil culture, self efficacy, engagement with school, behaviour and teacher support) can vary substantially between schools. This information is used confidentially by schools to evaluate their impact on pupils' experience at school. In other words, pupils' reported attitudes are employed as important outcomes in their own right, in addition to GCSE examination outcomes. Interestingly, recent evidence from the Lancashire project (Thomas & Smees, 1998) shows that the correlation between schools' GCSE value-added scores and pupil attitude outcomes is low (ranging from -0.24 to 0.16). These findings suggest that different dimensions of effectiveness may exist that relate to different—but equally important—aspects of schooling. Both the ISEP and the Lancashire project aim to expand the range of outcomes and VAMs used to evaluate school effectiveness, to reflect more fully the aims of schooling. In addition, head teachers in Lancashire have completed a questionnaire concerning the characteristics, educational processes, and quality of teaching and learning in their school as well as their views about effective schools in general. The aim is that this kind of information will assist schools in linking trends over time in value-added performance to each schools' individual context, targets and strategies for improvement.

TABLE 2. Lancashire LEA Value-Added Project: pupil outcome measures employed over five years 1993–97

1993 GCSE Value-added Scores	1994 GCSE Value-added Scores	1995 GCSE Value-added Scores	1996/97 GCSE Value-added Scores
Total score (by bands 1–3)	Total score (by bands 1–3)	Total score (by bands 1–3) 3 yrs Total score (by bands 1–3)	Total score (by bands 1–3) 3 yrs Total score (by bands 1–3)
English	English (by bands 1–3)	English (by bands 1–3)	English (by bands 1–3) 3 yrs English (by bands 1–3)
Maths	Maths (by bands 1–3)	Maths (by bands 1–3)	Maths (by bands 1–3) 3 yrs Maths (by bands 1–3)
	Science (by bands 1–3)	Science (by bands 1–3)	Science (by bands 1–3) 3 yrs Science (by bands 1–3)
			Best 5 (by bands 1–3) 3 yrs Best 5 (by bands 1–3)
			<b>Pupil Attitude Scales</b> Engagement with school Self Efficacy Behaviour Teacher Support

*Notes:* Bands 1–3 = Value-added scores are calculated separately for pupils categorized in three groups (approximately top 25% = 1, middle 50% = 2, bottom 25% = 3), in terms of prior attainment (at 11 years) in NFER Cognitive Abilities Test. 3 yrs = value-added score calculated using three GCSE cohorts including the current GCSE cohort (e.g. 1996) and two previous cohorts (e.g. 1994 and 1995).

## Moving from measurement of school effectiveness to school improvement

The task of linking school effectiveness measures to school improvement starts with the premise that analysis is the start—not the end—of the process. Monitoring alone does not improve performance, nor does it provide definite distinctions or comparisons. Therefore it is important that information about school, departmental and classroom effectiveness is continuously contrasted with current policy and practice. For example, some secondary schools in Lancashire have used separate value-added subject scores for the most and least able pupils to reflect on and evaluate their systems for setting GCSE pupils. Also schools in Northern Ireland involved in the Raising School Standards Initiative (funded by the Department of Education for Northern Ireland) aim to use VAMs as well as other evidence to evaluate the impact

TABLE 3. Lancashire Pupil Attitude Questionnaire Scales (1996)

	Mean Score*	Standard* Deviation
Scale 1. ENGAGEMENT WITH SCHOOL	2.85	0.08
Item 1 <i>I always like school</i>	[0.17]	
Item 3 <i>I always get on well with teachers</i>	[0.27]	
Item 5 <i>Teachers are always fair</i>	[0.33]	
Item 6 <i>School work is always interesting</i>	[0.19]	
Item 31 <i>Teachers are nearly always friendly towards pupils</i>	[0.20]	
Scale 2. PUPIL CULTURE	3.28	0.07
Item 2 <i>I always get on well with others in my year</i>	[0.34]	
Item 20 <i>I never felt left out of things</i>	[0.38]	
Item 33 <i>I never get bullied</i>	[0.17]	
Item 36 <i>I find it easy to make friends</i>	[0.28]	
Scale 3. SELF EFFICACY	2.93	0.08
Item 26 <i>My work in class is very good</i>	[0.51]	
Item 28 <i>I think I'm very clever</i>	[0.17]	
Item 29 <i>All my teachers think my work in class is good</i>	[0.37]	
Scale 4. BEHAVIOUR	3.25	0.10
Item 3 <i>I always get on well with teachers</i>	[0.07]	
Item 37 <i>How would you describe your behaviour in class? Good</i>	[0.40]	
Item 38 <i>How do you think teachers would describe your behaviour? Good</i>	[0.56]	
Scale 5. TEACHER SUPPORT	2.87	0.12
Item 8 <i>Teachers always help me to understand my work</i>	[0.19]	
Item 11 <i>Teachers always tell me I can do well</i>	[0.25]	
Item 14 <i>Teachers always tell me how I am getting on with work</i>	[0.31]	
Item 16 <i>Teachers always praise me when I have worked hard</i>	[0.29]	
Item 31 <i>Teachers are nearly always friendly towards pupils</i>	[0.14]	

Notes: The Lancashire pupil attitude scales were calculated using the LISREL statistical software [Linear Structural Equation Model for Latent Variables] (Joreskog & Sorbom, 1989). LISREL weights shown in brackets. Attitude scale range: 1 (most negative response category) to 4 (most positive response category, shown in table). \* estimates calculated across schools.

of particular improvement strategies at the school and classroom levels (Thomas, 1996).

Future developments in value-added research are likely to build on current findings that investigate the relationship between measures of school performance and the conditions that appear to enhance or hinder school effectiveness in different types of school contexts. This approach requires the use of both qualitative and quantitative data. For example, research by Sammons and colleagues (Sammons, Thomas & Mortimore, 1997) employed a value-added methodology to evaluate school performance and combined this with interview and questionnaire data to investigate factors and processes related to greater departmental and school effectiveness.

However, further attention is required on the crucial issue of which school improvement initiatives or strategies for improvement provide successful levers to the improved performance of schools over time. A particular issue relates to the variety of strategies that may be successful in different types of context, such as in areas of high versus low socio-economic disadvantage. The ISEP (MacBeath & Mortimore, 1994) is currently addressing these issues. The project's findings will provide important information regarding the implementation and impact of particular strategies for school improvement.

## Conclusions

This paper has mostly focused on the performance of schools in terms of examination and test results. On their own, such results are insufficient for proper judgements about schools' performance. Other pupil outcomes such as pupil attendance, capacity for independent learning, attitudes towards school and learning, behaviour and self-concept are also important. School effectiveness research aims to investigate a broad range of educational outcomes. Examples of this approach are provided by the Junior School Project (Mortimore et al., 1988), the Lancashire project (Thomas & Mortimore, 1996) and the ISEP in Scotland (MacBeath & Mortimore, 1994) which involves collecting pupil, teacher and parent attitude data in addition to academic outcomes. Nevertheless, the main performance indicator for schools continues to be how successful they are at ensuring that as many pupils as possible achieve their full academic potential.

Therefore monitoring of internal variations in performance in any year and across years should be encouraged to facilitate a review of standards for different groups of pupils, and at the departmental, subject and classroom levels as well as overall, to see how these standards may change over time. Results from research where schools are grouped into strictly defined effectiveness categories show that few schools perform both consistently across subjects and with stability over time (Thomas, et al., 1997a). These findings are of practical as well as theoretical importance. School performance that varies greatly over time or between departments in secondary schools has implications for whole school policies and may provide important evidence about the impact of school improvement initiatives. School performance that varies greatly for different groups of pupils (such as boys and girls) has implications for equal opportunities and pupil entitlement. Similarly, significant differences between subject



areas at the primary level have important implications for teacher effectiveness across the whole range of national curriculum subjects (Thomas, 1995; Sammons & Smees, 1997). Hopefully, evidence of this kind—employed as part of a confidential framework for school and teacher self-evaluation—will stimulate and inform teachers' evaluation of their own educational practices as well as the overall quality of teaching and learning in their school and its capacity for improvement in order to work towards raising educational standards for all children.

## Note

1. This research paper was completed via a grant awarded by the United Kingdom Economic and Social Science Research Council. I am very grateful for the helpful comments of my colleagues Louise Stoll, Pam Sammons, Harvey Goldstein and Hillary Street on a previous version of this paper.

## References

- Coleman, J.S., et al. 1966. *Equality of educational opportunity*. Washington, DC, United States Government Printing Office.
- Dearing, R. 1993. *The national curriculum and its assessment: interim report*. London, National Curriculum Council & School Examination Assessment Council.
- Department for Education and Employment. 1996. *Setting targets to raise standards: a survey of good practice*. London, DfEE.
- Fitz-Gibbon, C.T. 1995. *National value-added project. General report: issues to be considered in the design of a national value-added system*. London, School Curriculum and Assessment Authority.
- Goldstein, H. 1987. *Multi-level models in educational and social research*. London, Charles Griffin & Co.; New York, Oxford University Press.
- . 1993. Assessment and accountability, *Education* (TOWN), vol. 183, no. 3, p. 33–4.
- . 1995. *Multi-level statistical models*, 2nd ed. London, Sydney, Auckland, Edward Arnold. (Kendall's library of statistics vol. 3.)
- Goldstein, H.; Healy, M. 1995. The graphical presentation of a collection of means. *Journal of the Royal Statistical Society* (Oxford), vol. 581, no. 1, p. 175–77.
- Goldstein, H.; Sammons, P. 1997. The influence of secondary and junior schools on sixteen year examination performance: a cross-classified multi-level analysis. *School effectiveness and school improvement* (Lisse, Netherlands), vol. 8, no. 2, p. 219–30.
- Goldstein, H.; Thomas, S. 1996. Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society* (Oxford), vol. 159, no. 1, p. 149–63.
- Goldstein, H., et al. 1993. A multi-level analysis of school examination results. *Oxford review of education* (Oxford), vol. 19, no. 4, p. 425–33.
- Gray, J.; Goldstein, H.; Jesson, D. 1996. Changes and improvements in schools' effectiveness: trends over five years. *Research papers in education* (London), vol. 11, no. 1, p. 35–51.
- Gray, J.; Jesson, D.; Sime, N. 1990. Estimating differences in the examination performance of secondary schools in six LEAs: a multi-level approach to school effectiveness. *Oxford review of education* (Oxford), vol. 16, no. 2, p. 137–58.
- Gray, J., et al. 1995. A multi-level analysis of school improvement: changes in schools'

- performance over time. *School effectiveness & school improvement* (Lisse, Netherlands), vol. 6, no. 2, p. 97–114.
- Hill, N. 1994. *Value-added analysis: current practice in local educational authorities*. Slough, National Foundation for Educational Research in England and Wales, Education Management Information Exchange.
- . 1997. *Value-added and school performance: recent LEA initiatives*. Slough, National Foundation for Educational Research in England and Wales, Education Management Information Exchange.
- Hill, P.; Goldstein, H. forthcoming. Multi-level modelling of educational data with cross-classification and missing identification for units. *Journal of educational and behavioural statistics* (Washington).
- Jencks, C.S., et al. 1972. *Inequality: a reassessment of the effect of family and schooling in America*. New York, Basic Books.
- Joreskog, K.G.; Sorbom, D. 1989. *LISREL 7: user's reference guide*. Chicago, SPSS Inc.
- MacBeath, J.; Mortimore, P. 1994. *Improving school effectiveness—a Scottish approach*. (Paper presented at the British Educational Research Association 20th Annual Conference, September, St Anne's College, Oxford.)
- McPherson, A. 1992. Measuring added value in schools. *National Commission on Education Briefing No. 1*, February 1992, London.
- Mortimore, P.; Sammons, P.; Thomas, S. 1994. School effectiveness and value-added measures. *Assessment in education: principles, policy & practice* (Abingdon, U.K.), vol. 1, no 3, p. 315–32.
- Mortimore, P., et al. 1988. *School matters: the junior years*. Wells, Open Books.
- National Commission on Education. 1993. *Learning to succeed*. London, Heinemann.
- Nuttall, D. 1990. *Differences in examination performance*. London, Research and Statistics Branch, Inner London Education Authority.
- Nuttall, D., et al. 1989. Differential school effectiveness. *International journal of educational research* (Kidlington, U.K.), vol. 13, no. 7, p. 769–76. (Special issue Developments in school effectiveness research.)
- O'Donoghue, C., et al. 1997. *1996 DfEE study of value added for 16–18 year olds in England*. London, HMSO.
- Paterson, L.; Goldstein, H. 1991. New statistical methods of analysing social structures: an introduction to multi-level models. *British educational research journal* (Abingdon, U.K.), vol. 17, no. 4, p. 387–93.
- Rutter, M., et al. 1979. *Fifteen thousand hours: secondary schools and their effects on children*. London, Open Books.
- Sammons, P.; Nuttall, D.; Cuttance, P. 1993. Differential school effectiveness: results from a reanalysis of the Inner London Education Authority's Junior School Project data. *British educational research journal* (Abingdon, U.K.), vol. 19, no 4, p. 381–405.
- Sammons, P.; Smees, R. 1997. *Surrey LEA baseline screening*. London, Institute of Education.
- Sammons, P.; Thomas, S.; Mortimore, P. 1997. *Forging links: effective schools and effective departments*. London, Paul Chapman.
- Sammons, P., et al. 1994. *Assessing school effectiveness: developing measures to put school performance in context*. London, Office for Standards in Education (OFSTED).
- ; ———. 1996. Continuity of school effects: a longitudinal analysis of primary and secondary school effects on GCSE performance. *School effectiveness & school improvement* (Lisse, Netherlands), vol. 6, no. 4, p. 285–307.

- Scanlon, N. 1996. *Lancashire LEA value-added project*. Preston, Lancashire LEA.
- Scheerens, J. 1992. *Effective schooling: research, theory and practice*. London, Cassell.
- School Curriculum and Assessment Authority. 1994. *Value-added performance indicators for schools*. London, SCAA.
- . 1996. *Baseline assessment draft proposals*. London, SCAA.
- . 1997. *The value-added national project—report to the Secretary of State*. London, SCAA.
- Spours, K.; Hodgson, A. 1996. *Value-added and raising attainment—a formative approach*. London, Institute of Education.
- Thomas, S. 1995. Considering primary school effectiveness: an analysis of 1992 Key Stage 1 results. *The curriculum journal* (London), vol. 6, no 3, p. 279–95.
- . 1996. *Research proposal for DENI RSSI value-added project*. London, Institute of Education.
- Thomas, S.; Mortimore, P. 1996. Comparison of value-added models for secondary school effectiveness. *Research papers in education* (London), vol. 11, no. 1, p. 5–33.
- Thomas, S.; Nuttall, D.; Goldstein, H. 1992. The Guardian survey (of A-level examination results). *The Guardian* (Manchester), 20 October.
- . 1993a. The Guardian survey (of A-level examination results). *The Guardian* (Manchester; U.K.), 30 November.
- . 1993b. *Report on analysis of 1991 examination results*. London, Association of Metropolitan Authorities.
- Thomas, S.; Pan, H.; Goldstein, H. 1994. *Report on analysis of 1992 examination results*. London, Association of Metropolitan Authorities.
- Thomas, S.; Sammons, P.; Mortimore, P. 1995. Determining what ‘adds value’ to student achievement. *Educational leadership international* (Alexandria, VA), vol. 52, no. 6, p. 19–22.
- Thomas, S.; Smees, R. 1997. *Dimensions of school effectiveness : comparative analyses across regions*. (Paper presented at the Tenth International Congress for School Effectiveness and Improvement, Memphis, Tennessee, January.)
- .; —. 1998. *Examining the dimensions of secondary school effectiveness and improvement*. (Paper presented at the Eleventh International Congress for School Effectiveness and Improvement, Manchester, United Kingdom.)
- Thomas, S.; Smees, R.; McCall, J. 1997. *Room for improvement: analysis of ISEP primary baseline measures*. (Paper presented at a symposium Critical Issues in School Effectiveness Research at Tenth International Congress for School Effectiveness and Improvement, Memphis, Tennessee, January.)
- Thomas, S., et al. 1997a. Stability and consistency in secondary schools’ effects on students’ GCSE outcomes over three years. *School effectiveness and school improvement* (Lisse, Netherlands), vol. 8, no. 2, p. 169–97.
- . 1997b. Differential secondary school effectiveness: examining the size, extent and consistency of school and departmental effects on GCSE outcomes for different groups of students over three years. *British educational research journal* (Abingdon, U.K.), vol. 23, no. 4, p. 451–69.
- Tymms, P.; Henderson, B. 1995. *The value-added national project. technical report: primary*. London, School Curriculum and Assessment Authority.

---

# INSTITUTIONAL ARRANGEMENTS

---

## FOR EVALUATION

---

*Ernest R. House*

---

What institutional arrangements are best for conducting evaluations? To answer that question one must first consider why we have professional evaluation in the first place. Evaluation, as an institution for informing public decision makers, has emerged only in the last third of the twentieth century, and it has emerged first, and most extensively, in advanced capitalist societies.

Advanced capitalism—highly valued for its effectiveness in improving living standards—is also very destabilizing. Great masses of people migrate far from their homes and families to seek jobs in strange cities, cut off from their friends and family, their traditional associations and even their native language. When all human enterprises are subject to standards of productivity and efficiency, there is little room left for the traditional or communal. Yet, it has been the traditional forms of authority—the church, the community and the family—that have supported the legitimacy of governments.

As the traditional bases of support are weakened by the mobility of capital and labour, and as rapid changes in technology bring changes in lifestyle, governments must find new sources of legitimization for their actions other than traditional ones. There are several sources of legitimacy, but most important is a government's ability

---

*Original language: English*

*Ernest R. House (United States of America)*

Professor at the School of Education, University of Colorado, Boulder. His interests are in evaluation and policy. He has been a visiting scholar at UCLA, Harvard and New Mexico, as well as at universities in Australia, Austria, Spain, Sweden and the United Kingdom. Recent publications include: *Professional evaluation* (1993), *Schools for sale* (1998) and *Values in evaluation and social research* (with K. Howe). He has been a recipient of the Lasswell Prize (1989) for the article contributing most to the theory and practice of the policy sciences and of the Lazarsfeld Award (1990) for Evaluation and Theory, presented by the American Evaluation Association. E-mail: [ernie.house@colorado.edu](mailto:ernie.house@colorado.edu)

to sustain material improvement for its society, that is, its ability to increase the material prosperity of its population. Few regimes can survive without providing such material largess. Thus, the governments in these societies try to create economic conditions that facilitate capitalist endeavours.

Another source of legitimacy is the institution of professional science, which has assumed some of the authority once vested in religion. Professional evaluation, through its identification with science, provides legitimacy for government actions by evaluating government programmes and policies for their social worth. When President Lyndon Johnson advanced his controversial Great Society programmes in the United States in the 1960s, he secured legislative passage for them only by assuring the Congress that these programmes would be evaluated. Large-scale evaluation (of these programmes) began in the United States around 1967.

Now, *I am not saying that evaluation is undertaken purely for legitimacy purposes*. Evaluation does serve to determine the effectiveness of programmes, policies and personnel. It does inform government actions and educate public opinion. However, without the legitimating function that evaluation provides to a sceptical public, other information sources, such as the use of experts, would be adopted instead. Scientific evaluation has an advantage in that its findings are deemed credible by the general public.

## Credibility

To provide such credibility, the evaluation must be perceived as being independent and fair in its findings. If it is seen as being improperly biased, it loses its credibility. A loss of credibility is a problem facing some government commissions. Commissions lacking credibility are usually ones in which prominent citizens have produced findings desired by the government. Often such commissions are unbelievable because they lack independence and scientific credibility. In short, they appear to reflect only government influence and direction.

To be seen as credible, evaluations rely on two major practices. One is to employ an objective scientific methodology; the other is for the evaluators to be politically independent of the programmes and policies being evaluated. Within the professional evaluation community, there is constant debate about the efficacy of different evaluation methodologies. These discussions range from the best ways to collect data to how to present findings. Even though these debates are highly technical, they are often hotly contested.

Most judgements about whether methodologies are proper or not reside in the hands of the professional community, rather than with the government. The professional evaluation enterprise, like the professional medical enterprise, depends on a free marketplace of ideas to forge its methodologies. The government influences methodology indirectly by approving and using certain methods in its studies, but seldom does the government engage in the methodological debate directly.

For example, in a five-year study of the major evaluation office in the National Science Foundation (NSF), I found that government agencies tend to approve

traditional methodologies that are well understood and accepted by evaluators and non-evaluators alike, such as surveys and site visits. Government agencies are less likely to fund new methodologies on which people disagree. This methodological conservatism makes sense if the legitimacy of the study is the major concern.

The second major strategy for securing legitimacy is to locate the evaluators in a place that is protected from political interference. One does not want evaluation findings changed by those whose programmes are being evaluated. If the evaluation office is located in a secure place away from those being evaluated, there is less opportunity for interference with the findings. Hence, the evaluation office at the National Science Foundation is located in its own division. The status of its director is equal to that of programme directors in other divisions. Evaluation offices can also be protected from interference by locating them within a higher office of the organization. The Federal Bureau of Investigation is a good example. Its evaluation office is located within the office of the assistant director.

However, there is a price to pay for such political independence. When the evaluation office is removed from the programmes it studies, there is less communication between evaluators and programme personnel. This separation decreases the chances of the evaluation findings being used by those working in the programmes. A major concern of NSF evaluators is how to persuade programme personnel to use their evaluation findings. To promote use of the findings, one would want to engage the programme people early on, preferably when designing the study, to get them to 'buy in'. Yet the separation of the offices discourages such co-operation.

Another way to encourage the use of evaluation findings is to actively build an evaluation culture within the government agencies themselves. This can be done by holding seminars for and having individual meetings with programme staff, thereby helping programme personnel to understand their role in planning an evaluation and implementing its findings. However, building an evaluation culture within an agency is very time-consuming and only a few agencies in Washington have succeeded in this task. Developing an evaluation culture remains a task to be worked at rather than a task accomplished (House, Haug & Norris, 1996).

In addition to being scientific and independent, the evaluation must also be seen as being fair. Relevant interests and viewpoints should be represented. An evaluation will have little credibility with the general public if it puts forth only the interests and viewpoints of the government. Thus, the evaluation must aim to show the interests of all who are affected significantly by the programme. Currently, some of the most heated debates in the evaluation community are about how these interests could best be represented.

## Recent trends

Such have been the developments of the past few decades. What about other trends? One powerful trend has been for governments to be reduced in size. This reduction in size is part of an economic theory (of the Chicago school of economics) that governments should reduce their expenditures so that more money is available to

the private sector, which presumably can then use resources more effectively to produce more wealth. Governments should also deregulate the private sector according to this theory.

In addition, governments should be more businesslike and entrepreneurial (Osborne & Gaebler, 1992). They should plan and evaluate public services, but the services themselves should be produced by the private sector rather than by government agencies. The government should plan activities rather than produce them. Sometimes these policies are referred to as 'new public management'.

Such policies have resulted in decreases in government resources and have led to another role for evaluation. Governments want to make large budget cuts and also have those cuts be seen as legitimate. This has been especially the case in Europe, where the recent wave of evaluation has been closely associated with budget offices. The Thatcher government in the United Kingdom was perhaps the first to use evaluation primarily as a cost-cutting tool, but many other governments have followed, including even social democratic governments in the Nordic countries (Henkel, 1991).

This has resulted in changes to the professional evaluation community itself. Whereas the first evaluation communities in the United States consisted mostly of social researchers and government officials, in Europe accountants, economists and budget specialists have become prominent members of evaluation teams. Efficiency criteria have become as important as effectiveness criteria for evaluating programmes.

Another result has been that evaluation services themselves have been outsourced. To reduce their size, government agencies have outsourced many of their services to the private sector. Tasks once performed by government agencies are now being performed by private firms. The theory of new public management holds that private enterprises can do these jobs more efficiently. And, in fact, when government agencies reduce their staffs, often there are not enough people to do the tasks internally.

Evaluations have also become outsourced more frequently. For example, in the National Science Foundation office mentioned earlier, only three evaluation experts administer about US \$20 million in evaluation studies. Virtually all of the evaluation studies are outsourced to private firms. NSF personnel could not do the job, even if they wanted to. Similar policies have been adopted in Australia, New Zealand, the Netherlands and the United Kingdom (Boston et al., 1996).

Such outsourcing has resulted in the development of specialized evaluation markets. Both profit and non-profit firms conduct studies for the government. Worth noting, however, is that the evaluation market that has developed does not have the characteristics of a neo-classical market. In a neo-classical market, there are many buyers and sellers of goods and services. If buyers don't like one product, they can go to other sellers of the product. If sellers don't like the price offered for their product, they can find buyers with a better price. Hence, market relationships are impersonal, not personal. And, presumably, the market is kept in equilibrium at a high level of efficiency by this balance of impersonal buyers and sellers.

However, the developing evaluation market is not like this, at least not in Washington. For example, only four firms do all the business of NSF evaluation.

The personnel in these firms develop close personal relationships with the NSF personnel. After all, the NSF officials must have evaluation findings to show Congress and they must have dependable contractors to produce the evaluations on time. On the other side, the four contracting firms have become dependent on NSF funding, which accounts for a significant portion of their budgets.

So instead of the impersonal relationships of many buyers and sellers, the market becomes one of personal, face-to-face, bilateral relationships. The advantages of markets—low costs and high quality—do not necessarily apply in such marketplaces. The purchased services may actually cost more and product quality may be questionable, depending on the competency of the contractors.

This is what economists call an ‘imperfect market’. To make such a market work, the government must do certain things. The government agency must be able to define exactly what it wants of the firms contracted to do the evaluation. The agency must be able to determine which contractors are good ones. And it must be able to monitor the products closely, to tell good from bad evaluations (House, 1997). These tasks require monitoring arrangements to be built into the organization.

Outsourcing evaluations also raises issues, such as who has control over the findings. In the United Kingdom, in particular, the Thatcher and Major governments offered contracts that give considerable control over the evaluation findings to the government sponsors of the evaluations (Norris, 1995). In some cases, findings have been suppressed or changed; in others, evaluators have been denied the opportunity to make their findings public.

Although such control does not seem to be a major problem in all countries, one can see that widespread government suppression of findings would negate the credibility of the evaluation process altogether—or at least as far as the general public is concerned. Evaluators would produce reports that no one would believe. In the long run, such practices would be self-defeating for the government since they would contaminate a source of the government’s own legitimacy. But, of course, governments have been known to do unwise things before in an effort to secure short-term advantages at long-term costs.

In short, evaluation is a product of the interactions of the government, the professional evaluation community and the economic marketplace. The interactions among these three sectors of modern society—that is, the State, civil society and the economy—are complex, mutually dependent and often in conflict. But, such is the structure of advanced capitalist societies.

## Future challenges

What future challenges lie ahead for these institutional arrangements for evaluation? First, as mentioned before, there is the potential problem of political interference. Clearly, evaluation has political functions and is politically useful, but, to play its legitimating role for the government properly, it cannot have its processes politicized or it will lose all credibility. Having credibility means being able to publish negative



findings. What citizen anywhere would believe that government actions are always positive?

Second is the economic problem of specialized evaluation markets. Close connections between governments and firms result in close bilateral relationships that can affect the quality and credibility of work. Will firms that depend on government agencies produce reports favourable to the programmes of those agencies? There is some hint of this in the NSF evaluations. In one nationally visible evaluation, evaluators seemed reluctant to criticize government policy, although evaluators at NSF contend that this is not a widespread problem.

Of course, it is easy to see how contracting firms might think that if they produce negative reports they will be less likely to receive future contracts from the sponsoring agency. Ironically, in this case, agency officials wanted negative findings so that they would know how and where to cut some programmes. How the outsourcing of evaluations and the development of an evaluation marketplace affects the evaluations themselves is important. We do not have a clear understanding of this phenomenon at this time.

Third are methodological issues, which include that of fitting appropriate evaluation methodologies to policy issues. One looks for methods that are politically supportable and professionally acceptable—and they are not always easy to find. And it is important to understand how these issues might interact. For example, in any bureaucratic operation, the way information flows across technical, managerial and institutional levels is problematic (Wilson, 1989). Technical workers do the work, monitor the contracts and run the programmes, including the evaluation experts themselves. Managers, usually department and division directors, are concerned about the final products and may have only passing knowledge of the contractors doing the work. Institutional leaders at the top of the organization, who deal with legislators, the public and other political forces to obtain resources for their agency, are only dimly aware of the work being done. Communication among these levels can be dysfunctional.

The Challenger space shuttle 'O-ring' disaster is a striking example in which technical workers (the engineers) advised against the launch of the space shuttle because of cold weather. The contractor, Morton Thiokol Corporation, knew the O-rings that seal the fuel tanks could malfunction in cold weather. High administrators in the National Aeronautics and Space Administration (NASA) were informed of the danger, but never fully understood the likelihood of O-ring failure. They were so concerned about the political consequences of not launching the space shuttle that the disaster occurred; the shuttle was launched and blew up. Since each level of the agency was concerned about different things, the chances of miscommunication were high.

The same is true of evaluation inside the government. For example, legislators demand 'impact' information, a demand transmitted to evaluation experts through the chain of command. The evaluation experts and contractors work out what is technically feasible to do under the circumstances, that is, what kinds of data can be collected. However, what the Congress means by 'impact' data and what can

be produced are two different things. There will be many misunderstandings ahead.

Fourth, there is a practical issue: whose interests and viewpoints are to be included in the evaluation so that it is seen as being fair and impartial. Evaluations must appear to be fair and impartial if they are to have credibility. Being impartial means including the interests of all relevant parties in the evaluation, including those without power or voice. This touches on issues of social justice in evaluation.

I addressed these issues in *Evaluating with validity* (House, 1980), but will elaborate on them in a new book. In brief, my contention is that evaluation should be democratic, dialogical and deliberative. It should be democratic in that it includes the interests and viewpoints of all relevant stakeholders. It should be dialogical in that it engages in dialogue with stakeholders so that the evaluation represents stakeholder interests authentically. And the evaluation should be deliberative so that it provides opportunities for stakeholders to ascertain what their real interests are so as to arrive at wise decisions. Overall, evaluation should inform a deliberative democracy.

## Summary

Evaluations should be independent in terms of being able to arrive at impartial conclusions. They must be seen as being credible if they are to fulfil this function. Their impartiality is derived from using appropriate scientific methodology and by being protected from political interference. Within the government, evaluation offices should be lodged in independent agencies located close to decision makers, but not in the same office as the programme being evaluated. Scientific methods that protect against various forms of bias should be derived with assistance from the professional evaluation community.

Increasingly, evaluations are sponsored by government agencies and conducted through private firms causing a specialized marketplace for evaluation services to develop. These contractors must be seen as being professionally competent and government agencies must do certain things to make these arrangements work, including monitoring the quality of the evaluations produced and the capacity of the evaluation contractors themselves.

Finally, I realize that many people will think that talk of the political independence of evaluators from governments is naive, that government officials will always seek to interfere with evaluations and to influence findings to their own advantage and, of course, many government officials will do that. But the other significant consideration is the relative weakness of governments in advanced capitalist States. Governments need legitimacy in these societies because critical economic decisions have been placed beyond their reach. Governments need friends who can support their legitimacy, even if those friends are fair and impartial.

## References

- Boston, J., et al. 1996. *Public management: the New Zealand model*. Oxford, U.K., Oxford University Press.
- Henkel, M. 1991. *Government, evaluation and change*. London, Jessica Kingsley Publishers.
- House, E.R. 1997. Evaluation in the government marketplace. *Evaluation practice* (Greenwich, CT), vol. 18, no. 1, p. 37-48.
- . 1980. *Evaluating with validity*. Beverly Hills, CA, Sage Publications. [In Spanish: *Evaluación, ética, poder*. Madrid, Morata, 1994.]
- ; Haug, C.; Norris, N. 1996. Producing evaluations in large bureaucracies. *Evaluation* (London), vol. 2, no. 2, p. 135-150.
- Norris, N. 1995. Contracts, control, and evaluation. *Journal of education policy* (Basingstoke, UK), vol. 10, no. 3, p. 271-285.
- Osborne, D.; Gaebler, T. 1992. *Reinventing government*. Reading, MA, Addison-Wesley.
- Wilson, J.Q. 1989. *Bureaucracy*. New York, Basic Books.

---

# HOW CAN EVALUATION CONTRIBUTE TO EDUCATIONAL POLICY? THE USES OF INFORMATION IN AUSTRALIA<sup>1</sup>

---

*Barry McGaw*

---

Educational policy is formulated and implemented at the system and the school levels. Even where the system sets strong constraints on what schools do, there is always some level of discretion at the school level about the particular shape of policies and how they are implemented.

Some highly decentralized systems, such as those in the United States of America and the United Kingdom, are seeking to develop stronger national policies to set up a framework for school policies. Some traditionally centralized systems are devolving much more responsibility for decision-making about policy and practice to the school level, but may establish some kind of accountability framework to make clear to schools in which respects they must report to central authorities about their policies, programmes and performance.

---

*Original language: English*

*Barry McGaw (Australia)*

Ph.D. from the University of Illinois. Since 1986, Director of the Australian Council for Educational Research. Originally trained as a secondary science teacher. After completing a post-graduate B.Ed. Honours degree, he joined the Research and Curriculum Branch of the Queensland Department of Education. Following a period of further study in the United States, he became Head of that Branch. In 1976, he was appointed Professor of Education at the Murdoch University in Western Australia. Research interests are in measurement, learning and research policy. He has been engaged in a number of projects for UNESCO and the OECD. E-mail: mcgaw@acer.edu.au

## Use of system-level information

There are various ways in which education systems can evaluate their operations at the system level.

One way is to invite external review. The Department of School Education in New South Wales in Australia adopted this approach for a number of years, convening a small group of distinguished educators from the United Kingdom and the United States who met annually to review and report on some aspect of the system. The benefit that might flow from such a 'blue-ribbon' panel depends on the wisdom and status of the panel members, the scope of the issues they address and the adequacy of the information on which their judgements are based.

An alternative strategy is for the system to establish some routine methods for evaluating its own performance, collecting and analysing a variety of data about system performance.

### THE SHIFT TO A FOCUS ON OUTCOMES

In the 1960s and 1970s, many developed countries sought ways to increase funding for their education systems as they coped first with a population bulge and then with increased participation beyond the years of compulsory schooling.

By the 1980s, the benefits of increased funding were being questioned. As the Quality of Education Review Committee (1985) in Australia discovered, little information had been gathered about the effectiveness of the programmes introduced with the new funding so there was no strong story to be told about their benefits. Almost a decade later, that absence of evidence of benefit had come to be interpreted as the presence of evidence of no benefit. Further, it was argued that if funding had been increased without benefit, it could be reduced without detriment (Victorian Commission of Audit, 1993).

The absence of data about the effects of new programmes and other reforms did not serve the proponents of those initiatives well. However, many of them remained reluctant to support the implementation of any explicit monitoring programme on the grounds that the measures likely to be used would be too narrow to reflect the full range of their educational goals.

There was some willingness to shift the focus away from inputs but, in the first instance, only to intermediate measures such as participation. Increasing participation became a goal in its own right. For disadvantaged sub-populations, the goal became to close the gap in participation rates beyond the compulsory years between the population as a whole and the sub-population of concern.

Increasingly, the focus has shifted beyond participation to outcomes. In Australia, for example, a national report on the education of indigenous Australians (Department of Employment, Education and Training, 1988) built its recommendations for policy development and implementation around four themes:

- involvement of Aboriginal and Torres Strait Islander people in educational decision-making;
- equality of access to educational services;
- equity of educational participation; and
- equitable and appropriate educational outcomes.

Those themes reflect the shift in focus more generally from access and participation to outcomes. Access and participation are means, not ends. They are necessary conditions for achievement but they do not ensure achievement and certainly do not ensure equitable outcomes.

How the outcomes are expressed have also been subject to change. The debate about behavioural objectives in the 1960s and 1970s was an argument about the appropriate level of specificity in the statement of expected outcomes in a curriculum. Behavioural objectives were commended on the grounds that using them set down what teachers ought to expect in a form that would make achievement of the objective obvious, because the evidence would be observable in the behaviour of the student. Critics argued both that there were other important objectives beyond behaviour and that breaking up the objectives into small, observable pieces disaggregated learning to such an extent that the picture of the whole was lost for the parts.

The current move to outcome specifications has some of this risk of disaggregation but the approach is different in important respects. The most significant is that there is a more integrated approach to the conceptualization of the outcomes. Outcome sequences over successive years of schooling are typically specified to clarify the development expected in a student's learning in a particular subject. This approach has been adopted in the National Curriculum Framework (NCF) in Australia, in the specification of key stages in England and Wales and in various statements of standards in the United States.

In the Australian approach, for example, the curriculum has been broken into eight key learning areas, for each of which sequences of anticipated outcomes have been developed on several strands.

In English, there are three strands:

- speaking and listening;
- reading and viewing; and
- writing.

In studies of society and environment, there are six strands:

- investigation, communication and participation;
- time, continuity and change;
- place and space;
- culture;
- resources; and
- natural and social systems.

In the case of studies of society and environment, the first strand deals with key processes used in all studies in this area and the other five identify key concepts to be learned.

There is an ongoing debate about the place of content in the outcomes but it

is clear that they only specify the framework of knowledge, skills, attitudes and values to be learned. The outcome statements do not provide a curriculum through which the knowledge, skills, attitudes and values might be acquired. Curriculum development is a separate task to be informed by the outcomes framework.

#### FROM OUTCOMES TO BENCHMARKS

Specifying educational outcomes involves establishing expectations of learning. It is a significant move in the direction of setting standards but it may stop short of that.

The outcomes specified in the Australian NCF, for example, define a sequence of outcomes that students are expected to achieve during their first ten years of school. These outcomes are grouped into eight broad levels. There is no declaration of the level in the outcome sequence that students might be expected to reach at each stage of school. Level 8 was initially conceived as the level that students might reach in Grades 11-12. Because of the highly structured existing curricula in the states for Grades 11-12, with associated state-wide assessment, it proved extremely difficult to develop a common statement of outcomes for a Level 8 designed to cover those grades. In the end, Level 8 was retained to describe outcomes that the best students might achieve before the end of Grade 10.

The lack of a direct linkage between grade level and outcome level was quite deliberate. It was intended to make clear that, in any grade, students would be expected to be working at a considerable range of levels. Students in Australia advance through the grades by age, not by progress in achieving specific outcomes. Assuming that all students in a grade are at a particular stage in their learning is unreasonable and the school and class curriculum should take account of that.

Some of the state systems have since linked the eleven years from the primary year until Grade 10 to the eight levels in the outcome sequence. The state of Victoria deleted Level 8 and linked the other seven levels to its first eleven years of schooling. Introducing such links was, in part, a response to a perceived need for teachers to be given clearer guidelines of what they should be expecting in general of their students. It is also an attempt to set down expectations about the rate at which students are expected, in general, to learn the things being covered in the school curriculum.

Increasingly, explicit goals of this kind are being described as performance benchmarks. Precisely how such benchmarks should be set is not entirely clear. Past patterns of achievement of students at different grade levels provides one basis. Comparisons with levels of achievement expected at corresponding grade levels in other systems, including those in other countries, is another.

#### *Conceptions of benchmarks*

The notion of establishing benchmarks has been taken up from business and industry. There the discussion is usually of comparisons with 'best practice', even 'world's best practice'. Benchmarks place the emphasis on process not outcomes. Yet outcomes

are not irrelevant—they are relatively unambiguous and provide clear criteria against which to judge the process.

The outcome criteria need not be financial, though the ultimate measure of company performance in business and industry is financial. A purchasing department within a manufacturing organization, for example, could be compared to purchasing departments in similar organizations or against the purchasing component of a specialist organization that only buys and sells. Features of the purchasing processes in the external company could be introduced as a benchmark. Improvements in performance could be tested for, using explicit performance indicators such as transactions per staff member in some unit of time or cost per transaction.

Where outcomes are more ambiguous, as in the provision of human or social services, the focus of benchmarks tends to be on outcomes rather than on process. Comparisons with other organizations are made in terms of outcome measures. Best practice is defined in terms of outcomes, at least in the early stages.

Examples can readily be found in medical practice. In the field of public health, mortality and morbidity rates are used as indicators. The picture obtained can be quite complex, of course. If infant mortality rates decline, morbidity rates among children may well rise because children with impaired health survive. As mortality rates for one disease decline, those for another may rise, but with the average age of death greater than before. Those who do not die of the former cause may live longer but become more likely to die of the latter cause. Other indicators include average length of hospital stay for patients treated for a particular condition, rate of return of released patients and so on.

### *Benchmarks in education*

In education, most current discussions of benchmarks focus on outcomes. Benchmarks are typically defined as levels of achievement to be expected of all students at some stage of schooling. Those Australian states that have linked their school grades to levels in the NCF can claim that the outcomes at the level linked to a particular grade represent the benchmark for the grade by setting down what is expected of students in that grade.

It is not quite as straightforward as this seems, however, since there are eleven grades and eight outcome levels. The linkages defined in the state of Victoria are shown in Table 1.

TABLE 1. Linkage between outcome level and grade in Victoria, Australia

Outcome level	Grade	Outcome level	Grade
Level 1	End of Preparatory Year	Level 5	End of Grade 8
Level 2	End of Grade 2	Level 6	End of Grade 10
Level 3	End of Grade 4	Level 7	Enrichment of those exceeding Level 6
Level 4	End of Grade 6		



The Victorian adaptation of the NCF as its Curriculum and Standards Framework (CSF) clearly goes beyond the specification of sequences in which outcomes might be achieved to a declaration of intention about when outcomes are expected to be achieved.

The document for each key learning area advises schools that these linkages between school grade and the levels in the CSF are made:

to provide clear guidelines to schools about timing and progression of curriculum and standards. The Board recognizes that students do not learn at uniform rates. The CSF provides statements about the years of schooling within which the major elements of each of the key learning areas should be covered. ... In developing the standards, the Board is describing intended outcomes at particular levels. The standards are not simply informed estimates of current achievement, but relate specifically to curriculum focus aspirations (Board of Studies, 1995, p. 4).

The Curriculum Corporation is currently managing a project to set benchmarks for student achievement in literacy and numeracy at Grades 3 and 5. Draft statements have circulated for comment.

The public examination systems in the Australian states and territories at the end of Grade 12 provide an assessment of student performance in the subjects students choose to study. For many years, the systems have operated without a 'pass/fail' distinction. Students have to complete a particular number and pattern of courses to obtain a certificate on which subject results appear. In some cases, marks on a 0-100 scale appear, but 50 is not a 'passing' mark since a fixed proportion of candidates is usually assigned marks of less than 50.

In New South Wales, the government has adopted a standards-referenced approach to assessment and reporting of student performance, as recommended in a review of the Higher School Certificate (McGaw, 1997), but added a requirement that examination committees determine a minimum acceptable level of performance to which the mark of 50 will be assigned (New South Wales Minister for Education and Training, 1997). There will be no requirement that a particular proportion of students score below 50. Whether students score above or below will depend on whether they have reached the required standard or benchmark. In what sense these benchmarks might be comparable across subjects is yet to be determined.

#### SOURCES AND FORMS OF DATA

When the system-level focus was on inputs, various indicators of input quality were used to evaluate systems as a whole. Level of expenditure per student was the most obvious and frequently used. A major national review of schooling in the early 1970s showed Australian investment in school education to be low compared with other Organisation for Economic Co-operation and Development (OECD) countries and variable across systems within Australia (Interim Committee for the Australian Schools Commission, 1973). The review committee developed an index of resource

utilization as a target for increased investment that was to be reached by the end of the 1970s in all systems, with financial support from the federal Government. The target was reached.

Other input indicators, which reflected in large part the way in which financial resources were allocated, focused on the number of teachers and their level of training, numbers and types of ancillary staff, professional development opportunities for teachers and so on. The focus shifted to students in the 1980s as attention was increasingly given to raising rates of participation beyond the compulsory years (beyond Grade 10). This was seen as a desirable objective but it was not a direct indicator of outcomes. As indicated earlier, participation is a necessary but not sufficient condition for achievement.

Data on student achievement speak more directly to the question of whether an education system is enabling students to achieve the outcomes set down as the goals for student learning. All of the Australian education systems now have in place some form of annual data collection to monitor levels of student achievement. Most monitor performance in Grade 3, particularly to assess whether students have developed the basis skills in literacy and numeracy on which other learning will be built, and in Grade 5. Some monitor performance in the lower secondary years as well.

Most now assess the entire cohort of students in order to provide reports to parents. With this volume of assessments to be scored and reported, only limited forms of data collection are used. Since student answers need to be machine scored, response formats are largely limited, though not entirely, to multiple-choice forms.

In earlier years, and even today in Western Australia, monitoring involved only a representative sample of students and not the whole cohort. With a smaller volume of student responses to be scored and reported, it was possible to use more complex questions and open-ended responses. An example from a project monitoring students' performance in science in the state of Victoria is shown in Figure 1.

FIGURE 1. Stimulus material for assessment of students' science beliefs



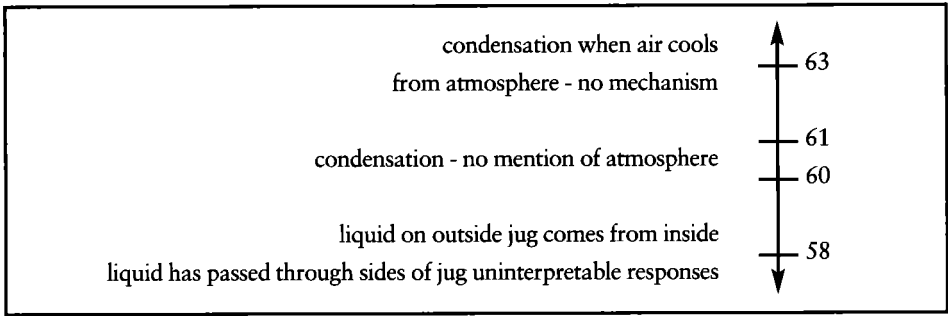
The question shown in Figure 1 was one of a number requiring an open-ended response that revealed something about the idea a student has about the nature of matter. For this item, the responses were classified into the categories shown in Table 2.

TABLE 2. Categories for student responses to condensation question

Category of response	Score	Category of response	Score
condensation when air cools	4	liquid on outside of jug comes from inside	1
from atmosphere—no mechanism indicated	3	liquid has passed through the sides of the jug	0
condensation—no mention of atmosphere	2	uninterpretable responses	0

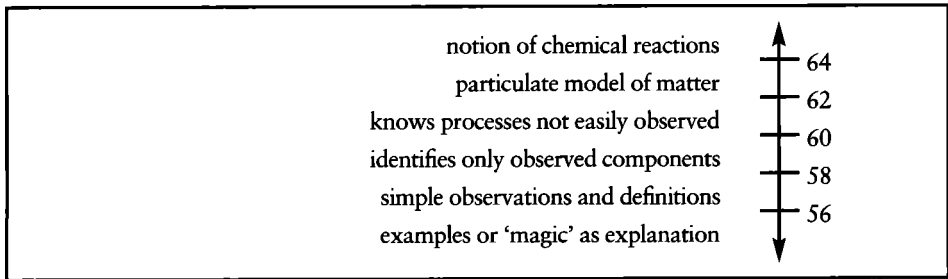
Using a partial-credit model from item-response theory, it is possible to calibrate the locations of the boundaries between the categories of responses to a question like this. The scale has interval properties and numerical values can be placed on it to indicate the scale units. The choice of numerical values is arbitrary, just as the choice of numerical values for the Fahrenheit and Celsius temperature scales is arbitrary. The results of the calibration are shown in Figure 2.

FIGURE 2. Calibration of boundaries for condensation question



From this and the other items drawing on students' conception of the structure of matter, it was possible to calibrate the structure of matter scale shown in Figure 3. The item on condensation involved only a physical change but some of the others to do with cooking involved chemical changes and so drew students whose model of matter provided it beyond a particulate theory of matter to the idea of the particles recombining in different ways as different substances are formed in a chemical reaction.

FIGURE 3. Structure of matter scale



Using a relatively large sample of students, the research team conducting this evaluation of student learning was able with a paper and pen test to undertake a detailed investigation of students' conceptual understanding in science of the type undertaken by science educators using much more intensive interview techniques (Adams, Doig & Rosier, 1991). The assessment instrument was subsequently published separately under the title *Tapping students' science beliefs* (Adams, Doig & Rosier, 1992).

#### DATA INTERPRETATION

Broad monitoring of the performance of students in an education system can be a fairly blunt instrument unless careful thought is given to the interpretation of the data gathered. That, in turn, requires consideration of the audiences for whom the data and interpretations are intended.

##### *Audience*

System-level monitoring is usually justified as a means of satisfying the requirements of public accountability. For this the audience is the general public. How well this is achieved depends on the openness and accessibility of the reporting of the overall student performances in the system.

Others with a legitimate interest in system-level monitoring are senior officials in the system, as this monitoring should provide them with feedback on the general impact of their policies.

Where the system-level monitoring is undertaken with full cohort testing, individual schools and individual parents can obtain information on individual school and individual student performance respectively, set in the context of the overall pattern of results.

##### *System-level data*

While participation data do not constitute an outcome measure, they do speak to questions of equity and they are important in any educational policy evaluation. Different participation rates by gender, socio-economic status and location (e.g. rural or urban) can all signal the need for policy initiatives to reduce inequities.

The effects can be more subtle too. It is not only a question of whether students are participating but in what they are participating. For example, in New South Wales English is offered currently through four different courses at Grades 11-12. These courses are hierarchically arranged by level of intellectual demand—the courses set different expectations for students and enrolments are presumed to reflect students' level of prior achievement. The recent review of the New South Wales Higher School Certificate found that expectations were differentiated by social class as well as by prior achievement (McGaw, 1997, p. 44).

This can be seen by comparing the course enrolments in Grades 11-12 of

students who did equally well in English in a state-wide assessment at the end of Grade 10 and who were in government non-selective high schools, but who differed in geographical origin—one group in the lower socio-economic suburbs of south-western Sydney and the other in the middle to upper socio-economic suburbs of northern Sydney. The details are shown in Table 3.

TABLE 3. Percentage of students in government non-selective schools enrolling in different English courses at Grade 12

	Top 30% in 1994 Grade 10 English state assessment				Top 10% in 1994 Grade 10 English state assessment			
	South-western Sydney		Northern Sydney		South-western Sydney		Northern Sydney	
	M	F	M	F	M	F	M	F
1996 Grade 12 English course								
Two most demanding courses	13	22	22	35	33	51	56	63
Second least demanding course	78	70	73	62	67	48	44	36
Least demanding course	9	8	5	4	0	1	0	1

Among those in the top 10% in Grade 10 English, students in south-western Sydney are more likely than those in northern Sydney to enrol in the second least demanding English course in Grades 11-12 and less likely to enrol in either of the two most demanding courses. A similar pattern occurs with students in the top 30% in Grade 10 English except that, for this broader group, there is a higher likelihood of students in the south-west opting down to least demanding English course.

It turned out that the students' choices were limited by the range of courses that were on offer in their schools. The highly differentiated course structure within English (and other subjects as well) in New South Wales was intended to offer courses appropriate for students with different levels of prior performance. In fact, course differentiation is mapped onto social class differentiation. The students from south-western Sydney shown in Table 3 do not differ from those in northern Sydney in prior achievement in English. They differ only in geography and socio-economic status.

The interpretation of system-level achievement data is not straightforward. The results can provide information about the level and spread of student achievements and comparisons across grade levels if more than one grade level is tested and if the tests are linked and calibrated onto a common scale. Whether the level of achievement should be seen as satisfactory or not is usually a key question in any public discussion of education.

A clear statement of the expected levels of achievement can serve as a benchmark against which to judge the actual levels of achievement. That can, of course, beg the question of whether the expectations are set at an appropriate level.

Repeated monitoring over time can provide evidence of any shift in the standards of performance, provided that the tests in different years are linked and can be calibrated onto a common scale. If the test materials are not secure, then the link between the tests for different years can be established, not by repeated use of some test content, but by developing a composite test with material from tests used in different years, and using it with comparable students in another system.

A recent national survey of literacy skills among primary school students in Australia illustrates these issues. This survey investigated literacy skills of pupils in Grade 3 and 5 in the five aspects of literacy set out in the NCF: writing, reading, viewing, speaking and listening.

The survey established the spread of performance levels of the students and mapped these against the outcome levels in the NCF (Management Committee for the National School English Literacy Survey, 1997). The percentage of students at the various levels of the five dimensions of literacy assessed are shown in Table 4.

TABLE 4. Percentage of pupils from Grades 3 and 5 achieving at each outcome level in the national framework

Outcome Level	Writing		Reading		Viewing		Speaking		Listening	
	Gr 3	Gr 5	Gr 3	Gr 5	Gr 3	Gr 5	Gr 3	Gr 5	Gr 3	Gr 5
5		5*		12*		4*		5*		13*
4	12*	33	12*	39	16*	37	17*	38	20*	47
3	47	47	42	28	42	47	55	50	65	36
2	35	15**	42	21**	41**	12**	26	7**	15**	4**
1	6**		4**				2			

\* at this level or above  
 \*\* at this level or below

These results provide a 1996 baseline of performance levels with which to compare performance levels of similar students in later years. The survey also provided information about the performance levels of sub-groups. Girls outperform boys in all aspects of literacy, with the gender difference being greatest in writing and smallest in viewing. Students from higher socio-economic backgrounds outperform those from lower socio-economic backgrounds, and the gender gap in favour of girls is greatest for students whose parents are from unskilled, manual occupations.

The Commonwealth Minister for Schools, Vocational Education and Training wanted a more precise comparison with a clearly defined benchmark than provided by this mapping onto the outcome levels. The Management Committee asked the researchers Masters and Forster from the Australian Council for Educational Research to see if they could locate points on their five National School English Literacy Survey (NSELS) scales that would correspond to the draft literacy

benchmarks being developed in the national project managed by the Curriculum Corporation.

The attempt to do this was restricted to reading and writing since they were the two aspects of literacy for which the draft benchmarks were most advanced. It proved not to be possible to move from the benchmarks as drafted to the determination of the corresponding locations on the NSELS reading and writing scales, so the task was turned the other way around. The benchmark writers were asked to judge whether each task in the NSELS reading and writing assessments illustrated adequate or less than adequate reading or writing performance for a Grade 3 or Grade 5 student. Their judgements were then combined to establish pass scores on the NSELS tests. The percentage of students whose performances were above and below the pass scores defined by the benchmark writers are shown in Table 5. These results were published in a separate report (Masters & Forster, 1997).

TABLE 5. Percentage of students reaching expected standards in literacy

	Writing		Reading	
Performance Level	Gr 3	Gr 5	Gr 3	Gr 5
Meeting standard	72	67	73	71
Not meeting standard	28	33	27	29

These results became the basis of a substantial and heated public debate. The media presented these results as revealing a literacy crisis in the schools. Although all the available evidence shows that there has been no decline in literacy standards, these results were not only interpreted as low but as lower than before, or certainly lower than critics could remember literacy standards being when they were at school.

The results reflect student performance levels in relation to the levels at which the benchmarks were set. The question of whether the benchmarks are set too high or too low can be addressed by looking at the kinds of performances in reading and writing that fall just above and just below the benchmark. If performances below the current benchmarks were judged to be satisfactory for students at the relevant grade level, then the benchmark could be lowered.

### *International comparisons*

Another way to form a view about whether performance levels are satisfactory would be to use comparative data from another system or country. The Third International Mathematics and Science Study (TIMSS) provided comparisons among the Australian state and territory education systems and also between Australia and other countries.

The international comparisons offered by TIMSS were generally encouraging for Australia. The results are summarized in Table 6. There were countries with superior performances to Australia but, nevertheless, Australian students

did well in the international comparisons. There were no English-speaking or Western European countries ahead of Australia. The state of Western Australia achieved performance levels equivalent to that of the highest performing country, Singapore.

TABLE 6. International rank of Australian students in mathematics and science

Categories of performance	Middle primary		Lower secondary	
	Maths	Science	Maths	Science
Number of countries better than Australia	6	2	8	4
Number of countries not different from Australia	5	6	13	14
Number of countries worse than Australia	14	17	19	22

Rankings shown apply to the upper grade of the two grade levels tested.

While these results may give Australians grounds for some satisfaction, they ought not to be taken as grounds for complacency. There are clearly higher standards being achieved elsewhere and the question of whether higher standards ought to be expected in Australia has to be faced. This introduces the possibility of trying to match international best practice, not in processes but in outcomes.

#### FROM PERFORMANCE MONITORING TO POLICY CHANGE

The task is to use productively the results of monitoring, to move from evaluation to action.

##### *From evaluation to action*

More fine-grained analysis than inspection and comparison of overall mean results is necessary if data from monitoring programmes are to be useful as a basis for evaluation.

Such detailed analyses of Australian student performance in TIMSS in comparison with those from countries that outperformed Australia (notably Singapore, Japan and Korea), point to Australia's deficiencies in mathematics. Australian students performed worse in computational tasks in mathematics, since that is given less emphasis in the Australian curricula. Australia might, on the other hand, have imagined its students would perform better in problem solving tasks but they do not. They sometimes do as well but are often significantly behind on this kind of task too (Stacey, in press).

Analysis of the data from sub-populations can also point to areas in need of action. It is always important to let such data speak to questions of equity in outcomes.



Disadvantaged groups can be identified and provided with additional resources or special programmes.

The state of New South Wales used its basic skills test results in literacy at Grade 3 at one time to identify those schools where significant numbers of students were performing poorly in order to allocate new specialist early literacy teachers to schools with the greatest need.

### *From action to evaluation*

Analysis of the data from system monitoring could then lead to considerations of curriculum reform to address deficiencies. Any curriculum initiatives introduced in this way ought to be evaluated seriously and not just implemented with a blind faith in success. In the long run, benefits ought to be seen in performance levels in system monitoring. In the short run, as the initiatives are being developed and implemented, the standard armoury of the formative evaluator should be deployed to provide the basis for modification and improvement of the programme.

Broad system monitoring is too blunt an instrument to be the measure by which specific curriculum initiatives are shaped or evaluated.

## **Use of institution-level information**

System-level data can yield institution-level information as well, particularly if whole cohorts of students are tested. Interpretation of institution level data is not simple, however, since they are the product not only of what an institution does but also of the characteristics of the students in the institution. This means that institutions should look beyond comparisons with other institutions, while not avoiding them, to a more detailed internal analysis of itself and its community.

### **A SYSTEM-LEVEL ACCOUNTABILITY FRAMEWORK**

The most elaborate institution-level evaluation system in Australia is the one implemented in Victoria in conjunction with a programme of substantial devolution of responsibility to schools for many decisions about curriculum, teaching and staffing that used to be the province of the central authority. The central authority passed responsibility to schools but also imposed an accountability framework under which the schools are required to report to their communities and the central authority.

Under the accountability framework, schools develop a charter in consultation with their community, provide an annual report to the community and the central authority in which performance and progress are reported in relation to the strategic plans contained in the charter, and conduct a triennial review with an external team in order to review and revise the charter for the next triennium.

This programme has been progressively introduced across government schools since 1993 when it commenced with a small group of volunteer schools. The

programme was increased by around 300 schools in 1994, another 500 in 1995 and the final 400 in a mid-year intake in 1995.

The accountability framework is based on five principles: client focus, performance orientation, local ownership and transparency, integration of all parts of the planning, policy and operational activities of the school, and a common set of flexible requirements for all schools in regard to data gathering and reporting.

#### SOURCES AND FORMS OF DATA

Accountability allows schools to develop their own priorities and plans, in line with overall system policies and programmes, and to monitor performance in relation to the local school programme that flows from its priorities and plans.

Data gathering and evaluation are focused on:

- student learning using student performance (on the state-wide monitoring tests and, in the case of secondary schools, on the Grade 12 Victorian Certificate of Education examinations and other assessments) and teachers' judgements of students' performances in relation to the CSF;
- curriculum, using time allocations to the eight curriculum areas; participation by gender in elective subjects in the secondary school; and parents' views about the quality of the curriculum;
- the school environment, using student attendance, student accidents and parents' views of the school environment;
- accountability, using student enrolment data (particularly changes over time), destinations of secondary school leavers, and parents' perceptions of the value of reports on their child's progress;
- management, using views of staff about the quality of school management, extent of professional development undertaken by staff and staff absences; and
- resources, using annual financial statement, and the school's total receipts and expenditure.

In setting their goals, schools are required to consider access and equity for students from non-English speaking backgrounds, students with disabilities and indigenous students.

Guidance is provided to schools about how to implement the accountability framework, including how to gather and analyse relevant data, by the Department of Education's Office of Review. Documents provided by the Office include *Developing a school charter*, *School annual report guidelines*, *School organizational health survey*, *Parent opinion questionnaire* and an associated *Monitoring parent opinion: user guide*.

#### INTERPRETATION OF DATA

##### *Audience for interpretations*

There are two key audiences for a school's annual reports and triennial review reports. One is the school community. The other is the Office of Review which

examines the reports but also combines data from reports across all schools to create a system-level report.

### *Institutional goals as criteria*

A primary focus in interpreting the data yielded by the reviews is the school's own goals defined in its charter. If a school's charter has set a priority on literacy, for example, this will feature strongly in the school's evaluation of its progress and its reports to its community and the Office of Review.

### *Comparisons across institutions*

The Office of Review uses the combined data from schools, either all schools or representative samples of schools, to establish distributions with which schools can compare themselves. These are described as benchmarks but they are, of course, only representations of current performance among schools.

Schools are not invited to compare themselves with all other schools but with 'like schools'. These groupings of 'like schools' are based on the proportion of the student population in the school using a language other than English at home, and the proportion of students who could be said to be 'at risk'—those who are in receipt of a government subsidy for parents with low income.

### *Estimates of 'value added'*

No attempts are currently being made to estimate the value added by primary schools or lower secondary schools. The possibility of such an approach being introduced is increasing now that regular full-cohort testing programmes have been in place for a number of years. There are now groups of students moving through the schools in some systems for whom there are achievement assessments at several grade levels. With these data, it is possible to use an earlier achievement measure to partition out of subsequent achievement those things that the school has not added in the intervening years, in order to estimate the school's influence. The procedures are not straightforward, however, since the measure of value added is a residual which typically has a low level of reliability. Any residual, even a simple difference between two measures, is much less reliable than the measures from which it is calculated.

The one place where value added has been estimated in Australia is at the end of secondary school with the Grade 12 assessments. The approach taken is conceptually flawed in the same way. The measure of non-school effects in both cases is a general developed abilities test actually taken by students towards the end of Grade 12. In Western Australia, the primary use of the test is to scale results in examinations in different subjects to express them on a common scale.

In Victoria, it is a General Achievement Test (GAT) designed to highlight school-based assessments which could be spuriously high (or low) in order to subject them to review. If a school's assessments and the results for the same students on the GAT

are sufficiently out of line, independent assessors are sent to the school to re-mark the students' work assessed originally by the school. If the school's assessments are higher than would have been expected but then survive this re-marking, the value-added estimation process will then re-label the discrepancy as evidence of value added, not as evidence of artificial elevation of the school's assessments.

If this practice becomes standard, there will be an incentive for schools to seek to lower the performance of their students on the GAT in order to create a discrepancy between GAT results and school assessments in a way that would not be removed by a review of the school assessments. Another response might be to try to inflate the school assessments sufficiently to reap some benefit in the value-added estimation but not sufficiently to provoke external review.

The trouble with both the Western Australian approach (now abandoned) and the Victorian approach is that they use a concurrent Grade 12 assessment of general achievement as a proxy for achievement at the point of entry to Grades 11-12 or, worse, as a proxy for 'intelligence'.

## Conclusion

There are volumes of data and much other information potentially available for school systems and schools for evaluating performance and modifying policy and practice. There is always a risk of injudicious data gathering that creates more information than can sensibly be used. We often know more than we understand. Careful selection of the data to be used and appropriate analysis and interpretation can turn information into understanding and perhaps, finally, wisdom.

## Note

1. Presentation at the joint UNESCO: International Bureau of Education and Brazilian Ministry of Education and Sport, *International Seminar on Educational Evaluation*, Rio de Janeiro, Brazil, 30 November-3 December 1997.

## References

- Adams, R.; Doig, B.; Rosier, M. 1991. *Science achievement in Victorian schools: 1990*. Hawthorn, Victoria, Australian Council for Educational Research.
- ; —; —. 1992. *Tapping students' science beliefs*. Hawthorn, Victoria, Australian Council for Educational Research.
- Board of Studies. 1995. *English—curriculum and standards framework*. Carlton, Victoria.
- Department of Employment, Education and Training. 1988. *Report of the Aboriginal education policy task force*. Canberra, Australian Government Publishing Service.
- Interim Committee for the Australian Schools Commission. 1973. *Schools in Australia*. Canberra, Australian Government Publishing Service.
- Management Committee for the National School English Literacy Survey. 1997. *Mapping literacy achievement: results of the 1996 National School English Literacy Survey*. Canberra, Department of Employment, Education, Training and Youth Affairs.

- Masters, G.N.; Forster, M. 1997. *Literacy standards in Australia*. Canberra, Commonwealth of Australia.
- McGaw, B. 1997. *Shaping their future: recommendations for reform of the higher school certificate*. Sydney, Department of Training and Education Co-ordination.
- New South Wales Minister for Education and Training. 1997. *Securing their future: the New South Wales Government's reforms for the higher school certificate*. Sydney, NSW Government.
- Quality of Education Review Committee. 1985. *Quality of education in Australia*. Canberra, Australian Government Publishing Service.
- Stacey, K. in press. Insights for mathematics. In: Lokan, J.; Greenwood, L., eds. *Raising Australian standards in mathematics and science: insights from the Third International Mathematics and Science Study*. Melbourne, Australian Council for Educational Research.
- Victorian Commission of Audit. 1993. *Report of the Victorian Commission of Audit*. Melbourne, Victorian Commission of Audit.

---

# TRENDS/CASES

---

# ASSESSING TEACHER EFFECTIVENESS IN INDIA: OVERVIEW AND CRITICAL APPRAISAL

---

*J.S. Rajput and K. Walia<sup>1</sup>*

---

## The scenario

The universalization of elementary education has been one of the major national concerns in the post-independence era of India. Initially, India faced a shortage of trained and qualified teachers, particularly in the remote regions of the country. Single-teacher schools and alternatives to formal schooling made it necessary to utilize the services of those who were not fully and adequately prepared to undertake the responsibility for teaching. Additionally, a large number of teachers who were underqualified or untrained were appointed as teachers in schools. This continued

---

*Original language: English*

### *J.S. Rajput (India)*

Has contributed in the areas of science education, elementary education and teacher education. Has guided research in the field of concept formation in science, education of minorities and tribal groups, and non-formal education. In addition to numerous research papers and articles, recent publications include *Experiences and expectations in elementary education* and *Universalisation of elementary education—role of the teacher education*. Appointed first Chairperson of National Council for Teacher Education.

### *K. Walia (India)*

Doctorate in teacher education from the Central University of Jamia Millia Islamia (1992). Interest in elementary-stage teacher education and in the development of a competency-based teacher-education curriculum at elementary stage. Has conducted a major study on the profiles of teacher educators in India. Is co-ordinating the development of a curriculum framework for teacher education in India, a task undertaken by the National Council for Teacher Education where she is presently working as a research officer.

for two to three decades. During the last ten years, the position has changed appreciably. Access to education and retention of children in schools have remained significant factors. A third major dimension, learning attainments and quality of education in schools, has received ample attention. The issues of retention of children in schools and higher learning attainments require a congenial learning environment, as well as professionally qualified and committed teachers. Enhancing teacher and school effectiveness has become a major issue for policy-makers, as well as for those responsible for implementation strategies. Before examining the relevant parameters, it would be worthwhile to have an idea of the system as it has developed over the last few decades.

The teacher-education system in India serves more than 4.5 million teachers in over 600,000 primary schools and 170,000 secondary schools. There are more than 1,200 primary teacher-training institutions and some 600 secondary teacher-training institutions. Nearly 90% of teachers in schools possess a training qualification acquired either before joining the system or during their in-service period. The geographical distribution of trained teachers and training institutions is not uniform. Several states, particularly in the north eastern region, have a shortage of trained teachers while in certain other states, the percentage of trained teachers is around 97% to 98%. While there is a surplus of trained teachers in certain regions, others are struggling to prepare more teachers using various alternative strategies. Pre-service teacher training is provided through the university departments of education, colleges of teacher education affiliated to universities and primary teacher-training institutions managed and recognized by the state governments. All training qualifications must be approved by either universities or other state-level agencies. Admission to teacher-training institutions is made on the basis of marks obtained in the school or university qualifying examination. In some cases, specifically designed admission tests are also conducted.

Not all of those who seek admission to pre-service teacher-education programmes are necessarily interested in teaching. This is a consequence of pressures on the job market in various sectors. Further, the teaching profession is not considered to be very lucrative in terms of salary, promotional opportunities and social status. In the majority of cases, the chances of vertical advancement in the profession are few. Schemes for enhancing professional qualifications are also limited. These are some of the factors which dissuade brighter candidates from entering this profession.

School education is the responsibility of state governments. Private agencies and organizations can also open schools and training institutions. However, all institutions have to seek recognition from the state-designated bodies and have to follow the prescribed syllabus. There are also set recruitment procedures.

There is a general feeling that the quality of teacher-education programmes has suffered as a result of expansion. Teacher-training strategies utilizing correspondence courses have been severely criticized in terms of lack of rigour, duration, supervision and academic inputs. Emphasis on in-service education has grown in recent years. Earlier, the situation for in-service education at both the primary and secondary stages was not encouraging. Steps are needed to devise



alternative strategies utilizing new educational technologies and electronic media to upgrade in-service training. To reach a larger number of teachers and to provide orientation for them in these programmes, attempts are being made to adopt diverse strategies, including institution-based training, home study, self-directed learning and interactive television. In-service teacher education has moved from an exploratory phase to that of systematization and institutionalization.

## Teachers at the primary and secondary stages

Teachers in India are appointed for teaching at primary (including upper primary), secondary and senior secondary stages separately. The primary and upper primary stages are referred to as the elementary stage which consists of eight years. It is followed by two years in secondary and another two years in senior secondary. (In subsequent pages 'primary' includes upper primary and 'secondary' includes senior secondary.) Those appointed to teach at the elementary stage are required to have twelve years of schooling followed by two years of pre-service training in teacher-training institutions. Teachers of the secondary stage are required to be graduates: twelve years in school and three years in college/university followed by one year of teacher training. Those teaching at the senior secondary stage are supposed to possess two years of post-graduate qualifications in their respective discipline. The one-year teacher-training programme remains the same for secondary and senior secondary (see Table 1).

Generally, most primary and secondary-level teachers work in conditions of

TABLE 1. Pre-service training of teachers for different levels of education.

Level of education	Type of training	Admission requirements	Duration of the course	Diploma/ degree awarded
Pre-primary education	Private unaided	12 years of schooling	One to two years	Certificate in pre-school education
Primary education	Government Private aided Private unaided	10/12 years of schooling	One to two years	Certificate/ diploma in elementary education
Secondary education	Government Private aided Private unaided	Graduation	One year	Bachelor of education

Source: India. Ministry of Human Resource Development. *Development of education in India, 1995-96*. Delhi, Department of Education, 1996.

deficiencies and even deprivations in various aspects. The existing situation at the primary stage requires greater attention and inputs. It is encouraging to note that policy focus has shifted to the primary stage in view of the national resolve to achieve universalization of elementary education and also to ensure acceptable levels of elementary education quality. Currently, the percentages of trained teachers at the primary and secondary stages are 88 and 89% respectively. At the primary stage, while large number of teachers are qualified and trained, the number of those who are under-trained, underqualified or ill-trained is also considerable. New developments in curriculum transaction and educational technologies are being attempted through different innovative approaches to upgrade qualifications and to train teachers.

Teacher effectiveness at the primary stage suffers considerably due to certain practical factors like the inevitability of multi-grade teaching. To address this concern, nearly 130,000 primary schools which were previously single-teacher schools have now become two-teacher schools and these two teachers look after children in the first five grades. This has been made possible because of a massive scheme launched by the central government to support more than half a million primary schools in terms of staff, equipment and other infrastructure. Under this scheme, known as Operation Blackboard, a second teacher was provided to these single-teacher schools. Efforts are now focused on providing a third teacher. Additionally, teachers have to undertake several other assignments, such as various campaigns launched by the government, or they must participate in conducting elections or censuses. At the secondary stage, teachers may have to teach subjects for which they may not be fully qualified. This is due to the unavailability of adequately qualified teachers. These conditions exercise a demoralizing influence on teachers, thereby reducing their effectiveness and sometimes even their interest.

During the last few years, Operation Blackboard has provided every primary school with some reading material. Secondary stage teachers generally have access to libraries in their own schools. These are generally located in urban areas and small towns, thus the teachers have better opportunities for interaction, access to information and enhanced awareness as compared to those of the primary stages. At both stages of school education, steps are being undertaken to provide teachers with a positive school environment to help motivate and nurture their creative capabilities.

## **Need for assessing teacher effectiveness**

Every successful educational enterprise requires the optimum utilization of the human capabilities available to the system. Consequently, every such enterprise or activity needs periodic assessment and review, followed by a search for better conceptual understanding, implementation strategies and practices. It is now well understood and appreciated that the role of teachers shall change in the twenty-first century for obvious reasons. While it will be necessary for the teachers and the teacher preparation systems to ensure regular acquisition of new skills and upgrading of existing skills, the assessment of teachers' performance shall also remain an essential pre-condition

for enhancing the efficacy of educational processes (Rajput, 1994).

Assessment of teacher performance is essentially based on the idea and understanding that performance can be improved upon. A willing teacher who assists in this exercise is likely to gain more and would probably help to make the evaluation more objective. The teacher has to be realistic and objective in revealing his or her areas of strengths and weaknesses to the evaluator, along with his positive contributions and academic interests. Such an approach also requires the assessor to be well versed, not only in the routine tasks assigned to the teacher but also in the professional practices being adopted by the school to meet the community's expectations. Any system of assessment would remain weak if it does not take these aspects into consideration. The question of who should assess teachers remains an active issue of discussion.

One of the outcomes of assessment could be the possibility of extending vertical mobility, rewards or incentives to the teacher. Through this, apart from professional improvement, a certain amount of motivation could also be achieved. In view of large-scale expansion, the unstated objective of teacher assessment often becomes ensuring the 'maintenance' of standards and norms. Once this is achieved, assessment could be utilized to enhance the effectiveness of the system.

The process of assessment and appraisal often gets confined to administrative inspections. There has been a serious effort to change this. Teacher participation in the assessment process could be assured through generating awareness of the positive aspects that could follow such assessments. Most of the teachers would appreciate a transparent system of appraisal which would result in improvement in establishing communication and rapport between staff at different levels in the hierarchy in the system. Many of them would appreciate receiving critiques on their own performance. Obviously teachers would be more in favour of an appraisal scheme which would ensure promotions or professional advancement as compared to one that could lead to disciplinary measures. However, there would be a number of teachers who would perceive appraisal as a threat utilized as an instrument of accountability. Proponents of greater teacher autonomy consider appraisal as a bureaucratic measure intended to keep a firm grip on them. They also consider it contrary to professionalism in teacher education. The general feeling amongst teachers is that so long as the focus of the appraisal system is on what teachers do efficiently and competently rather than on only highlighting their deficiencies, it would receive their support and achieve the expected goals.

## Existing appraisal systems

The National Commission on Education (1964–66) recommended the introduction of a new system of supervision in schools (National Council of Educational Research and Training, 1970). The National Policy on Education (NPE) recommended the transformation of the system of inspection and supervision into an academic support system. Since the NPE of 1986, several initiatives have been taken to improve upon the existing system. Specific programmes to orient and reorient supervisory staff

(such as the heads of the institutions, principals or headmasters, vice-principals or deputy headmasters, education officers, block level officers, extension officers, district level officers and others) are being regularly conducted by national and state level institutions.

While supervisory systems differ at the provincial level, a broad picture emerges if one considers these on the basis of general approach and guidelines. While the heads of institutions are supposed to supervise the process of classroom transaction informally in every situation, in certain states these tasks have been assigned to them formally. The heads of institutions are often very busy and have many duties other than the academic and the professional tasks expected to be performed by them. Hence academic appraisal by them is not necessarily regular or comprehensive.

Every teacher is supposed to fill out annually a self-appraisal form which indicates the targets set by him in areas of curriculum transaction, extracurricular school activities and interaction with parents and the community. He is supposed to maintain a diary in which the targets achieved are recorded regularly. The specific reasons for any significant gaps between the targets set and those achieved are to be recorded. The diaries are to be submitted to the headmaster or the principal on a fortnightly or monthly basis. At this stage the necessary guidance is expected to be provided by the head of the institution.

In addition to self-appraisals and regular supervision by the head of the institution, periodical inspections and supervision are also conducted by senior-level officers who are designated as inspection officers, education officers, deputy directors or directors of education. With the large-scale expansion of schools at the district level, a separate official mechanism has been set up for the primary and secondary level.

The periodical judgement is often supplemented by a well-established procedure of an annual panel inspection at secondary schools. The panel usually consists of senior teachers of the same level, representatives of the school education departments and subject experts. The panel devotes two to three days to a school and goes through the performance record of every teacher, including the teachers' own diaries/self-appraisals. The panel also observes classroom teaching and pays particular attention to the use of educational technology, presentation of the subject matter, communication skills and the extent of participation evoked from the learners. As a necessary prerequisite it also observes the nature of assignments given to the learner and subsequent guidance given in this aspect after evaluating the assignments. They also review these aspects in light of other administrative and management assignments given to the teacher by the school. This panel submits a report on the infrastructural facilities available or the inputs necessary to be provided to the teachers and school on various accounts.

The panel reports are examined in the department by professionals and the salient features of these are communicated to the head of the school who in turn is required to communicate the same to the individual teachers. Unfortunately, in the majority of cases only negative aspects are promptly communicated to the teacher while positive aspects (which could motivate them) are often preserved 'safe' in the files.

One strategy to enhance teacher effectiveness could be through mutual assessment. Unfortunately, within the system, this has not found an adequate place practically anywhere. Mutual assessment, however, is utilized in the pre-service teacher-education programmes of the elementary as well as secondary levels. While delivering a lesson in the classroom, the trainee is supervised by a senior teacher educator. In addition, the performance of the trainee is also observed by a group of eight to ten fellow trainees who make observations on criteria which includes various aspects of the teaching and learning transaction. Subsequently, the observations of the teacher educator and the peer group are discussed in the presence of the trainee. This has been found to be highly beneficial not only to the individual trainee but also to others who observe.

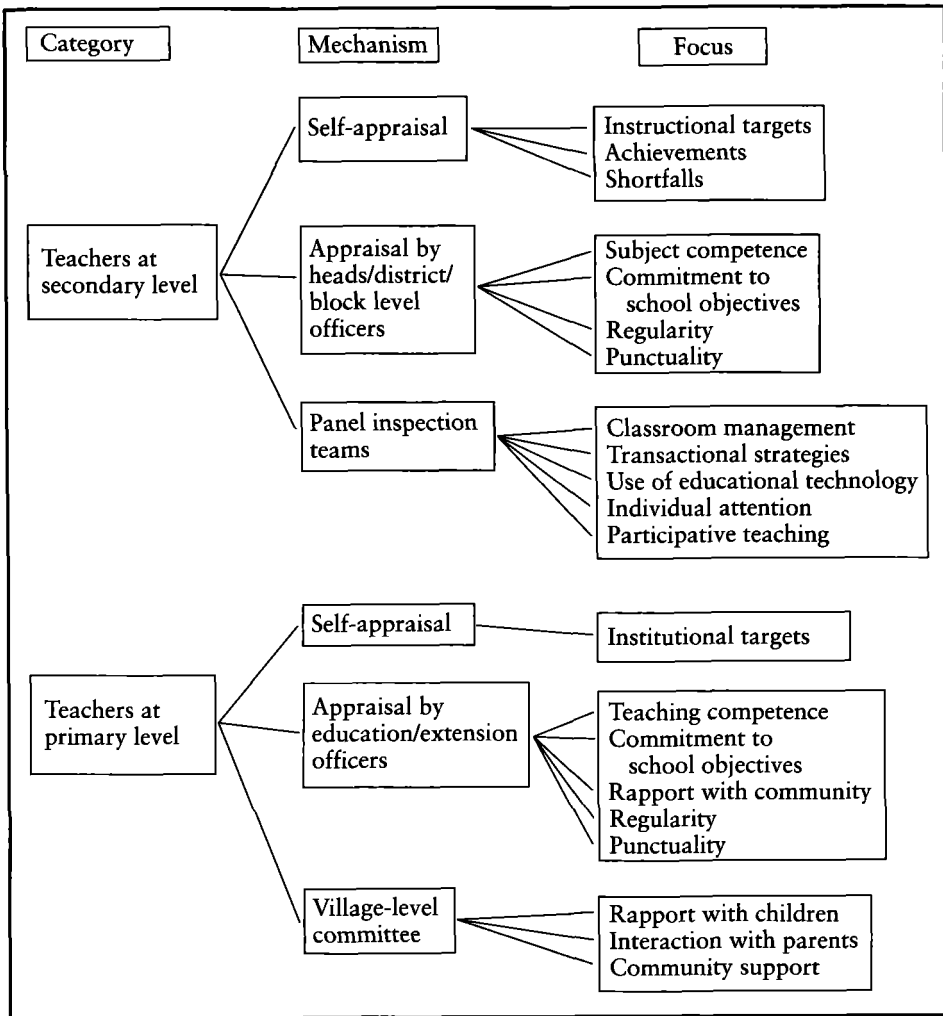
At the primary stage, inspection is often confined only to the strategy of hierarchical supervision which too takes place only rarely. One of the increasingly accepted methods is that of associating the community in evaluating the general performance of learners specifically in areas of affective domain and in cognitive skills. This results in school/community rapport being achieved at the desirable level. It also provides motivation to teachers and in several cases has led to higher learning attainments and enhanced school efficacy. Figure 1 presents an overview of the assessment strategies for the primary and secondary levels.

It should be interesting to assess how teachers react to the existing systems of supervision and inspection. Usually they are apprehensive of the assessment process, particularly of its implications. They also complain of inadequate feedback, particularly on the professional front. They generally favour the concept that appraisal of effectiveness ought to focus upon two-way communication. They were willing to opt for an appraisal system if its outcome could provide feedback on how they are functioning inside and outside the classroom. Teachers are willing to listen to criticism and probably accept judgements, but they also expect appreciation of whatever innovations they attempt. An important function of appraisal for many teachers is to improve staff communication and strengthen the channels of communication within the institution. The appraisal systems in India are such that active discussions with teachers are limited, if not absent.

Assessment should be designed in the sense of helping teachers to solve their problems through a two-way exchange of views. No assessment should be conducted without detailed discussion with the individual concerned. It would also pave the way for openness in the assessment system.

Teachers are particularly concerned about negative remarks going into their records. Every teacher would like to have his/her say about such aspects before these are finally recorded or deleted. It is increasingly appreciated in the Indian education system that there are a considerable number of committed and active teachers who seek appraisal in terms of constantly updating their practice and avoiding falling into routine. The majority of such teachers still favour appraisals and find them conducive to the enhancement of their performance and efficacy. Most of them have their own suggestions about the modalities of supervision and follow-up action.

FIGURE 1. Modalities of teacher assessment.



## Research

Research in various aspects of teacher education is conducted in university departments of education, colleges of teacher education, state-level institutions like the state Councils of Educational Research and Training, the State Institutes of Education and others. At the national level there are several organizations like the National Council of Educational Research and Training (NCERT), the National Institute of Educational Planning and Administration (NIEPA) and the National Council for Teacher Education (NCTE).

Over the last three decades considerable investigative work has been done on teaching, teaching efficiency and teaching effectiveness. Most of the studies have taken samples in the specific context of the stage of teacher education, region, teacher

qualifications, socio-economic background, cultural contexts, and the rural-urban divide as **linking and de-linking** factors in these studies. Teaching aptitude, academic grades, socio-economic status, teaching experience and age appeared to be sound predictors of teacher effectiveness. Separately the main predictors were identified as home, health, dominance, submission, verbal and non-verbal intelligence, and social, emotional and overall adjustment. Highly effective teachers were more intelligent, more self-confident, less suspicious, less prone to guilt and less radical. Intelligence and knowledge in their respective subject areas were also found to be the best predictors of teacher effectiveness. Teacher effectiveness as rated by colleagues and by teachers themselves is significantly co-related with the teacher's ability to do research and publish. It was also found that more effective teachers were characterized by a higher level of differentiation and integration in their cognitive and perceptual functioning. They have a higher capacity for imaginative and original thinking. More effective teachers have a well-developed value system and self-awareness. Professional dignity, altruistic nature, professional involvement, democratic values and family background were found to be co-related with teaching effectiveness. Factors such as region, rank, age, experience and the size of the teacher's family could also significantly influence the level of teaching effectiveness (Buch, 1987).

Predictors of teacher effectiveness which have been researched and investigated include nature, adaptability, attitudes towards children, working conditions, mental ability, professional preparedness, personal contacts and job satisfaction. In addition, various independent variables such as sex, locality, organization, grade, experience and qualification have also been studied. A significant and highly positive association has been found between teaching effectiveness and teaching aptitude. Likewise, a significant difference between effective and ineffective teachers with respect to their level of intelligence has been observed in most of these studies. Effective teachers scored significantly higher mean values than ineffective teachers on all of the individual dimensions of job satisfaction, including nature of work, salary, supervision, promotion, working conditions, and intellectual, emotional and aesthetic self-concept variables.

The majority of the teachers in Indian schools are appointed after having undergone pre-service teacher-education programmes of one or two years duration. The first effort to convert the lay individual into a professional practitioner is made during his/her stay in training institutions. Apart from gaining a basic theoretical understanding of psychological and philosophical context, they are also familiarized with various aspects of pedagogy. In these programmes, the trainees are required to deliver thirty-five to forty lessons in schools under the guidance of expert teacher educators. This aspect of teacher training (commonly known as practice teaching or a teaching internship) is the major component that provides the teacher educator with a chance to observe behavioural changes in the trainee. An assessment is made at the end of the practice teaching of the effectiveness of the training itself in terms of the trainee's attainments.

Research studies have been conducted taking into account practically every variable and the findings, while generally being similar, have also revealed the diversity

that exists in the country in terms of socio-economic, cultural, linguistic and religious multiplicity. It can be inferred that superior teachers are conscientious, controlled, emotionally stable, more intelligent, willing to work together, and aware of the learning needs and emotional requirements of their pupils. The majority of them were found to be innovative, practical and venturesome. The findings of some of the research studies have been taken note of in policies, programmes, initiatives and in implementing long-term strategies (Arora, 1978; Kulandavel and Rao, 1968; Chaya, 1974; Shah, 1995; Sohoni et al., 1977).

## **Support systems to optimize teacher effectiveness**

During the last decade effective steps have been taken to enhance and upgrade all aspects of school education. The National Policy on Education 1986 clearly highlighted the need for the government and the community to create conditions 'which will help motivate and inspire teachers on constructive and creative lines'. The policy further elaborates that 'teachers should have the freedom to innovate, to devise appropriate methods of communication and activities relevant to the needs and capabilities of and the concerns of the community' (India, 1992). Several innovative programmes have been launched as a follow-up to these policy directions. Those directly connected with teachers and teacher effectiveness are briefly mentioned below.

### **INSTITUTIONAL UPGRADE**

With nearly 2,000 teacher-training institutions catering to the pre-service education and in-service education requirements of school teachers, it was decided to establish centres of excellence<sup>2</sup> in teacher education for the elementary and secondary stages throughout the country. A District Institute of Education and Training (DIET) has been set up in every district. These institutions impart pre-service teacher training, conduct in-service teacher-training programmes and function as resource centres for curriculum development, field studies, management and planning, vocationalization of education, non-formal and adult education and educational technology. The necessary infrastructure, faculty and equipment have been provided to all of these institutions.

For the secondary teacher-education programmes, Institutes of Advanced Studies in Education (IASEs) and Colleges of Teacher Education (CTEs) have been established. While the CTEs are supposed to impart pre-service and in-service training at the secondary stage, the IASEs are supposed to guide both the DIETs and the CTEs, conduct research, innovate and prepare resource persons in critical areas such as curriculum development, evaluation techniques, field studies, programme evaluation, educational technology and others.



## CONTINUITY OF TEACHER EDUCATION

It has been increasingly realized that the current practice of imparting a fixed duration of pre-service training followed by sporadic participation in in-service education will not serve the purpose of enhancing teacher effectiveness in the future. The inseparability of pre-service teacher education and that of in-service orientation has been clearly emphasized in the National Policy and has been effectively reflected in the programmes developed subsequently. Various models of in-service education such as institution-based training, self-learning materials, cluster-based approaches and short-term programmes are being attempted extensively. In addition, nationwide programmes oriented to specific changes have also been launched and successfully completed. During the period 1986–90, the Programme of Mass Orientation of School Teachers (PMOST) acquainted nearly 1.7 million teachers with the new trends in the policy imperatives. The Special Orientation Programme for Teachers (SOPT), launched in 1993, trained 417,000 teachers by 1996.

## SETTING STANDARDS IN TEACHER EDUCATION

Due to various factors (particularly the expansion in terms of enrolments and institutions), a consistent concern has been the quality of teacher-education programmes. It was considered necessary to establish a national-level body which could monitor standards, accredit institutions and provide necessary guidelines and professional support in all aspects of teacher education throughout the country. By an Act of Parliament, the National Council for Teacher Education (NCTE)<sup>3</sup> has been set up and vested with the necessary authority to undertake planned and co-ordinated development of teacher education throughout the country and ensure maintenance of norms and standards. The NCTE has been assigned several tasks which include advising the state governments on eligibility criteria of admission to training institutions, recruitment of teachers in schools and consideration of aspects of manpower planning. It is also mandated to look after the areas of research, innovation, studies and surveys on its own as well as in collaboration with other institutions. This unique effort is bound to contribute to upgrading the efficacy of the teaching/learning process and achieve a higher level of teacher effectiveness.

## CURRICULUM RENEWAL

A massive nationwide consultation has been launched by the NCTE to visualize a curriculum framework of teacher-education programmes that would respond to the emerging needs in teacher education (National Council for Teacher Education, 1996). The context in which teachers shall be discharging their duties will become extremely relevant as will be the role of communication and information technologies. The very concept of a teacher in the classroom shall change—consequently, mere cosmetic changes in teacher-education programmes shall not help. This requires a fresh look and in-depth thinking on the components of teacher preparation programmes at the

pre-service as well as in-service stages (Rajput, 1996). At this juncture all teacher-training institutions have been requested to develop their perceptions on the structure, design, duration, content and transactional strategies of teacher-education programmes which they feel would be futuristic in nature. Subsequent to this consultation, the NCTE will propose a suggested framework for teacher-education programmes.

#### AWARDS AND INCENTIVES

Both at the provincial level and the national level, selected teachers are given recognition through awards and incentives for innovative practices and outstanding contributions. The procedure for selection of such teachers has been standardized and leaves no scope for favouritism. Teachers working in remote areas are assessed by a committee consisting of supervisory staff and community leaders. Their recommendations are reviewed at the state level where a panel for state awards is prepared. Recommendations for the national awards emerge at the state level and are finalized at the national level. Annually on National Teachers Day (5 September), nearly 250 teachers receive awards from the President of India. These awards have been found to be highly motivating and have resulted in teachers making sincere efforts to be candidates for the awards.

## Conclusion

Assessing teacher effectiveness is a complex issue with social and historical dimensions. Assessment of teacher effectiveness was an alien concept in ancient and medieval India. The teacher's moral authority, scholarship, wisdom and role in shaping the lives of the youth and the society was unquestionable. The place of the teacher ('the Guru') was always considered much higher than that of the parents.

In the modern context, especially in the context of 'Education For All', things have changed. Teachers are now part of a large, established sector which has its own management system, are accountable to educational authorities and open to formal assessment. The existing system of teacher preparation has its own strengths and weaknesses. It pays scant attention to self-motivated learning. Thus many teachers see assessment of their competencies as a threat rather than as a conduit of feedback for improving their skills and competencies. The teacher preparation programmes are confined to classroom transaction strategies, some theoretical inputs and a few skills (such as asking questions or blackboard work). The teacher/learner relationship becomes trivialized. Under these conditions teachers take a dim view of the necessity of their assessment.

A fresh look at the existing appraisal system is necessary. One of the latest developments is the establishment of village education committees<sup>4</sup> vested with responsibility for schools, recruitment of teachers and teacher assessment. This innovation has the potential of restoring the traditional relationship between schools and the community. Additionally, assessment practices will change due to the explosive impact of information technologies which have the capacity and capability of changing

the very concept of classrooms and transforming the role of teachers. Efficiency shall also become increasingly linked to teachers' attitudes and skills in utilizing and developing an effective rapport with the learners and the community. At present these aspects do not receive attention in appraisals but will become essential in the future.

The existing assessment procedures tilt towards internal indicators and rely heavily upon examination results and teaching performance. They also seek details on operating indicators such as the use of equipment and other resources. However, indicators of innovation, community acceptability, interactive ability and productivity (in terms of publications, material development, design of activities, development of experiments, etc.) should receive due attention. Therefore, the NCTE has recently identified ten competency areas, five commitment areas and five performance areas. The competency areas are contextual, conceptual, content, transactional, other educational activities, developing teaching/learning materials, evaluation, management, working with parents, and working with the community and other agencies. The commitments, to be internalized and put to practice, are commitment to the learner, to community/society, to the profession, to knowledge/excellence and to values. The performance areas are the classroom, the school level, outside the school, parent-related and community-related. These three areas taken together could very well be the guiding factors in assessing teacher effectiveness through well-designed and professionally developed appraisal systems. Teachers then would accept assessment as an opportunity to receive desirable remedial inputs to themselves and to their institutions, thereby improving their worth, credibility and professionalism.

In order to enhance the efficacy of the teacher education system, the Delors Commission makes a very comprehensive recommendation regarding 'a system of evaluation and supervision that helps diagnose difficulties and surmount them, and that uses inspection as a means of recognizing good teaching and encouraging it' (Delors, 1996). This message should guide and inform all future efforts for increasing and enhancing teacher effectiveness.

## Notes

1. The authors gratefully acknowledge the help rendered by Prof. O.S. Dewal in the preparation of this paper. Several ideas and suggestions offered by R.H. Dave, S.C. Behar and R. Govinda are also acknowledged. John A. Smyth of UNESCO provided the opportunity and motivation for developing this paper.
2. The central government's approval of the centrally sponsored scheme 'Restructuring and Strengthening Teacher Education' was given on 27 October 1987. The establishment of District Institutes of Education and Training (DIETs) and Institutes of Advanced Studies in Education (IASEs), and the improvement of Colleges of Teacher Education (CTEs) were major components of this scheme. These institutions were perceived as centres of excellence and the process of establishing and operationalizing them is still in progress. The scheme is fully funded by the Government of India.
3. The central government Act to establish the National Council for Teacher Education (NCTE) was passed by the Parliament in December 1993. The process of operationalizing

the NCTE began in August 1994 when the chairman took charge of his post. The Act came into force with effect from 17 August 1995.

4. The National Educational Policy 1986/1992 envisaged a major role for local committees in programmes of education and school improvement. This was elaborated in the Programme of Action (POA) document of 1986/1992. The POA envisaged the constitution of village education committees (VECs) which would be responsible for the administration of education programmes at the village level. The major responsibility of VECs was perceived as the operationalization of micro-level planning and school mapping in the village through systematic house-to-house surveys and periodic discussion with the parents. Their primary mandate is to ensure universal elementary education to each and every child in the age group of 6–14 years.

## References

- Arora, K. 1978. *Differences between effective and ineffective teachers*. New Delhi, Chand & Company Limited.
- Buch, M.B. 1987. *Third survey of research in education*. New Delhi, National Council of Educational Research and Training.
- Chaya. 1974. *An investigation into certain psychological characteristics of an effective teacher behaviour*. (Ph.D. thesis, University of Kanpur.)
- Delors, J., et al. 1996. *Learning: the treasure within*. Paris, UNESCO. (Report to UNESCO by the International Commission on Education for the Twenty-First Century—The Delors Report.)
- India. Ministry of Human Resource Development. 1992. *National policy on education*, revised. New Delhi, Department of Education.
- Kulandalvel, K.; Rao, T.R.B. 1968. *Qualities of good teachers and good students (A study of students' rating)*. Coimbatore, Ram Krishna Mission Vidyalaya.
- National Council of Educational Research and Training. 1970. *Education and national development—Report of the education commission*. New Delhi, NCERT.
- National Council for Teacher Education. 1996. *Discussion document teacher education curriculum framework*. New Delhi, NCET.
- Rajput, J.S. 1994. The facets of educational change. *Teacher today* (Bikaner), vol. 36, no. 4, p. 9–23.
- . 1996. Search for futuristic curriculum framework in teacher education. *University news* (New Delhi), vol. XXXIV, no. 13.
- Shah, B. 1995. *Determinates of teacher effectiveness*. Ambala Cantt, The Indian Publication.
- Sohoni, B.K., et al. 1977. *A study of the development of teacher effectiveness through teaching practices*. (Ph.D. thesis, SNDT College of Education for Women, Pune.)

---

# PROFILES OF FAMOUS EDUCATORS

## FLORENCE NIGHTINGALE

---

# FLORENCE NIGHTINGALE

---

(1820–1910)

---

*Alex Attewell*

---

## The legend

The popular perception of Florence Nightingale as a romantic heroine ignores her achievements as an educationist. The Nightingale legend nevertheless had an educational impact. It popularized nurse training, which led to the establishment of a new profession for women. While the legend has remained an important part of nursing culture worldwide, it has served to hinder a deeper understanding of Florence Nightingale.

Florence Nightingale rose to fame for nursing the sick and wounded during the Crimean War (1854–56). After the war, she might have taken a high-profile post as a hospital matron and superintendent of nurse training. Instead, she retired from public life to use her influence to campaign and promote educational schemes. Her impact was probably greater for choosing to influence policy rather than exercising power. Florence Nightingale wrote 200 books, reports and pamphlets after the Crimean War, which had a profound effect on army health, welfare in India, civil hospitals, medical statistics and nursing. Her greatest educational contributions were in the establishment of new institutions for the training of army doctors and hospital nurses, but some of her lesser-known educational schemes are illuminating.

Florence Nightingale has been studied as a reformer, statistician, administrator

---

*Original language: English*

*Alex Attewell (United Kingdom)*

Assistant curator of a hospital museum in the west of England before joining the Florence Nightingale Museum, London, in 1989. He became an Associate of the Museums Association in 1993 and Curator of the Florence Nightingale Museum in 1994. He often lectures, gives broadcasts and organizes temporary exhibitions in the area of his expertise. E-mail: [a.attewell@florence-nightingale.co.uk](mailto:a.attewell@florence-nightingale.co.uk)

and researcher, but studies of her impact on education have remained limited. The main history of Nightingale nurse training (Baly, 1986) ignores the broader context of Florence Nightingale's educational views to focus on the often fraught administrative history of the early Nightingale School for nurses.

It is not surprising that there should exist many common threads uniting the various aspects of education with which Florence Nightingale was involved. Her own education and training had no outlet until the age of 31. She felt driven to make some practical use of her education and, consequently, her early letters, notes and booklets continually refer to the purposes of education and criticize the education available to women in her age. Taking these early writings and her later promotion of training schemes, it is possible to claim that Florence Nightingale was a great, if unrecognized, educationist.

## Her early life

Florence was born in 1820, the second child of William and Frances Nightingale, a wealthy couple who both came from Unitarian backgrounds (i.e. they were dissenters from the established Church). She grew up at a time of intense social change surrounded by liberal and reforming ideas. Her maternal grandfather, William Smith, was a Member of Parliament for forty-six years and was one of those who campaigned successfully for the rights of religious dissenters and for the abolition of the slave trade. With a number of family members connected to politics, she naturally had a keen sense of involvement in the great affairs of her day.

The education of Florence and her elder sister, Parthenope, was supervised by their father, who had been educated at the University of Cambridge. His curriculum included Latin, Greek, history, philosophy, mathematics, modern languages and music. William Edward Nightingale held advanced ideas about the improvement of society and the education of women, consequently his daughters' education was invested with a seriousness more often accorded to sons. While Parthenope preferred sketching, Florence was a natural academic. It is a mark of the depth of Florence's learning that in later life she was able to give substantial help to Benjamin Jowett in his translation of Plato's *Dialogues*.

For a young woman of Florence's social standing, the opportunities to make use of her education in a practical sense were strictly limited. From her teens, she began to sense a conflict between the pleasures of the family's lively social life and her inner desire for action.

## Her 'calling'

At the age of 17, a private note records, she had a mystical experience,<sup>1</sup> her 'calling', a pivotal experience which gave her strength in her conviction that she was not destined for a conventional life. In her twenties, she increasingly came into conflict with her family over the question of marriage, but she tenaciously maintained her independence. It was difficult to find a suitable sphere of activity into which Florence

Nightingale could channel her intellect and abilities. At the age of 30, she recalled the frustration: 'During the middle portion of my life, college education, acquirement [of knowledge] I longed for—but that was temporary'. When Queen's College was opened in 1848, offering a new higher education for women, Miss Nightingale had not been interested; she had already been seeking a sphere of action, an outlet for her already considerable education. She wrote:

[...] the first thought I can remember & the last was nursing work & in the absence of this, education work, but more the education of the bad than of the young. But for this, I had no education myself (Vicinus & Nergaard, 1989, p. 30).

In 1845, in search of some practical means of learning about nursing, Florence had asked her parents' consent to nurse at Salisbury Infirmary, where a family friend was head physician. Permission was refused, not because of specific objections to the hospital but because her parents considered the work unsuitable for a lady of Florence's social standing. 'It was as if I had wanted to be a kitchen maid,' she commented. Her bitter conclusion was that only widowhood or poverty could give an educated woman a reason to work. In this bleak period she was given some encouragement by Dr Samuel Gridley Howe, the American pioneer of schooling for blind people. He advised her to persevere with nursing despite the consternation of her family and friends.

While in London in 1848, Florence had the opportunity to teach poor children (her 'little thieves') for several months at the Ragged School in Westminster. The experience opened her eyes to poverty and she felt she could do some good, but was once again held back by family objections: 'if only education could be conducted without reference to what people think or do not think but only to abstract right and wrong, what a difference it would make!' (O'Malley, 1930, p. 151).

## A turning point

In 1849, Florence embarked on a cultural tour of Egypt and Greece, taking time to make some detailed notes on the social conditions and archaeological sites. While returning through Germany, the party visited Kaiserswerth, near Dusseldorf, where in 1836 Pastor Theodor Fliedner had founded a hospital, orphanage and school. The institution was staffed by 'deaconesses' trained by Fliedner and his wife Caroline. At the age of 30, 'the age at which Christ began his mission', Florence returned to train as a nurse at Kaiserswerth against strenuous family opposition.

Miss Nightingale proved herself a pupil of outstanding ability and after three months of training Pastor Fliedner invited her to publish an account of life at Kaiserswerth for the English readership (Nightingale, 1851). For her part, Florence was keen to promote Kaiserswerth as a place where women could find useful education. The booklet, published anonymously, begins with a critique of contemporary women's education:



[...] while the intellectual foot has made a step in advance, the practical foot has remained behind. Woman stands askew. Her education for action has not kept pace with her education for acquirement (ibid., p. 3).

As early as 1846, in correspondence with her father, Florence had developed this theme with regard to education in general and it is interesting that she foresaw no easy formula to adjust the imbalance between theory and practice. Her idea was that 'trials must be made, efforts ventured—some bodies must fall in the breach for others to step upon . . .' (Vicinus & Nergaard, 1989, p. 30). It is worth bearing this remark in mind when we examine the origins of the Nightingale School fourteen years later.

Florence Nightingale found no immediate outlet for her new-found training and on her return from Kaiserswerth in 1851 she drafted what she referred to as her 'religion to the working tailors'—a philosophical discourse privately printed ten years later as the three volumes of *Suggestions for thought for searchers after religious truth* (Nightingale, 1860b).

In a semi-autobiographical chapter entitled 'Cassandra', which remains a key text in nineteenth century women's history, she made an impassioned plea for a new type of education: 'Women long for education to teach them to teach, to teach them the laws of the human mind and how to apply them . . .' (Nightingale, 1860b, p. 391). It was typical of Florence Nightingale that, after voicing her idealism, she went on to sound a note of pragmatism: 'and knowing how imperfect, in the present state of the world, such an education must be, they long for experience, but experience followed up and systematized.'

Between 1851 and 1854, she supplemented the practical experience gained at Kaiserswerth by visiting hospitals throughout the United Kingdom and Europe and collecting information. She systematized her experiences by analysing and reflecting on hospital reports and government publications on public health.

When she visited the newly-built Lariboisière Hospital in Paris in 1853 she was favourably impressed by the wards, which were built on the pavilion plan. The wards were specifically designed to admit light and fresh air while allowing 'noxious airs' or 'miasmas' to disperse between the long, narrow ward blocks. Her research into the reduced mortality at the Lariboisière served to confirm the validity of the 'miasma' theory. The theory stated that disease arose spontaneously in dirty and enclosed spaces. Since the 1830s, this had been the basis for improvements in public health in the United Kingdom, such as the laying of sewers and the provision of clean water supplies to the cities. Few of the public health officials or 'sanitary reformers', as they were known, were doctors; many were civil engineers, and Edwin Chadwick, the chief sanitarian of the day, worked in insurance. In 1858, Louis Pasteur identified germs and proved that disease did not simply arise spontaneously. From that point, medical scientists contested the reformers' agenda but, although the sanitarians' premises were wrong, it can still be claimed that their conclusions were correct and that their reforms were valid.

Florence Nightingale's emphasis on hygiene in the Crimean War (1854-56)

and the importance she placed on the nurse's role in the management of the environment can be traced largely to her understanding of the causes of disease. She may be distinguished from the miasmatisists of her day by the idiosyncratic connection she made between her scientific and religious views. She believed that God created miasmatic disease in order that Man should learn its causes through observation, and then prevent its recurrence through management of the environment. Consequently, she believed that nurses, with their responsibility for maintaining hygiene, had a unique opportunity for spiritual advancement, discovering the nature of God by learning his 'laws of health' (Nightingale, 1873). Florence Nightingale considered that she had not been taught about the nature of disease, not even at Kaiserswerth, but had learnt through experience, observation and reflection. Thus, when called upon to organize nurse training, she sought to replicate the conditions by which she had learned the 'self-evident' facts of disease.

It was not until August 1853 that Florence Nightingale took up her first employment, a post which at last gave her the scope to apply her knowledge and training. She became Lady Superintendent of an Institution for Sick Gentlewomen at no. 1 Upper Harley Street in the West End of London, where she remained till the outbreak of the Crimean War. At Harley Street she proved herself a brilliant manager. While working on the wards she was careful to subordinate herself and her staff to the doctors on all matters of treatment, but when dealing with the committee she frequently questioned and on occasion overturned policy in the interest of the patients.

## The Crimean War

Florence Nightingale was looking for a fresh challenge, such as the matronship of a London hospital, when the Crimean War broke out in 1854.

The British hospital arrangements during the war were possibly no worse than they had been when last put to the test during the Napoleonic Wars, forty years earlier. However, society had higher expectations during the Crimean episode and, with front-line reporting in the newspapers, the populace were more aware of developments. A tide of public concern for the soldiers' well-being allowed the Secretary of State for War, Sidney Herbert, to take a radical step. The appointment of Florence Nightingale to superintend a group of nurses was unprecedented. No woman had previously held an official position in the Army and Florence Nightingale was an interesting choice for superintendent of nurses: experienced in nursing and highly intelligent, but hardly cut out to accept the orders of a bungling hierarchy.

She immediately grasped the situation at Scutari, the principal British military hospital. Not wishing to imperil the prospect of reform by alienating the doctors, her first action was to place her nurses under the doctors' orders and to establish a hospital laundry. Within a month she had secured improvements in the upkeep of the wards, provided new bedding and clothing for the soldiers and improved the hospital diets.

In addition to supervising the nursing of the men, she wrote letters on their

behalf, instituted a scheme for remitting money to their families and provided reading rooms and games for the convalescents. Florence Nightingale battled with the military authorities, the purveyor's department and was a thorn in the side of the Superintendent of Army Medical Services. A growing public interest in her endeavours gave her voice an impetus not available to reformers within the ranks of the army. Many of the recommendations made by the Superintendent of Nurses to the Secretary of State for War rapidly emerged as new army regulations.

While Florence Nightingale's administrative genius brought her the respect of Queen Victoria and of many of those in the government, it was the individual care and attention that she gave to the sick and wounded soldiers that engendered the affection of the British people. She reputedly walked the four miles of hospital corridors every night and one grateful soldier recalled how he kissed her shadow as the 'lady with the lamp' passed by. Florence Nightingale became a symbol of hope during what was otherwise a disastrous military campaign.

In November 1855, at the height of Florence Nightingale's Crimean fame, a group of her supporters held a public meeting in London to raise funds to enable her to carry on the reform of civil hospitals upon her return to England by establishing an institute for the training of nurses and hospital attendants. The Nightingale Fund took off with little involvement from its namesake, who was still inundated with the concerns of the war. She was not to turn her attention to it until 1860, and even then she was faced with other pressing concerns.

On her return to England, she was horrified that the government appeared satisfied with its limited enquiries into the disastrous mismanagement which had caused 16,000 deaths from disease against 4,000 deaths from battle. She immediately set about campaigning for a full commission of enquiry and the whole process was not exhausted till 1860.

According to the critical biographer Lytton Strachey, 'Scutari had given her knowledge; and it had given her power too: her enormous reputation was at her back—an incalculable force' (1918). In truth, the Nightingale 'power' was a rather more subtle force than Strachey implied, but it remained irresistible.

## **Military education**

Florence Nightingale used her influence to champion the education of British soldiers and of Army doctors. Despite the limitations of the instructions she received before going to the Crimea, at the height of the war she had pointed to the lack of practical experience of many of the junior surgeons, suggesting that during the conflict itself they should attend lectures on pathology and related subjects. A pathology laboratory was indeed briefly set up at Scutari as a result of her suggestion. Her concern for practical medical education later surfaced as one of the four central areas of the Royal Commission on the Health of the British Army.

Florence Nightingale was instrumental in obtaining the Royal Commission in 1857, and with Sidney Herbert as Chairman and a majority of supporters on the board, she began the work of marshalling her evidence of hospital mismanagement

and collating statistics of mortality. (It was on the strength of her Crimean statistics that Florence was elected the first female Fellow of the Royal Statistical Society in 1860.)

Miss Nightingale laid down plans for military medical education in her privately printed *Notes on matters affecting the health, efficiency and hospital administration of the British Army* (1858). The purpose of the training was clearly grounded in the lessons of the recent war:

[...] whatever amount of scientific information appears to be presented by the civil student on his entrance into the Army, they convey little or no evidence of his practical knowledge. But as his entrance into the Army instantly introduces him into practice, and in a very short space of time submits patients to his charge, it seems necessary that a school of that kind which exercises the pupil in practical knowledge should intervene between his entrance into the army and his regimental service (Nightingale, 1858a, p. 43).

It is a mark of the force of her proposals that they were implemented by senior physicians and surgeons who were veterans of the Crimean campaign. The first Army Medical School in the United Kingdom was established in 1860 at Fort Pitt, Chatham.

Her concern for Army education extended from the teaching of doctors to the provision of educational facilities for the troops. A recent article (Calabria, 1994) on this little known aspect of Florence Nightingale's work has shown that she was in advance of her time in thinking of the common soldier as educable at all. Like many of her contemporaries she was aware of the debilitating effects on the army of drink and of prostitution; she was, however, exceptional in believing that the soldier's condition was to be blamed on the environment, rather than the soldier's nature. She wrote:

I have never been able to join in the popular cry about the recklessness, sensuality, helplessness of the soldier. On the contrary I should say [...] that I have never seen so teachable & helpful class as the Army generally. Give them opportunity promptly & securely to send money home—and they will use it. Give them a School & a Lecture & they will come to it.

Give them a book & a game & a Magic Lanthorn [sic] and & they will leave off drinking (Goldie, 1987, p. 21).

The success of the Scutari reading room encouraged Florence Nightingale to campaign for similar rooms to be introduced in larger army barracks after the war, and she had a measure of success in this venture.

The Crimea gave Florence Nightingale the opportunity to put her ideas to the test, and after the war she felt obliged to publish her account (Nightingale, 1858a; 1858b; 1859). She knew that the opportunity for learning from the lessons of the war had to be consolidated immediately: 'we cannot try this experiment over again for the benefit of enquirers at home, like a chemical experiment. It must be brought forward as a historical example' (McDonald, 1993). If the post-Crimean reforms demanded urgent attention, the reform of nursing did not have the same immediacy.

It was therefore not until 1860, four years after the Crimean War, that she turned to nurse training, the area with which her name is most closely associated.

## The training of nurses

Nurse training in the United Kingdom was not an entirely new idea by the mid-nineteenth century. In antithesis to Charles Dickens' stereotype of the drunken ignorant nurse, before the Crimean War there was a resurgence of nursing sisterhoods, producing many competent and moral nurses. A number of training houses had been founded in the United Kingdom in the 1830s and 1840s as a result of new religious freedoms. St John's House, an Anglican sisterhood established in 1848, trained women for three months to nurse poor, sick people in their own homes. Six St John's nurses accompanied Florence Nightingale to the Crimea, but although she established close friendships with Mary Jones, Superintendent of St John's House and the Reverend Mother Clare Moore, Mother Superior of the Convent of Mercy in Bermondsey, who also supplied Crimean nurses, she insisted on establishing secular nurse training. While devising a scheme for nurse training, she was sensitive to potential opposition. During the Crimean War, claims in the press that certain of the nurses had attempted the religious conversion of soldiers on their deathbeds had nearly upset her mission. The threat of religious controversy was probably an important factor in influencing Florence Nightingale towards secular nurse training.

There were already vociferous opponents of reformed nursing within the hospitals. In 1856, John Flint South, surgeon at St Thomas' Hospital, London, let it be known that he considered nurses needed no more qualifications than housemaids. Nevertheless, in 1859 Florence Nightingale and the Nightingale Fund began negotiations to establish a training school at St Thomas' Hospital. Opposition from certain quarters of the medical profession was inevitable.

It has already been stated that Florence Nightingale preferred to influence policy rather than direct it, yet with regard to the Nightingale School there were other reasons why she may have preferred to keep a lower profile in the school's affairs. The illness which had continued to afflict her since the Crimean War limited her activity; it was therefore logical to delegate the heavy workload of superintending the school to an active hospital matron. Even had she been well, it is doubtful that she would have taught. Her private correspondence indicates that she did not consider herself a successful teacher of women. In December 1861, she wrote, with some exaggeration, to Mary Mohl:

My doctrines have taken no hold among women. Not one of my Crimean following learnt anything from me—or gave herself [...] to carry out the lesson of that war (Vicus & Nergaard, 1989, p. 230).

Also, she was quite clear that the best practitioners made the best teachers:

the writer, who has herself seen more of what may be called hospital nursing, i.e., practical manual nursing, than, perhaps, anyone in Europe, honestly believes that it is impossible to learn it

from any book, and that it can only be thoroughly learnt in the wards of a hospital; and she also honestly believes that the perfection of surgical nursing may be seen practiced by the old-fashioned 'Sister' of a London Hospital as it can be seen nowhere else in Europe (Nightingale, 1860a).

Although Florence Nightingale considered textbooks inappropriate for teaching the 'handicraft' of nursing, she did concede that books could teach the environmental management or sanitary aspects of nursing. The fact that she insisted on each probationer having her own private room in the Nightingale Home for study and reflection shows that she was not just concerned with the practical side of nursing.

The characteristics which distinguished the Nightingale School in its early years were:

- The training school was independent but linked to a hospital;
- The hospital matron had sole authority over the probationer nurses;
- The training school provided a secure 'home' for the probationers;
- The teaching of probationers was by hospital staff: sisters and doctors;
- The probationers were assessed by the sisters and matron;
- The probationers were paid a basic wage during their training;
- The probationers' contract bound them, after their training, to accept a position in a hospital of the Fund's choice and it was the Fund's policy to send out groups of trained nurses to spread the Nightingale system of training to other hospitals.

There were many difficulties associated with the new endeavour. There was a reliance on the sisters who were themselves untrained; the doctors could not have been expected to understand the special requirements of nursing as opposed to medical education; the matron, Mrs Sarah Wardroper, being responsible for nursing in the hospital, used the probationers as extra pairs of hands; and it proved difficult to recruit probationers of a suitable calibre.

According to the historian of the Nightingale Fund, Monica Baly,

There was no sudden beam from Miss Nightingale's lamp; reform came slowly and painfully and what became known as the Nightingale system was not an ideal scheme of Miss Nightingale's devising but a pragmatic experiment and the result of enforced compromise (Baly, 1986, p. 230).

Undoubtedly, Dr Baly's overall assessment is correct: the school's development was not as regular as earlier historians have suggested. Its first decade was particularly difficult. However, the system that began to emerge in the school's second decade was greatly improved, mainly as a result of a series of initiatives by Florence Nightingale herself in the early 1870s. There is little reason to believe that she conceived of nurse training as anything other than an experiment. She knew from bitter experience with the Royal Commission on the health of the army that reforms were not produced by swift victories. From 1872 until her powers began to fade, she kept in close contact with the school's development, getting to know many of the probationers and sending an annual printed address full of practical and moral advice.

Another point which is worth bearing in mind when assessing the success of the Nightingale School is that its reputation spread far and wide, despite its awkward beginnings. This may be attributed in part to the lasting impact of the Nightingale legend, but it was also the product of hard work. Florence Nightingale's cousin, Henry Bonham-Carter, was secretary to the Nightingale Fund from 1861 to 1914 and his dedication helped to ensure that the Nightingale School received recognition for its achievements. By the time of Mrs Wardroper's retirement in 1887, Bonham-Carter was able to proclaim that the school had provided forty-two hospitals with matrons and 520 nurses had completed their training. The school's successes made it easier to recruit probationers of a higher calibre, and in turn better-trained Nightingale nurses began to establish their own nursing schools.

Early migrations of Nightingale nurses to Australia, Canada, India, Finland, Germany, Sweden and the United States of America led to a network of training schools on the Nightingale system. As nursing became a respectable profession for women across the world, Florence Nightingale's lamp became the profession's emblem, symbolizing, on the one hand, the hope given to the Crimean wounded and, on the other, literacy and learning. When the Florence Nightingale International Foundation was established in 1934 to provide a suitable educational memorial to Florence Nightingale, the 'lamp' was naturally its symbol.

## Florence Nightingale's theories

If a beam can be described as coming from Florence Nightingale's lamp, it was in 1882 when she produced two articles for *Quain's dictionary of medicine* entitled 'Nurses, training of' and 'Nursing the sick.' In the first article she expounded for the first time the requirements of an ideal nurse-training school, distilled from the experience of the Nightingale School. One of the essential requirements in a training school was the 'home sister'. Her role was to consolidate the learning from the wards and to oversee the probationers' moral development. The home sister was in effect the first specialist nurse-teacher. It seems surprising that Florence Nightingale had conceived of nurse training in 1860 without the home sister, considering that this could be conducted by the matron, sisters and doctors alone.

Miss Nightingale also set out her theory of learning, with its emphasis on acquiring practical skills:

Observation tells *how* the patient is; reflection tells *what* is to be done; training tells *how* it is to be done. Training and experience are, of course, necessary to teach us, too, *how* to observe, *what* to observe; *how* to think, *what* to think (Nightingale, 1882).

Florence Nightingale considered that once a nurse had 'learned to learn', the process should be continued beyond formal training. On this subject, her views are extraordinarily up to date: 'every five or ten years [...] really requires a second training nowadays' (Seymer, 1954, p. 333). It was hardly surprising that in her old age Florence Nightingale argued against the registration of nurses. It seemed to her that

the attainment of this registered status for nurses would signify a cut-off point in their training. She warned that registered status would lead to conceit and that it was merely mirroring the professional path taken by doctors. She emphasized the separate requirements of a nurse and her particular responsibility for the well-being of the patient which, in her view, was best secured if the nurse regarded her work as a higher calling or a vocation rather than as a profession. Her arguments eventually, and perhaps inevitably, went unheeded.

## Promoter of education

Education entered into almost every area of Florence Nightingale's life. A common thread was her concern that educational methods should be practical and reflect the purposes to which education might be put.

She took a keen interest in the village elementary school near the family home in the county of Derbyshire. She procured books for the school library, but was also keen on learning through other means. Given the rich geology of Derbyshire, she recommended the use of rock and mineral specimens as a prompt for learning in the classroom. This was a far cry from the deathly dull teaching methods of Mr Gradgrind, Charles Dickens' caricature of a Victorian school-teacher.

Her interest in schools extended to the British colonies. Of particular concern to her was the effect of schooling on the health of children. In 1863, with the sponsorship of the Duke of Newcastle, she conducted a statistical survey of 143 colonial schools in Australia, Canada, South Africa and Ceylon (as Sri Lanka was then). She was concerned that European educational methods were not suited to the teaching of native populations. In correspondence with Sir George Grey, governor of New Zealand, she explained:

keeping a certain number of children a great deal of each day in a classroom, cramming and exciting them with formulae, [would be] fatal to a race exposed to it for the first time. In their children it will produce bad health, scrofula, and consumption and is, in reality, death by slow torture (Keith, 1995).

According to Jocelyn Keith, her advice seems to have gone unheeded.

In the late 1860s, Florence Nightingale's attention was drawn to the subject of education in workhouses for the poor. Her trenchant criticism of the punitive regime suffered by the paupers in residence there received widespread acclaim. The thrust of her argument was that paupers should not be punished, but taught to help themselves. Consequently, it was important to establish practical education that would teach manual skills. She was keen to take children away from the workhouse environment and to teach them in the recently established industrial schools.

Through her long-standing friendship with Dr Benjamin Jowett, Master of Balliol College, University of Oxford, she was drawn into questions of higher education. In the 1870s, she had supported the idea of instituting a medal for achievement in statistics, in memory of Adolphe Quetelet, the founder of modern



statistics. In the early 1890s, Jowett revived her interest in promoting statistics and he introduced her to the mathematician Professor Francis Galton. Together, Nightingale and Galton formulated plans for a new Chair of Statistics at Oxford. In a letter to Galton of 7 February 1891 Florence Nightingale suggested that the professorship should address the need for statistics relating to education, penology, workhouses and India. The proposals came to nothing and historians have debated the reasons for the failure. It should be noted that Florence Nightingale's concern for the practical application of statistics to social problems was not shared by the majority of academics at the time. Karl Pearson, the father of modern applied statistics, recognized the virtues of Miss Nightingale's ideas on the subject, so her contribution was not entirely wasted.

## Conclusion

Florence Nightingale once quoted from an address on education delivered at the Universities of St Andrew's and Glasgow, which perfectly reflected her own standpoint: '[...] education is to teach men not to know, but to do' (Nightingale, 1873, p. 576). It would seem fair to judge Florence Nightingale's contribution to education by the practical effect which her reforms had. A letter written to her by Benjamin Jowett should stand as her epitaph:

There was a great deal of romantic feeling about you 23 years ago when you returned home from the Crimea [...] and now you work on in silence, and nobody knows how many lives are saved by your nurses in hospitals; how many thousand soldiers [...] are now alive owing to your forethought and diligence; how many natives of India in this generation and in generations to come have been preserved from famine and oppression and the load of debt by the energy of a sick lady who can scarcely rise from her bed. The world does not know all this or think about it. But I know it and often think about it (31 December 1879).

## Note

1. It was the first of four such experiences which Florence Nightingale recorded in her private diaries.

## References

- Baly, M.E. 1986. *Florence Nightingale and the nursing legacy*. Beckenham, U.K., Croom Helm.
- Calabria, M. 1994. Florence Nightingale and the libraries of the British Army. *Libraries and culture* (Austin, TX), vol. 29, no. 4, p. 367-88.
- Goldie, S.; ed. 1987. *'I have done my duty': Florence Nightingale in the Crimean War, 1854-56*. Manchester, U.K., Manchester University Press.
- Keith, J. 1995. What if they had listened to Florence?: an essay in contrafactus. In: Bryder, L.; Dow, W., eds. *New countries and old medicine*, p. 340-46. Auckland, New Zealand, Pyramid Press.

- McDonald, L. 1993. *Women founders of the social sciences*. Ottawa, Carleton University Press.
- Nightingale, F. 1851. *The Institution of Kaiserswerth on the Rhine, for the practical training of deaconesses, under the direction of the Rev. Pastor Fliedner, embracing the support and care of a hospital, infant and industrial schools, and a female penitentiary*. London, Colonial Ragged Training School. 32 p.
- . 1858. *Subsidiary notes as to the introduction of female nursing into military hospitals in peace and in war. Presented by request to the Secretary of State for War*. London, Harrison & Sons. 133 p.
- . 1858. *Notes on matters affecting the health efficiency and hospital administration of the British Army founded chiefly on the experience of the late war. Presented by request to the Secretary of State for War*. London, Harrison & Sons. 567 p.
- . 1860. *Notes on nursing: what it is and what it is not*. London, Harrison. 70 p.
- . 1873. A 'note' of interrogation. *Fraser's magazine*. May, p. 567-77.
- O'Malley, I.B. 1930. *Florence Nightingale, 1820-56*. London, Thornton Butterworth. 416 p.
- Seymer, L., ed. 1954. *Selected writings of Florence Nightingale*. New York, MacMillan. 397 p.
- Strachey, L. 1918. Florence Nightingale. In: *Eminent Victorians*. London, Chatto & Windus. (Now available from Penguin Books.)
- Vicinus, M.; Nergaard, B., eds. 1989. *Ever yours, Florence Nightingale*. London, Virago Press. 461 p.

## Works by Florence Nightingale

1851. *The Institution of Kaiserswerth on the Rhine, for the practical training of deaconesses, under the direction of the Rev. Pastor Fliedner, embracing the support and care of a hospital, infant and industrial schools, and a female penitentiary*. London, Colonial Ragged Training School. 32 p.
- 1858a. *Notes on matters affecting the health, efficiency and hospital administration of the British army founded chiefly on the experience of the late war. Presented by request to the Secretary of State for War*. London, Harrison & Sons. 567 p.
- 1858b. *Subsidiary notes as to the introduction of female nursing into military hospitals in peace and in war. Presented by request to the Secretary of State for War*. London, Harrison & Sons. 133 p.
1859. *Notes on hospitals: being two papers read before the National Association for the Promotion of Social Science, at Liverpool, in October 1858. With evidence given to the Royal Commissioners on the State of the Army in 1857*. London, John W. Parker & Sons. 108 p.
- 1860a. *Notes on nursing: what it is and what it is not*. London, Harrison. 70 p.
- 1860b. *Suggestions for thought for searchers after religious truth*. 3 vols. London, Eyre & Spottiswoode. (Privately printed.)
- 1863a. *How people may live and not die in India*. London, Emily Faithfull. 11 p.
- 1863b. *Notes on hospitals*. Third edition. London, Longmans. 187 p.
- 1863c. *Sanitary statistics of native and colonial schools and hospitals*. London. 67 p.
1865. *The organisation of nursing in a large town (an account of the Liverpool nurses' training school). With an introduction by Florence Nightingale*. Liverpool. 103 p.
1867. 'Suggestions on the subject of providing, training, and organizing nurses for the sick poor in workhouse infirmaries. In: *Report of the committee on cubic space of*

- metropolitan workhouses with papers submitted to the committee.* Parliamentary Blue Book, p. 64-79.
1871. *Introductory notes on lying-in institutions, together with a proposal for organising an institution for training midwives and midwifery nurses.* By Florence Nightingale. London, Longmans, Green & Co. 110 p.
- 1873a. A 'Note' of interrogation. *Fraser's magazine*. May, p. 567-77.
- 1873b. A sub-'Note' of interrogation. What will our religion be in 1999? *Fraser's magazine*, p. 25-36.
- 1874a. *Life or death in India. A paper read at the meeting of the National Association for the Promotion of Social Science, Norwich, October, 1873. With an appendix 'life or death by irrigation'.* London, Harrison & Sons. 63 p.
- 1874b. *Suggestions for improving the nursing services of hospitals and on the method of training nurses for the sick poor.* 18 p.
1876. *Metropolitan and National Association for providing trained nursing for the sick poor: on trained nursing for the sick poor.* London, Spottiswoode & Co. 12 p.
1879. A missionary health officer in India. In: *Good words*, (three articles) July, August and September 1879, p. 492-96, p. 565-71, p. 635-40.
1882. 'Nurses, training of' and 'Nursing the sick'. In: *Quain's dictionary of medicine*, p. 1038-43; p. 1043-49.
1892. Three letters. In: *Report of the training of rural health missioners and of their village lecturing and visiting under the Bucks County Council: 1891-92.* Winslow, U.K., E.J. French. 50 p.
1893. 'Sick-nursing and health-nursing.' In: *Woman's mission: a series of congress papers on the philanthropic work of women by eminent writers.* London, Sampson Low, Marston & Co., p. 184-99.
1894. *Health teaching in towns and villages. Rural hygiene.* London, Spottiswoode & Co. 27 p.

Bishop and Goldie's *Bio-bibliography* remains the best guide to Florence Nightingale's 200 books, reports, articles and pamphlets. Only the two important new editions that have appeared in English since the publication of *the Bio-Bibliography* are also listed below.

- Bishop, W.J.; Goldie, S. 1962. *A Bio-bibliography of Florence Nightingale.* London, Dawson's of Pall Mall.
- Nightingale, F. 1991. *Cassandra and other selections from suggestions for thought.* (Edited by M.A Poovey.) London, Pickering & Chatto.
- Skretkiewicz, V., ed. 1992. *Florence Nightingale's notes on nursing* (revised with additions). London, Scutari Press.

## Works about Florence Nightingale

- Cook, E.T. 1913. *The life of Florence Nightingale*, 2 vols. London, Macmillan.
- O'Malley, I.B. 1930. *Florence Nightingale, 1820-1856.* London, Thornton Butterworth. 416 p.
- Woodham-Smith, C. 1992. *Florence Nightingale.* London, Constable.

To place your subscription to  
***PROSPECTS:***  
the quarterly review of comparative education

Complete the order form below and send it by fax or post to:

**Jean De Lannoy, UNESCO Subscriptions Service, Avenue du Roi, 202, 1190 Brussels, Belgium.**

Tel.: (32) 2-538.43.08; Fax: (32) 2-538.08.41. E-mail: [jean.de.lannoy@infoboard.be](mailto:jean.de.lannoy@infoboard.be)

Internet: <http://www.jean-de-lannoy.be>

YES, I would like to subscribe to *Prospects*, UNESCO's quarterly review of comparative education.

*Language edition:*

☐ English edition

☐ French edition

☐ Spanish edition

*Annual subscription rates:*

☐ Individuals or institutions in developed countries, **180 French francs** (single issue: 60FF)

☐ Individuals or institutions in developing countries, **90 French francs** (single issue: 30FF)

I enclose payment in the form of:

☐ a cheque in French francs, drawn on a bank with branches in France, made out to 'DE LANNOY—UNESCO Subscriptions';

☐ Visa/Eurocard/Mastercard/Diners/American Express credit card:

No.: ..... Expiry date: .....

☐ UNESCO coupons.

Name:

.....

Address:

.....

.....

(Please type or print clearly)

Signature: ..... Date: .....

Further information and enquiries about other language versions of *Prospects*:

International Bureau of Education, PUB, P.O. Box 199, 1211 Geneva 20, Switzerland.

E-mail: [j.fox@ibe.unesco.org](mailto:j.fox@ibe.unesco.org); Internet: <http://www.unicc.org/ibe>

Further information on UNESCO publications from: UNESCO Publishing, 1, rue Miollis, 75732 Paris Cedex 15, France. Tel.: (33) 1-45.68.10.00; Fax.: (33) 1-45.68.57.41;

Internet: <http://www.unesco.org/publishing>

The  
international  
quarterly review  
of comparative education

# P R O S P E C T S

**PROSPECTS** is edited by the International Bureau of Education.

The contents pages of *Prospects* can be viewed on Internet at:

<http://www.unicc.org/ibe/Publications>

**'Open files' for PROSPECTS in 1998—Volume XXVIII:**

- No. 1, March 1998: *The evaluation of education systems*
- No. 2, June 1998: *Learning to live together through the teaching of history and geography*
- No. 3, September 1998: *Higher education for the twenty-first century*
- No. 4, December 1998: *The control of discipline in schools*

Annual subscription rates:

- ❑ Individuals or institutions in developed countries, 180 French francs (single issue: 60FF)
- ❑ Individuals or institutions in developing countries, 90 French francs (single issue: 30FF)

All correspondence on subscriptions to *PROSPECTS* should be addressed by fax or post to:  
Jean De Lannoy, UNESCO Subscription Service, Avenue du Roi, 202, 1190 Brussels, Belgium.  
Tel.: (32) 2-538.43.08; Fax: (32) 2-538.08.41.

E-mail: [jean.de.lannoy@infoboard.be](mailto:jean.de.lannoy@infoboard.be)

Internet: <http://www.jean-de-lannoy.be>

Further information and enquiries about other language versions of *Prospects*: PUB, IBE,  
P.O. Box 199, 1211 Geneva 20, Switzerland. E-mail: [j.fox@ibe.unesco.org](mailto:j.fox@ibe.unesco.org)

---

## National distributors of UNESCO publications

---

**AUSTRALIA:** Hunter Publications, 58A Gipps Street, Collingwood, Victoria 3066, (P. O. Box 404, Abbotsford, Victoria 3066), tel.: (61-3) 9417 53 61, fax: (61-3) 9419 71 54; e-mail: jpdavies@ozemail.com.au.

**Gray International Booksellers**, 3/12 Sir Thomas Mitchell Road, Bondi Beach, New South Wales 2026, tel./fax: (61-2) 30 41 16.

*For scientific maps and atlases:* **Australian Mineral Foundation Inc.**, 63 Conyngham Street, Glenside, South Australia 5065, tel.: (61-8) 379 04 44, fax: (61-8) 379 46 34.

**AUSTRIA:** Gerold & Co., Graben 31, A-1011 Vienne, tel.: (+43) 1533 5014, fax: (+43) 15 33 5014-12; e-mail: buch@gerold.telecom.at.

**BAHRAIN:** United Schools International, P. O. Box 726, Bahrain, tel.: (973) 23 25 76, fax: (973) 27 22 52.

**BANGLADESH:** Karim International, G. P. O. Box 2141, 64/1 Monipuri Para, Tejgaon, Farmgate, Dhaka 1215, tel.: 32 97 05, fax: (880-2) 81 61 69.

**BARBADOS:** University of the West Indies Bookshop, Cave Hill Campus, P. O. Box 64, Bridgetown, tel.: 424 54 76, fax: (809) 425 13 27.

**BELGIUM:** Jean De Lannoy, Avenue du Roi 202 Koningslaan, 1190 Bruxelles, tél.: 538 43 08, fax: (32-2) 538 08 41; e-mail:

jean.de.lannoy@infoboard.be; Internet: <http://www.jean-de-lannoy.be>

**BOTSWANA:** Botswana Book Centre, P. O. Box 91, Gaborone.

**BRAZIL:** Fundação Getúlio Vargas, Editora, Divisão de Vendas, Praia de Botafogo 190 - 6º andar, 22.253-900 Rio de Janeiro (RJ), tel.: (21) 536 9195, fax: (55-21) 536 9155. **Books**

**International Livros Comércio Exterior Ltda.**, Rua Pamplona, 724 conj - 67, Cerqueira Cesar, 01405-001 Sao Paulo, SP, tel.: (55-11) 283 5840 & 288 0692, fax: (55-11) 287 1331; e-mail:

booksint@mandic.com.br.

**BULGARIA:** Hemus, 7, Levsky Street, 1000 Sofia, tel.: (359-2) 87 0365, **Kantora Literatura**, Boulevard Rousky 6, Sofia.

**CAMEROON:** Commission nationale de la République du Cameroun pour l'UNESCO, Boîte postale 1600, Yaoundé.

**Librairie des Éditions Clé**, Boîte postale 1501, Yaoundé.

**CANADA:** Renouf Publishing Company Limited, 5369 Canotek Road, Unit 1, Ottawa, Ontario K1J 9J3, tel.: (613) 745 2665, fax: (613) 745 7660, Internet: <http://fox.nstrn.ca/renouf/>. **Bookshops:**

711/2 Sparks Street, Ottawa, Ontario tel.: (613) 238 89 85, fax: (613) 238 60 41 and 12 Adelaide Street West, Toronto, M5H 1L6, tel.: (416) 363 31 71, fax: (416) 363 59 63; **Les Éditions La Liberté Inc.**, 3020, Chemin Sainte-Foy, Sainte Foy, (Québec) G1X 3V6, tél./fax: (418) 658 3763; e-mail: liberte@mediom.qc.ca;

**CHINA:** China National Publications Import and Export Corporation, P. O. Box 88, 16 Gongti East Road, Beijing, 100704, tel.: 506 6688, fax: (861) 506 3101.

**CROATIA:** Mladost, Likica 30/11, Zagreb.

**CYPRUS:** 'MAM', Archbishop Makarios 3rd Avenue, P. O. Box 1722, Nicosia.

**CZECH REPUBLIC:** SNTL, Spalena 51, 113-02 Prague 1; **Artia Pegas Press Limited**, Palac Metro, Narodni trida 25, 110-00 Prague 1; **Intes-Praha**, Slavy Hornika 1021, 15006 Prague 5, tel.: (422) 522 449, fax: (422) 522 443.

**DENMARK:** Munksgaard Book and Subscription Service, Norre Sogade 35, P. O. Box 2148, DK-1016, København K, tel.: (+45) 33 12 85 70, fax: (+45) 33 12 93 87; e-mail:

subscription.service@mail.munksgaard.dk.

**EGYPT:** UNESCO Publications Centre, 1 Talaat Harb Street, Cairo, fax: (202) 392 25 66; **Al-Ahram Distribution Agency**, Marketing Dept., Al-Ahram New Building, Galaa Street, Cairo, tel.: 578 60 69, fax: (20-2) 578 60 23, 578 68 33; **Al-Ahram Bookshops:** Opera Square, Cairo.

**Al-Bustan Center**, Bab El-Look, Cairo; **The Middle East Observer**, 41 Sherif St., Cairo, tel.: (20-2) 3939-732, 3926-919, fax: (20-2) 3939-732, 3606-804, e-mail: fouda@soficom.com.eg.

**ETHIOPIA:** Ethiopian National Agency for UNESCO, P. O. Box 2996, Addis Ababa.

**FINLAND:** Akateeminen Kirjakauppa, Pl 23 (Stockmannsvägen 1H) 00371 Helsinki, tel.: (+358) 0 12141, fax: (+358) 0 1214450. **Suomalainen Kirjakauppa Oy**, Koivuvaarankuja 2, 01640 Vantaa, tel.: (+358) 0 852 751, fax: (+358) 0 852 7900.

**FRANCE:** Librairie de l'UNESCO, 7, place de Fontenoy, 75352 Paris 07 SP, tél.: 01 45 68 22 22 et 1, rue Miollis, Paris 75732 Cedex 15, tél.: 01 45 68 28 48;

**GERMANY:** UNO-Verlag Vertriebs- und Verlags GmbH, Dag-Hammarskjöld-Haus, Poppeldorfer Allee 55, D-53115 Bonn 1, tel.: (0228) 94 90 20, fax: (0228) 21 74 92, Internet: [www.uno-verlag.de](http://www.uno-verlag.de).

**S. Karger GmbH**, Abt. Buchhandlung, Lörracher Strasse 16A, D-W 7800 Freiburg, tel.: (0761) 45 20 70, fax: (0761) 452 07 14; *For scientific maps:* **Internationales Landkartenhaus GeoCenter**, Schockenriedstr. 44, Postfach 800830, D-70565 Stuttgart, tel.: (0711) 788 93 40, fax: (0711) 788 93 54.

**GHANA:** The University Bookshop of Legon, P. O. Box 1, Legon.

**GREECE:** G. C. Eleftheroudakis S.A., 17, Panepistimiou Str., 105 64 Athens, tel.: (301) 331 4180/3, fax: (301) 323 98 21; e-mail: elebooks@netor.gr; Internet:

[www.netor.gr/elebooks/](http://www.netor.gr/elebooks/); **H. Kauffmann Bookshop**, 28, rue du Stade, Athens, tel.: (3) 322 21 60, (3) 325 53 21, (3) 323 25 45. **Greek National**

**Commission for l'UNESCO**, 3, Akadimias Street, Athens. **John Mihalopoulos & Son S.A.**, 75 Hermou Street, P.O. Box 73, Thessalonique, tel.: (031) 27 96 95 and (031) 26 37 86, fax: (031) 26 85 62.

**HONG KONG: Swindon Book Co. Ltd.**, 13-15 Lock Road, Tsim Sha Tsui, Kowloon, tel.: 2366 8001, 2366 8555, fax: (852) 2739 4975. **HUNGARY: Librotrade Ltd.**, **Bookimport**, Pesti út. 237, H-1173, Budapest, tel.: (36-1) 257 77 77, tel./fax: (36-1) 257 74 72.

**ICELAND: Bokabud, Mals & Menningar**, Laugavegi 18, 101 Reykjavik, tel.: (354-1) 552 4242, fax: (354-1) 562 35 23.

**INDIA: UNESCO Regional Office**, 8, Poorvi Marg, Vasant Vihar, New Delhi 110057, tel.: (91-11) 611 00 37, (91-11) 611 00 38, fax: (91-11) 687 33 51 or (91-11) 6872724; **Oxford Book & Stationery Co.**, Scindia House, New Delhi 110001, tel.: (91-11) 331 58 96, 331 53 08, fax: (91-11) 332 26 39; **The Bookpoint (India) Limited**, 3-6-272, Himaytnagar, Hyderabad 500 029, AP, tel.: 23 21 38, fax: (91-40) 24 03 93; **The Bookpoint (India) Limited**, Kamani Marg, Ballard Estate, Bombay 400 038, Maharashtra, tel.: 261 19 72; **Allied Publishers Limited**, 751 Mount Road, Madras 600 002, tel.: 852 3938, 852 3984 and 852 3958, fax: (91-44) 852 0649.

**INDONESIA: PT Bhratara Niaga Media**, Jalan. Oto Iskandardinata III/29, Jakarta 13340, tel./fax: (62-21) 81 91 858.

**IRAN, ISLAMIC REPUBLIC OF: Iranian National Commission for UNESCO**, Shahid Eslamieh Bldg, 1188 Enghelab Avenue, P.O. Box 11365-4498, Tehran 13158, tel.: (98-21) 640 83 55, fax: (98-21) 646 83 67.

**IRELAND: Educational Company of Ireland Ltd**, P. O. Box 43A, Ballymount Road, Walkinstown, Dublin 12, tel.: (353) 1 450 0611, fax: (353) 1 450 0993; **TDC Publishers**, 12 North Frederick Street, Dublin 1, tel.: (353-1) 74 48 35, 72 62 21, fax: (353-1) 74 84 16.

**ISRAEL: R.O.Y. International**, 17, Shimon Hatarssi Street (postal address: P.O. Box 13056), Tel Aviv 61130, tel.: (972-3) 546 1 423, fax: (972-3) 546 1442, e-mail: royi@Netvision.net.il; **Steimatzky Ltd.**, 11 Hakishon Street, P.O. Box 1444, Bnei Brak 51114, tel.: (972-3) 579 4579, fax: (972-3) 579 4567; **Neighbouring territories and countries: INDEX Information Services**, P. O. Box 19502, Jerusalem, tel.: (972-2) 627 16 34, fax: (972-2) 627 12 19.

**ITALY: LICOSA (Libreria Commissionaria Sansoni S.p.A.)**, via Duca di Calabria, 1/1, 50125 Firenze, tel.: (39-55) 64 54 15, fax: (39-55) 64 12 57, e-mail: licosa@ftbcc.it; Internet: <http://www.ftbcc.it/licosa>; and Via Bartolini 29, 20155 Milano, tel.: 2/382 65 083-32 72513; **FAO Shop**, viale delle Terme di Caracalla, 00100 Roma, tel.: 52 25 57 27, fax: (39-6) 5225 33 60, e-mail: publications-sales@fao.org; Internet: <http://www.fao.org>

**JAPAN: Eastern Book Service Inc.**, 3-13 Hongo 3-chome, Bunkyo-ku, Tokyo 113, tel.: (81-3) 3818-0861, fax: (81-3) 3818-0864, e-mail: orders@svt-ebs.co.jp.

**JORDAN: Jordan Distribution Agency**, P. O. Box 375, Amman, tel.: (962-6) 63 01 91, fax: (962-6) 63 51 52; **Jordan Book Centre Co. Ltd**, P. O. Box 301,

Al-Jubeiha, Amman, tel.: (962-6) 67 68 82, 60 68 82, fax: (962-6) 60 20 16.

**KENYA: Africa Book Services Ltd.**, Quran House, Mfangano Street, P. O. Box 45245, Nairobi, tel.: 223641, 330272; **Inter-Africa Book Distributors Ltd**, Kencom House, 1st Floor, Moi Avenue, P. O. Box 73580, Nairobi, tel. (254) 212 764 and 211 183, fax: (254) 213 025.

**KUWAIT: The Kuwait Bookshop Co. Ltd**, Al Muthanna Complex, Fahed El-Salem Street, P. O. Box 2942, Safat 13030, Kuwait, tel.: (965) 242 42 66, 242 42 89, fax: (965) 242 05 58.

**LEBANON: Librairies Antoine S.A.L.**, B. P. 11-656, Beirut, tél.: (961) 1 48 10 72/8 et (961) 1 48 35 13, fax: (961) 1 49 26 25.

**LESOTHO: Mazenod Book Centre**, P. O. Box 39, Mazenod 160.

**MALAWI: Malawi Book Service**, Head Office, P. O. Box 30044, Chichiri, Blantyre 3.

**MALAYSIA: University of Malaya Co-operative Bookshop**, P. O. Box 1127, Jalan Pantai Bahru, 59700 Kuala Lumpur, fax: (603) 755 44 24; **Mawaddah Enterprise Sdr. Brd.**, 75, Jalan Kapitan Tam Yeong, Seremban 7000, N. Sembilan, tel.: (606) 71 10 62, fax: (606) 73 30 62.

**MALTA: L. Sapienza & Sons Ltd.**, 26 Republic Street, Valletta.

**MAURITIUS: Nalanda Co. Ltd.**, 30 Bourbon Street, Port-Louis.

**NETHERLANDS: Roodvelt Import b. v.**, Brouwersgracht 288, 1013 HG Amsterdam, tel.: (31-20) 622 80 35, fax: (31-20) 625 54 93, e-mail: roodbook@euronet.nl; **De Lindeboom Inor Publikaties**, M. A. de Ruyterstraat 20 a, Postbus 202, 7480 AE Haaksbergen, tel.: (53) 574 00 04, fax: (53) 572 92 96. **Kooyker Booksellers**, P. O. Box 24, 2300 AA Leiden, tel.: (31-715) 16 05 60, fax: (31-715) 14 44 39.

*For scientific maps:* **Rudolf Müller**, P. O. Box 9016, NL-1006 AA Amsterdam & Overtoorn 487, 1054 LG Amsterdam, tel.: 616 59 55, fax: (20) 683 86 51.

**NEW ZEALAND: GP Legislation Services**, Bowen State Building, 34 Bowen Street, P. O. Box 12418, Wellington, tel.: 496 56 55, fax: (644) 496 56 98; **Bennetts Government Bookshop** Auckland, Level One, Whitcoulls Corner Store, Queen St, P. O. Box 5513, Wellesley St, tel. (9) 356 5402, fax (9) 356 5409 & Hamilton, Whitcoulls, Centreplace, Bryce St, P. O. Box 928, tel. (7) 839 6305, fax (7) 834 3520; **Bennetts Bookshop**, Palmerston North, 38-42 Broadway Ave, P.O. Box 138, tel. (6) 358 3009, fax (6) 358 2836; **Bennetts University Bookshop**, Palmerston North, Massey University, tel. (6) 354 6020, fax (6) 354 6716; **Bennetts Government Bookshop**, Wellington, Bowen House, Cnr Lambton Quay and Bowen St, P. O. Box 5334, tel. (4) 499 3433, fax (4) 499 3375; **Bennetts Government Bookshop Whitcoulls**, Dunedin, 143 George St, P. O. Box 1104, tel. (3) 4377 8294, fax (3) 477 7869 and 111 Cashel St. Private Bag, Christchurch, tel. (3) 379 7142, fax (3) 377 2529.

**NIGERIA: UNESCO Office**, 9 Bankole Oki Road, Off. Mobolaji Johnson Avenue, Ikoyi, P. O. Box 2823, Lagos, tel.: 68 30 87, 68 40 37, fax: (234-1) 269 37 58.

**NORWAY: NIC Info A/S**, P. O. Box 6512, Etterstad, N-0606 Oslo, tel.: (47) 22 97 45 00, fax:

(47) 22 97 45 45; **Akademika A/S**, P. O. Box 84, Blindern 0314, Oslo, tel.: 22 85 30 30, fax: (47) 22 85 30 53, e-mail: ebryn@sio.uio.no, Internet: <http://www.akademika.no>.

**PAKISTAN: Mirza Book Agency**, 65 Shahrah Quaid-E-Azam, P. O. Box 729, Lahore 54000, tel.: 66839, telex: 4886 ubplk; **UNESCO Publications Centre**, Regional Office for Book Development in Asia and the Pacific, P. O. Box 2034A, Islamabad, tel.: 82 20 71/9, fax: (92-51) 21 39 59 or 82 27 96.

**PHILIPPINES: International Booksource Center, Inc.**, Rm 719, Cityland 10, Condo Tower 2, H.V. de la Costa corner Valero, Makati City, tel.: (632) 817 17 41, fax: (632) 817 96 76.

**POLAND: ORPAN-Import**, Palac Kultury, 00-901 Warszawa; **Ars Polona-Ruch**, Krakowskie Przedmiescie 7, 00-068 Warszawa; **A. B. E. Marketing**, Plac Grzybowski 10/31A, 00-104 Warszawa, tel.: (482) 638 25 60, fax: (482) 666 88 60.

**PORTUGAL: Livraria Portugal**, rua do Carmo 70-74, 1200 Lisboa, tel.: 347 49 82/5, fax: (351) 347 02 64 (*postal address: Apartado 2681, 1117 Lisboa Codex*).

**QATAR: UNESCO Regional Office**, P. O. Box 3945, Doha, tel.: 86 77 07/08, fax: (974) 86 76 44.

**REPUBLIC OF KOREA: Korean National Commission for UNESCO**, P. O. Box Central 64, Seoul, tel.: 776 39 50/47 54, fax: (822) 568 74 54; *Bookshop: Sung Won Building, 10th Floor, 141, SamSung-Dong, KangNam-Ku, 135-090 Seoul.*

**RUSSIAN FEDERATION: Mezhdunarodnaja Kniga**, Ul. Dimitrova 39, Moskva 113095.

**SEYCHELLES: National Bookshop**, P. O. Box 48, Mahé.

**SINGAPORE: Select Books Pte Ltd.**, 19 Tanglin Road No. 3-15, Tanglin Shopping Centre, Singapore 247909, tel.: 732 15 15, fax: (65) 736 08 55, e-mail add: selectbk@cyberway.com.sg.

**SLOVAKIA: Alfa Verlag**, Hurbanovo nam. 6, 893-31 Bratislava.

**SLOVENIA: Cancarjeva Založba, Kopitarjeva 2**, P.O. Box 201-IV, 61001 Ljubljana.

**SOUTH AFRICA: David Philip Publishers (Pty) Ltd**, Cape Town Head Office, P. O. Box 23408, Claremont 7735 (208, Werdmuller Centre, Newry Street, Claremont 7700), tel.: (021) 64 41 36/7/8, fax: (021) 64 33 58, e-mail: dpp@iafrica.com; **Praesidium Books (South Africa)**, 801, 4th Street, Wynberg 2090, Johannesburg, tel.: (011) 887 59 94, fax: (011) 887 81 38, e-mail: pbooks@iafrica.com.

**SRI LANKA: Lake House Bookshop**, 100 Sir Chittampalam Gardiner Mawata, P. O. Box 244, Colombo 2, fax: (94-1) 43 21 04.

**SYRIAN ARAB REPUBLIC: Librairie Sayegh**, Immeuble Diab, rue du Parlement, B.P. 704, Damas.

**THAILAND: UNESCO Regional Office, Prakanong Post Office**, Box 967, Bangkok 10110, tel.: 391 08 80, fax: (662) 391 08 66. **Suksapan Panit**, Mansion 9, Rajdamnern Avenue, Bangkok 14, tel.: 281 65 53, 282 78 22, fax: (662) 281 49 47; **Nibondh & Co. Ltd.**, 40-42 Charoen Krung Road, Siyaeg Phaya Sri, P. O. Box 402, Bangkok GPO, tel.:

221 26 11, fax: 224 68 89; **Suksit Siam Company**, 113-115 Fuang Nakhon Road, opp. Wat Rajbopith, Bangkok 10200, (P. O. Box 2 Maharthai, Bangkok 10206), tel.: (662) 225 95 31/2, fax: (662) 222 51 88.

**TRINIDAD AND TOBAGO: Trinidad and Tobago National Commission for UNESCO**, Ministry of Education, 8 Elizabeth Street, St Clair, Port of Spain, tel./fax: (1809) 622 09 39; **Systematics Studies Limited**, 9 Watts Street, Curepe, Tel.: (1809) 645 3475, fax: (1809) 662 5654, e-mail: tobe@trinidad.net.

**TURKEY: Haset Kitapevi A. S.**, Istiklâl Caddesi n°. 469, Posta Kütüsü 219, Beyoğlu, Istanbul; **Dunia Infotel Basın Yayın Habercisne San.**, Istiklâl Cad. N° 469, 80050 Beyoğlu, Istanbul, tel.: (212) 251 91 96, fax: (90-212) 251 91 97.

**UNITED ARAB EMIRATES: Al Mutanabbi Bookshop**, P. O. Box 71946, Abu Dhabi, tel.: 32 59 20, 34 03 19, fax: (971-2) 31 77 06.

**UNITED KINGDOM: (ex HMSO) The Stationery Office Books**, Publications Centre, 51 Nine Elms Lane, London SW8 5DR, (P. O. Box 276, London SW8 5DT), tel. (44) 171 873 9090/0011, fax orders: (44) 171 873 8200, Internet:

<http://www.thesoinfo.gov.uk>. **The Stationery Office Bookshops**: London: 49 High Holborn, WC1V 6HB, tel.: 0171-873 0011 (*counter service only*); Edinburgh: 71 Lothian Road, EH3 9AZ, tel.: 0131-479 314; Belfast: 16 Arthur Street, BT1 4GD, tel.: 0123-223 8451; Manchester: 9-21 Princess Street, M60 8AS, tel.: 0161-834 7201; Birmingham: 68-69 Bull Street, B4 6AD, tel.: 0121-236 9696; Bristol: 33 Wine Street, BS1 2BQ, tel.: 0117-926 4306; The Friary Cardiff, CF1 4AA, tel.: 01222 395548; *For scientific maps: GeoPubs (Geoscience Publications Services)*, 43 Lammas Way, Ampthill, MK45 2TR, tel.: 01525-640 58 14, fax: (44) 1525-640 53 76.

**UNITED REPUBLIC OF TANZANIA: Dar es Salaam Bookshop**, P. O. Box 9030, Dar es Salaam.

**UNITED STATES OF AMERICA: Berman Associates (ex UNIPUB)**, 4611-F Assembly Drive, Lanham, MD 20706-4391, tel.: 1 800-274-4447, fax: (301) 459-0056; **United Nations Bookshop**, General Assembly Building, New York, NY 10017, tel.: (212) 963 76 80, fax: (1-212) 963 4910, e-mail: bookshop@un.org.

**ZAMBIA: National Educational Distribution Co. of Zambia Ltd**, P. O. Box 2664, Lusaka.

**ZIMBABWE: Textbook Sales (Pvt) Ltd**, 67 Union Avenue, Harare; **Grassroots Books (Pvt) Ltd**, Box A267, Harare.

#### UNESCO BOOK COUPONS

Can be used to purchase all books and periodicals of an educational, scientific or cultural character. For full information, please write to: UNESCO Coupon Office, UNESCO, 7, place de Fontenoy, 75352 Paris 07 SP (France).



# Oxford Review of Education

## EDITORS

**David Phillips**, *University of Oxford, UK*

**John Wilson**, *University of Oxford, UK*

Supported by an Editorial Board

The object of the *Oxford Review of Education* is to advance the study of education. It especially wishes to promote the elaboration and evaluation of a body of speculative and empirical theory, the development of which might improve educational practice. The journal publishes papers on the theory and practice of education from scholars throughout the world in a variety of disciplines: philosophy, political science, economics, history, anthropology, sociology, psychology and medicine.

The Editorial Board seeks to provide a common forum and, on occasion, a focal point of controversy for the discussion of research findings, of historical and contemporary issues and of the functioning of educational institutions.

## SUBSCRIPTION RATES

1997 - Volume 23 (4 issues). ISSN 0305-4985.

Institutional rate: EU £144.00; Outside EU £164.00; North America US\$280.00

Personal rate: EU £58.00; Outside EU £62.00; North America US\$108.00

## ORDER FORM

Please send a completed copy of this form, with the appropriate payment, to the address below.

Name .....

Address .....

.....

.....



CARFAX

Carfax Publishing Company, PO Box 25, Abingdon, Oxfordshire OX14 3UE, UK



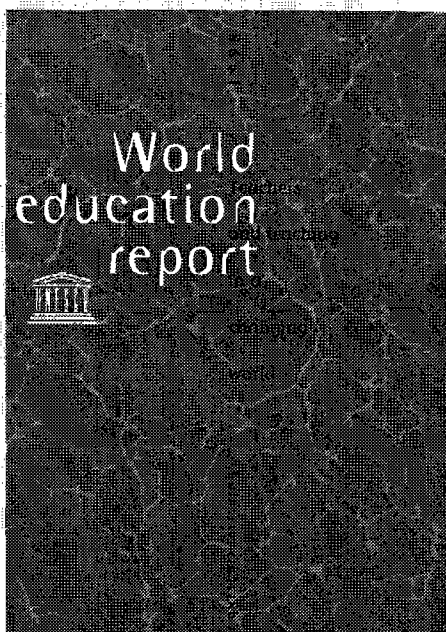
# World Education Report 1998

## Teachers and Teaching in a Changing World

*A UNESCO report on the situation  
and prospects of the world's largest profession*

- the changing status and profile of teachers
- trends in public expenditure on education
- the alarming situation of teachers in the less-developed countries
- the challenges of new technologies in the classroom

Also supplies country-by-country data on key aspects of education in over 180 countries



### World Education Report 1998

4th edition, biennial  
178 pp., tables and figures in colour  
ISBN 92-3-103450-2 / **150 FF**

To order:  
UNESCO Publishing  
7, place de Fontenoy  
75732 Paris 07 SP  
Fax : 01 45 68 57 37  
Internet :  
<http://www.unesco.org/publications>

## A NEW QUARTERLY

# World Heritage Review

**Contribute to protect the planet's most valuable treasures  
while enjoying quality reading**

*World Heritage Review* • Published simultaneously in three separate language editions (English, French and Spanish) by UNESCO Publishing and INCAFO Archivo Fotográfico (Spain) • Quarterly, 80 pages, full-colour high-quality photographs and carefully selected commercial publicity • Feature articles written by specialists with the general public in mind

- Available only on subscription.

*A portion of the income from sales is deposited in the UNESCO World Heritage Fund, created for the safeguard and protection of World Heritage sites.*

### The World Heritage Review

- ❖ An excellent introduction for responsible cultural travellers to the most spectacular destinations
- ❖ An ideal working tool for history, geography, biology or intercultural-communication teachers
- ❖ A prime source of knowledge of other peoples and cultures
- ❖ A genuine documentation source on natural parks, landscapes and sites of exceptional scientific and historical value
- ❖ A must for public and school libraries

#### SUBSCRIPTION FORM

YES, I wish to subscribe to the *World Heritage Review*

Language chosen:

☐ English ☐ French ☐ Spanish

☐ 1 year 140 FF/£18/US\$28 plus airmail charges:  
(4 numbers) European Union countries: 64FF/£8.15/US\$13  
Other European countries: 72FF/£9.15/US\$14.65  
Other countries: 96FF/£12.20/US\$19.55

☐ 2 years 250 FF/£32/US\$51 plus airmail charges:  
(8 numbers) European Union countries: 128FF/£16.30/US\$26  
Other European countries: 144FF/£18.30/US\$29.30  
Other countries: 192FF/£24.40/US\$39

Family name:

First name:

Address:

City/Town:

Postcode:

Country:

Tel/Fax:

Payment enclosed:

- ☐ Cheque (except Eurocheques, and only in French francs)  
☐ Visa ☐ MasterCard  
☐ Eurocard ☐ American Express  
☐ Diners

Card number:

Expiry date:

Date and Signature

Subscription forms and payment (cheques to the order of *Jean De Lannoy*) should be sent to UNESCO Publishing's subscription agent:

**Jean De Lannoy** - Service abonnements

Avenue du Roi, 202

B-1190 Brussels (Belgium)

Tel: (32-2) 538 51 69/538 43 08, fax: (32-2) 538 08 41.

DEP-WHR-9601A



## **Education as a humanitarian response**

**Edited by Gonzalo Retamal and Ruth Aedo-Richmond**

**London, Cassell. ISBN: 0-304-70193-9. Price: £50.00**

With humanitarian emergencies becoming increasingly drawn out, we are now witnessing devastating situations in countries such as Rwanda, Somalia and Bosnia lasting years rather than months.

In this extraordinary book, the contributors (all of whom have been actively engaged in humanitarian crises) outline how best to set up and manage practical education programmes under extreme mental, physical and financial pressures. The book's clear and important message echoes UNESCO's urgent appeal to build a 'culture of peace' which accepts human diversity.

Available from Cassell, Wellington House, 125 Strand, London WC2R 0BB, United Kingdom.

# ONLINE FOR 1998 Comparative Education

## EDITOR

Patricia Broadfoot, *University of Bristol, UK*

Supported by an International Editorial Board

*Comparative Education* presents up-to-date analyses of significant educational problems and trends throughout the world. It seeks to publish articles which use a comparative approach to explore both the formation and implementation of policy and practices in teaching and learning. Articles which are well developed theoretically and empirically are of wide international interest. It draws on a wide range of disciplines in the fields of government, management, sociology – and indeed technology and communications – as these affect educational policy and decisions.

Articles submitted to the editorial board are assessed by all members and are discussed in detail at Board meetings three times a year. In addition, expert advice is taken on particular papers.

## ELECTRONIC ACCESS

In 1998 Carfax is making *Comparative Education* available over the Internet at no extra charge; subscribers to the journal in 1998 will be able to access 1998 issues via the Internet using *RealPage* software with no further payment. This applies to the 1998 volume of the journal only. For further details on how to access the electronic version please visit the Carfax Home Page at

<http://www.carfax.co.uk/cfxnews.htm>

## SUBSCRIPTION RATES

1998 - Volume 34 (3 issues). ISSN 0305-0068.

Institutional rate: EU £244.00; Outside EU £280.00; North America US\$498.00

Personal rate: EU £74.00; Outside EU £94.00; North America US\$168.00

## ORDER FORM

Please send a completed copy of this form to the address below and we will send an invoice to you.

Name .....

Address .....

.....

.....



CARFAX



Carfax Publishing Limited, PO Box 25, Abingdon, Oxfordshire OX14 3UE, UK