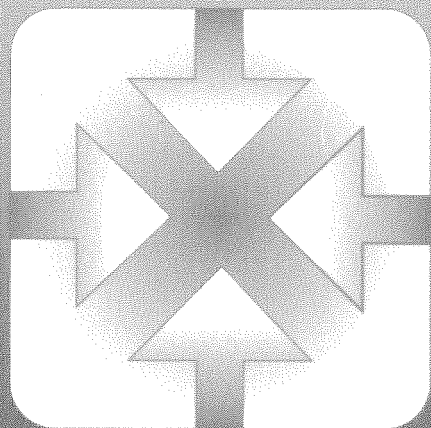# THE ROLE OF RESEARCH IN MATURE EDUCATION SYSTEMS

Edited by Seamus Hegarty

nfer

The role of research in mature education systems

Proceedings of the NFER International Jubilee Conference
Oakley Court, Windsor
2–4 December 1996

Edited by Seamus Hegarty

National Foundation for Educational Research in England and Wales

Contents

List of contributors

Professor **Albert Beaton** is Director of Boston College's Center for the Study of Testing, Evaluation, and Educational Policy and International Study Director of the Third International Maths and Science Study.

Mr **Jarl Bengtsson** is Head of the Centre for Educational Research and Innovation at the Organisation for Economic Cooperation and Development.

M **Gérard Bonnet** is Chargé de Mission, Office for Assessment and Forecasting, Ministry of Education, France.

Professor **Sally Brown** is Deputy Principal and Professor of Education at the University of Stirling, Scotland.

Professor **Susan Fuhrman** is Dean of the Graduate School of Education at the University of Pennsylvania.

Professor **Caroline Gipps** is Dean of Research at the University of London Institute of Education.

Professor **John Gray** is Director of Research at Homerton College, Cambridge.

Professor **Karl-Heinz Gruber** is Professor of Comparative Education at the University of Vienna.

Professor **David Hargreaves** is Professor of Education at the University of Cambridge.

Professor **Wynne Harlen** is Director of the Scottish Council for Research in Education.

Professor **Seamus Hegarty** is Director of the NFER.

Professor **Maurice Kogan** is Professor of Government and Social Administration (Emeritus) and Joint Director of the Centre for Evaluation of Public Policy and Practice at Brunel University.

Professor **Barry McGaw** is Director of the Australian Council for Educational Research.

Ms **Maris O'Rourke** is Senior Adviser Education at the World Bank.

Mme **Luce Pépin** is Head of EURYDICE European Unit, Commission of the European Communities, Brussels.

Professor **Tjeerd Plomp** is Professor of Education at the University of Twente and Chair of IEA.

Professor **Neville Postlethwaite** is Professor of Comparative Education (Emeritus) at the University of Hamburg.

Professor **David Raffe** is Director of the Centre for Educational Sociology at the University of Edinburgh.

Mr **Erhart Schulte** is National Expert, Education and Training DG XII, Commission of the European Communities.

Mr **Ramsay Selden** is President of the Educational Statistics Services Institute in Washington DC.

Professor **Malcolm Skilbeck** is Deputy Director of Education at OECD.

Mr **John Smyth** is Chief Editor of UNESCO's World Education Report.

Mr **Arnold A.J. Spee,** is Senior Consultant, Ministry of Education, The Netherlands.

Dr **Sheila Stoney** is Assistant Director, Evaluation and Policy Studies at the NFER.

Professor **Alejandro Tiana** is Professor Titular of History of Education and Comparative Education at the Spanish Open University.

Mr **Ralph Tabberer** is Assistant Director, Information at the NFER.

Professor **Margaret Wang** is Director of the Center for Research in Human Development at Temple University, Pennsylvania.

Mr **Chris Whetton** is Assistant Director, Assessment and Measurement at the NFER.

Foreword

Peter Mortimore, University of London Institute of Education

These fiftieth anniversary conference papers encompass a wide range of educational topics. They have been written by eminent international scholars and exemplify the contribution made by research to the development of educational ideas. The Foundation should be justifiably proud that its achievements over the past 50 years are being commemorated by such a prestigious publication.

The role of an independent research organisation with a national remit is challenging. Carrying out government commissions while retaining independence and, simultaneously, competing in the marketplace for a range of other work are matters to test to the limits the skills of research management. The close links between the NFER and government agencies, local authorities, universities and international bodies demonstrate that collaboration between policy makers, practitioners and scholars remains a highly prized activity.

The shape of the education systems of the next century can only be guessed at, but the collection of reliable and valid data, the employment of sophisticated analyses and the formulation of well-supported theories will be crucial to societies which have as their prime task the creation of knowledge. The Foundation, with such an established world-wide network of distinguished 'graduates', will be well placed to make a major contribution to such a future.

As the head of the institution from which the Foundation grew and as a successor to Sir Fred Clarke, who played such an important role in the early days of its life, I am pleased to extend the congratulations of my colleagues at the Institute of Education to all those who have contributed to the successful research endeavours of the past 50 years. We wish the current Director — who is continuing to foster links with the Institute through his position as a Visiting Professor of Education — and all his staff even greater success in the future.

Acknowledgements

1.

Educational research in context

Seamus Hegarty, NFER

It may have seemed presumptuous to some that a modest research body like NFER, situated on one of Europe's offshore islands, should propose to celebrate its fiftieth birthday with an event like this. There are in fact two reasons why we have done so. The less important one is that we are not *so* modest. With a staff of more than 200 and an extensive programme of work, we are one of the largest research bodies anywhere dedicated to education and training issues. We have a very wide portfolio of research and development, which we seek to disseminate by every means available to us. We have achieved a reputation for being thorough, impartial and willing only to conclude what can be supported by evidence. (It was a particular delight in the course of last year to be accused simultaneously of being left-wing ideologues and government poodles. I wonder how many others can claim such impeccable credentials of impartiality!) So, perhaps we can be allowed a little party on reaching 50.

But that is not the only reason why we are gathered here. Educational research is — once again — going through a period of reflection and challenge and, hopefully, renewal. Our forebears of 50 years ago faced the task of rebuilding a Europe devastated by war. But their task was not just to rebuild; they also sought to create new institutions that responded to different ways of regarding society and to changes in understanding how decisions affecting society at large should be taken, developments ironically that were accelerated by the social and political upheaval of the war years.

In this country — as in others — we owe a debt of gratitude to the educators of the 1940s; despite the enormous challenges they faced, they had the vision and the practical commitment to create a national research capacity in education. We can best discharge that debt by clear-headed analysis of the situation of educational research in our time as we approach a new millennium and a period of unprecedented change in the generation and distribution of knowledge, and by being prepared to follow through the implications of that analysis by action at local, national and transnational levels.

The programme we have put together represents an effort to structure that analysis and to provide a framework for the conversations and reflections

that we hope will take place here. We have assembled a very strong line-up of speakers and respondents. Between us, we encompass a huge diversity of perspectives on, and levels of engagement with, educational research.

As an aside, can I observe what a luxury it is to have 48 hours' focused reflection on what we are centrally about, to be able to stand aside from day-to-day concerns and concentrate attention on the essentials of our craft. I meet with many of you in other contexts where our time is taken up with matters of organisation, procedure, constitution, resourcing and so on. I am sure these are all very important but I imagine I am not the only one who has sat through meetings wondering if this is an appropriate way for a grown-up person to be spending his time.

So I can promise that over these days there will be no motions or calls to order. What there will be, I hope, is focused reflection on the nature and role of educational research at a time when the demands on educational systems — and their underpinning structures such as research — increase in scope and complexity. Research is only one strand in creating excellent schools and in maximising learning opportunities for everybody, but it is an important strand that is sometimes devalued, and it behoves us who have some responsibility for the conduct of research to have a clear view of where it fits in and what it can contribute.

What I propose to do in this paper is to address two matters: the knowledge base that underpins educational action and how research fits into that base; and some of the key challenges currently facing educational research.

I shall be looking at these issues, naturally enough, from the perspective of an independent research centre. This is not simply because that is the perspective I know best or because I want to be a bit provocative at the outset. I also want to broaden the terms on which these debates are usually conducted. There is a widespread assumption that research is conducted just in universities — a situation, incidentally, which I shall argue is not without its difficulties — and the very large body of research conducted outside the university sector is often referred to as an afterthought, if at all.

I would like first to read a short extract, from a government official, on the relevance of educational research:

> *In the present ferment of opinion as to national education, the work of a well-equipped educational intelligence office might do much to provide unity of educational effort, understanding of the needs of different types*

of schools and a clearness of educational aims. In the past, much
public money has been wasted through failure to make a careful study
of educational methods and problems before embarking on schemes
entailing large expenditure. It is probable that great savings would be
effected by the supply of timely information on many educational
subjects for the consideration of those locally interested in the supply
and management of schools. The aim of the writers of such reports
should be to give practical help to educational workers without being
narrowly statistical or doctrinaire... It should be their aim to
disentangle what is valuable from what is obsolete or antiquated in our
English educational traditions and to preserve all that is good in our
present educational arrangements (Board of Education, 1903; quoted
in Taylor, 1973).

This was written in 1903, as part of the resignation statement of the Director of the Office of Special Enquiries and Reports. Aside from its historical interest, you may find it interesting in determining how much things have changed for the better since then!

Knowledge base for educational action

The nature of the knowledge base that underpins educational action and particularly the relationships between the different kinds of knowledge are matters of key concern. These issues have gained a certain attention, even notoriety, in this country in regard to the relationship between research evidence and inspection evidence. Regarding the teaching of reading in inner-city schools, for instance, or the reasons for educational underachievement, inspection evidence has been adduced by the chief inspector of schools to support conclusions quite different from the findings arising out of research.

The issue is much broader than this, however. The requisite knowledge base for intelligent action encompasses a diversity of knowledge types, each with its own epistemology, truth criteria and so on. I shall be arguing that educational research is just one strand of this knowledge base. And for this reason too it is not for research to provide recipes for action or instant solutions to problems. Its function is to contribute to the knowledge base which enables intelligent action to take place. When we lose sight of the complex and dynamic nature of these cognitive structures, we sell research short. It is either made to appear irrelevant or has unrealistic demands made on it — either way, it is made impotent.

If we take as the central question, What knowledge underpins educational action, in the sense of clarifying options, informing decisions, enabling expert action? we must first appreciate how complicated the question is. To answer it, we have to break it down along at least two major dimensions. The first relates to the different actors involved. Research is not just a matter for teachers, however central their role. Aside from the students, who are key actors in any learning process, there are also policy makers, managers, support staff, teacher educators, examiners, inspectors and parents, all of whom need to act intelligently. It is not that each group is entirely separate in terms of the knowledge requirement to underpin appropriate action; there are of course overlaps, but there are also significant differences. A psychologist, an administrator and a teacher will look at the same pupil through different lenses, and indeed the strength of multidisciplinary assessment is precisely that it can bring different disciplinary lenses to focus on a single subject. The key point is that there will be distinctive patterns of knowledge/action relationships depending on the actor's role in the educational drama, as it were.

A second source of differentiation relates to the substantive topics toward which action is directed. Consider a secondary school curriculum team facing three tasks: develop an integrated approach to teaching science; revise the teaching of the history of World War II; and devise an anti-bullying policy for the school. It is clear that the knowledge required to carry out the tasks successfully will be very different in each case.

So the underlying question is not as simple or as unidimensional as is sometimes portrayed.

To make the issues more concrete, I propose to take an example and sketch some of the knowledge inputs that bear on action in relation to it. The topic is the education of pupils with special needs in ordinary schools, what is often and somewhat misleadingly referred to as integration, and I shall look at it from just two perspectives — the policy maker's and the teacher's.

For the policy maker, the questions have to do with where certain pupils should be educated and what kind of education they should be offered. There have been major policy changes in these regards over the past 30 years. In crude terms, we have moved from a situation where some young people were deemed to be ineducable, others could only attend segregated schools, often to be offered an impoverished curriculum in those schools, and yet others were allowed to attend regular schools but little was expected or demanded of them, to a situation where all children are expected to benefit from education and where the regular school and mainstream curriculum frameworks are the

4

options of first choice for everybody. The shift in rhetoric has been dramatic, though not everywhere matched by consonant practice.

For our purposes, the question is, What knowledge base has underpinned these policy changes? What kinds of knowledge have fuelled the drive from segregation to inclusive education? These are sketched out in Table 1, where the left-hand column sets out factors which have sustained the move to inclusive education and the right-hand column outlines the different knowledge bases implicated in each factor.

Table 1: Movement toward inclusive education: motive factors and the underpinning knowledge bases

FACTORS	KNOWLEDGE BASE
New concepts of handicap	Theory
Backlash against categories	Theory Research Pedagogic practice
Improved assessment	Theory R & D Pedagogic and assessment practice
Comprehensive schooling	Ideology Public debate Research Pedagogic experience
Information on practice elsewhere	Visits/descriptions Evaluations
Comparative studies	Research
Human rights awareness	Public debate Campaigning

5

So much for the policy of educating pupils with special needs in ordinary schools. What about the practice? What knowledge inputs does a teacher need in order to teach a class containing pupils with particular learning difficulties? These are set out in Table 2, in terms of key components of good practice in this area and the knowledge inputs that inform them.

Table 2: Teaching pupils with learning difficulties: key components and knowledge inputs

KEY COMPONENTS	KNOWLEDGE INPUTS
Positive attitudes	Experience Research evidence
Existing teaching competence	Professional development Experience
INSET	Research evidence Distilled experience Theory
Assessment instruments	R & D
Differentiated teaching	Research evidence Experience Materials Theory

What these examples illustrate is the multiplicity of knowledge inputs required for intelligent action, whether in policy making or in teaching, and the complicated relationships between them. Research evidence is a recurrent strand in these knowledge inputs. It is a direct input throughout, but it is implicated indirectly as well — in assessment instruments and curriculum materials, in theory testing and development, in the distillation of experience and in the targeting of professional development.

Enough has been said to illustrate the complicated links between research and practice so far as this topic is concerned. This complexity has been illustrated for a single topic only. If one took a different topic and looked at the knowledge requirements in relation to it, the knowledge inputs and the

relationships between them could be quite different. And these differences would be magnified if the perspectives of the different actors — other than policy makers and teachers — concerned with that topic were brought into play.

It is reasonable to conclude that research findings do not have a simple, linear application. Research takes its place alongside other sources of knowledge in a structured knowledge base, and the relationships between research and the other components of the knowledge base depend critically on the substantive topic and the nature of the action being taken.

A further way of understanding this comes from looking at the nature of intelligent behaviour in general. If we can answer the question, How is intelligent behaviour possible in everyday life? we can see more clearly how research fits into the knowledge base for action.

One person who has addressed this topic very suggestively is the Canadian Bernard Lonergan. I cut my philosophical teeth many years ago on his magnum opus *Insight*. His purpose was to interpret and develop scholastic philosophy in the light of twentieth century science and patterns of knowledge use. One of his ideas which always seemed particularly fecund relates to common sense and its unique role in bridging theory and action.

His starting point was to seek an understanding of mathematical and scientific insights, which are relatively well structured and accessible. But people also behave intelligently in everyday life, and a comprehensive account of human understanding must encompass this intelligent behaviour even if it does not exhibit the precision, replicability or generalisability that characterise scientific enquiry.

Lonergan's notion of common sense was that of a 'habitual but incomplete set of insights that is completed with appropriate variations in each situation requiring action' (Lonergan, 1957, p. 180). In other words, common sense or intelligent behaviour in a practical context is the creative enhancement of existing knowledge. Its distinguishing characteristic is the way it combines the generic with the specific. It requires both a body of relevant, general understanding and the capacity to apply it in concrete situations. This entails making a selection of insights appropriate to the situation in question and generating further, situation-specific insights to enable an intelligent response to it. These latter insights are no longer relevant when that situation has passed — though they may of course enrich the individual's general store of knowledge — and the set of insights reverts to its original, incomplete state

ready for the next situation, which may require a different closure. Different situations need to be approached with different sets of insights, which need in turn to be completed in different ways. Intelligent practical behaviour therefore calls for a large repertoire of incomplete sets of insights, skill in selecting appropriately from them and the ability to generate fresh insights that complete the set in an illuminating way.

Figure 1
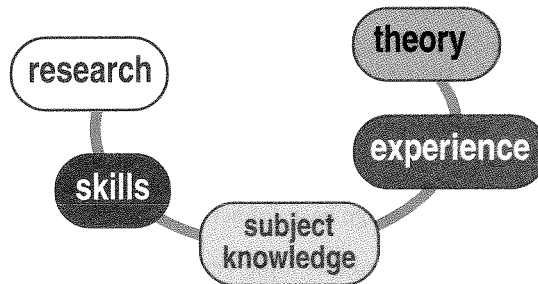
If we apply this to intelligent behaviour on the part of classroom teachers, we have to posit the existence of a number of incomplete sets of relevant insights and competences which would be variously completed in classroom practice. These penannular rings, as it were, are composed of theoretical inputs, experience-based insights, subject knowledge, pedagogical skills and research findings. They derive from formal training, on-the-job learning, personal reflection and general life experience. Figure 1 outlines a schematic pennanular ring as applied to teaching. This is an idealised structure since not every component is present in every case. Figure 2 outlines an instance of successful classroom intervention where the teacher has completed the ring with a fresh insight that combines the various elements into a coherent, practical response.

Figure 2
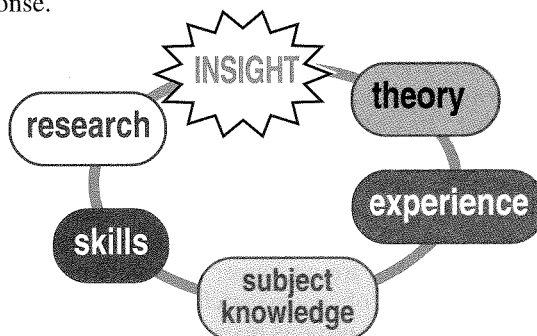
To conclude, what I hope this model offers is an authentic way of conceptualising the relationship between research and practice. Research findings are one of a number of knowledge inputs, none of which is sufficient on its own to sustain intelligent action but all of which are necessary. Research can neither be dismissed as irrelevant nor enlisted in a simplistic, linear way. Its role is to contribute, alongside other elements, to the conceptual repertoire with which teachers approach classroom situations.

Challenges facing educational research

I shall examine briefly here three of the challenges that currently face educational research: how it is organised; the training requirement for research; and research and the public domain.

Organisation of educational research

A starting point here is to ask how educational research is organised in our different countries. Are there a few prevailing patterns, with significant advantages and disadvantages associated with each? Two recent reports, Stoney *et al.* (1995) and OECD (1995), provide some answers, albeit uncomfortable ones. Where one might have expected a few prevailing patterns, adopted for clear reasons or related to significant features of the underlying educational systems, what emerges is a spread of organisational arrangements that is so diverse as almost to defy categorisation.

The United Kingdom is a particularly rich example of diversity. Apart from bodies like the National Foundation for Educational Research and the Scottish Council for Research in Education, research is conducted by university teachers, local authority staff, practitioners such as teachers and psychologists, various official bodies, management consultants, market research organisations, individuals and so on. One could regard this as a creative response to the exigencies of information gathering in a complex system where intelligence needs and responsibility for action are highly diversified. Efficiency in this view is enhanced by democratising the conduct of research and, in a sense, matching it to the structure of educational praxis.

An alternative view would be to regard it as a glorified cottage industry, with lots of small and idiosyncratic producers, some of them undoubtedly doing excellent work, but hardly in their totality constituting a coherent system of knowledge acquisition, distribution and utilisation. It is all very well letting a thousand flowers bloom — but not if you want a top-class garden. Bad

gardeners who cannot tell their weeds from their flowers or their annuals from their perennials and never know when something will come into bloom or what colour it will be when it does, stand to be constantly surprised, but their gardens are likely to be a hotchpotch of colours where everything flowers in June and nothing in September, and there are lots of bare patches and other spots are overcrowded.

If we want coherence and guaranteed quality from research, if we see it as articulating a cumulative understanding of our education systems, it may be necessary to sacrifice some individual enterprise and accept a degree of planning in the disposition of the limited resources available to educational research. The case for organising research in a businesslike way is taken for granted in agriculture, aerospace, defence, medicine, science and so on — and there is no reason to believe that these sectors lose out on creativity because the conduct of research in them is tightly structured.

A big part of the problem in educational research is that it is full of sacred cows, some of them giving all the appearance of sanity. Top of the list for culling are the notion that teaching and researching are integrally related and its corollary that high-quality teaching depends on having a direct engagement in research, i.e. doing it oneself as opposed to using others' findings. These go along with an implicit devaluation of teaching as an activity: unlike the practice of law, medicine or other knowledge-based crafts, excellence in teaching does not seem to be enough on its own without the added patina of research involvement.

This should not be taken as a rejection of practitioner research in principle, but two points may be made about it. First, if research is conducted in a professional way and communicated clearly to appropriate audiences, it does not matter who does it. There is no reason why people cannot do quite different jobs provided they have the requisite expertise, energy and time. It is important though to appreciate that teaching and researching are different skills, and competence in one is no guarantee of competence in the other.

Secondly, we need to be careful in what we designate research. It is tempting to regard as research what is simply good reflective teaching practice (almost as if the research label is required for added credibility). In my early years of teaching, I happened to have two similar classes, so I thought I would try a little experiment. I was covering the same work with both groups, but handled them quite differently: with the first group I sought to be pleasant and agreeable whereas with the second I was a real stickler — the slightest step out of line, and I came down like a ton of bricks. As the year went on, I had

an excellent relationship with the second group and had them working well, but constantly struggled to control and motivate the first. I found this a good learning experience that helped to shape my teaching practice, but was it research? I hardly think so — it was a useful bit of professional development but it would be misleading to go beyond that and style it as research.

A major source of our difficulties in this area derives from the prevailing assumptions in higher education about the mutual interdependence of teaching and research. These assumptions emerged in the German universities in the nineteenth century and gradually became established as a defining characteristic of universities, particularly in North America and Britain. In point of fact, universities were founded originally to transmit established bodies of knowledge and to provide vocational training — for lawyers, doctors, theologians and so on. There was of course a long tradition of scholarship which preceded the focus on research and its elevation into a central organising principle of the university.

I want to argue that this concept of the university needs to be set aside, for a number of reasons but in particular because of the problems its creates with regard to organising research coherently. First of all, it is probably not affordable. Higher education is expanding in all our countries — by more than 50 per cent in the UK from 1989 to 1994 — but the scope for doing research will not expand in step. This disparity alone will make it impossible to have a close alignment between higher education teaching and research activity. Secondly, the assumption that research productivity and teaching effectiveness are linked is not supported empirically. Feldman's (1987) meta-analysis of available studies found little correlation between the two, a conclusion echoed by Neumann (1996). (Fairweather's (1993) study of the reward structure in US colleges and universities is particularly apposite here. It found a strong correlation between pay and research — the more time staff spent on research as opposed to teaching, the higher their income, regardless incidentally of discipline or type of university.) Thirdly, universities of the twenty-first century are going to have to focus much more sharply on their teaching role. A recent IES study for the Committee of Vice-Chancellors and Principals of UK universities concludes: 'Universities are going to have to respond more effectively to the changing demands of students and society. To be successful, they are likely to have to develop more distinctive and differentiated roles to attract different groups of students and meet their distinctive needs' (Connor *et al.,* 1996, p. xvii).

The main point, however, relates to the conditions for conducting high-quality research on a sustained basis. These include:

- critical mass of dedicated researchers
- staff development and career progression
- infrastructure
- dissemination capacity.

There is no reason why these conditions cannot be met in universities, but it would be unrealistic to expect them to be met for every discipline by every university. And as a matter of fact, they are frequently not met at the moment. In a comment on the structural obstacles facing social science and other research within universities, the dean of a major social sciences faculty recently drew attention to 'the poor terms and conditions of service for many contract researchers, lack of investment on research infrastructure, opaque funding regimes within universities ... the lagging commitment in many universities ... to staff development and training' (Goddard, 1996).

If we are concerned to establish and maintain a national research capacity in a cost-effective way, there is no alternative to dedicated research centres. This is the only way of meeting the requirements cited in terms of coherence, critical mass, training, infrastructure and so on. Such centres could be in universities or independent of them; in an ideal arrangement, there would perhaps be some of both to capitalise on the respective advantages of independent and higher education status.

Training

The second challenge I want to consider is the training requirement for research and how it is best met. This is not a simple matter. Most of us here, especially those of a certain age, probably received relatively little formal training in the conduct of research at the outset of our careers. To the extent that we have become expert, it is because of on-the-job learning or mid-career training. Some may argue that this is the best way to do it. It is a system that has served us well and should be perfectly adequate for the next generation of researchers.

There are strong counter-arguments, however. First of all, working conditions now are quite different from what they were in the past. Projects tend to be more tightly specified, shorter and offering less leeway for on-the-job learning. Secondly, the job has become more demanding in technical terms. The methodological armoury available to the researcher, whether it be multilevel modelling or qualitative analysis of textual databases or complex hybrid techniques, is considerably more sophisticated and demands correspondingly

greater expertise. Thirdly, the quality imperative is greater, from sponsors who — rightly — demand high-quality work on every single occasion, from academic and research peers, and from researchers themselves who seek to ensure that their work does make a difference to practice. Finally, there is a greater premium on the communication skills — written and spoken — needed to convey research findings to diverse audiences.

Where are all these skills going to come from? An obvious place to look is the PhD, which has been seen as the traditional route into research. Public funding for doctoral programmes is based on the assumption that they prepare people for research careers.

Their accomplishment of this task must be questioned, however. In terms of providing a systematic, comprehensive training for a research career, many doctoral programmes would do well to be modest in their claims. There are of course a growing number of taught courses, especially at masters level, that focus on research methods. The fact remains, however, that the current training regime is inadequate in terms of producing a professional cadre of career researchers. A common refrain among directors of social science research institutions — the problem is not confined to education — when they get together is the amount of training which is routinely necessary for newly appointed staff.

Doctoral programmes can be excellent in deepening individuals' understanding of a subject area, introducing them explicitly to the structure of knowledge in it and equipping them with heuristic frameworks for tackling new issues. But if the aim is to prepare people for a research career, perhaps the relative isolation of the PhD will not suffice. Just as in teacher education and most other areas of professional training, it may be that substantial periods of time have to be spent in a structured, professional environment. Such environments are best provided by research centres. It does not matter whether these are university-based or independent. What does matter is that they be reasonably large and have a strong focus on the business of doing research.

Research and the public domain

Finally, I want to speak briefly about research and its public context. However privately a research project is pursued, it has an inescapable public dimension. At a simplistic level, what has been discovered cannot be made unknown again — as some of the atomic physicists working on the Manhattan Project in the early 1940s found out to their dismay.

Our problem is rather the opposite: how to get our findings adequately into the public domain. The most immediately visible aspect of this concern is dissemination, which we are regularly told we all do badly. Now, I am sure we could communicate our findings more clearly, though I believe the criticisms are sometimes overstated, not least in the light of the complex knowledge–action relationships I have referred to.

I want, however, to draw attention to two other matters, viz. restrictive research contracts and incomplete reporting of research studies. Much of the sponsored research in this country is subject to contracts which do not permit publication without the express consent of the sponsors, usually a government department or agency. While these gagging clauses are rarely invoked, their existence is not consonant with the public nature of research, particularly research funded by public money. To the best of my knowledge, such restrictions are not common practice in many other countries. A different sort of difficulty is raised by incomplete reporting of studies, whether this is an outcome of the trend toward publishing executive summaries only or of other less easily justified pressures. The possibility of independent scrutiny of the methods by which findings were obtained is essential to scientific verification. Progress in science has depended on this, and it is imperative for bodies who wish to communicate distilled findings to ensure that there is adequate access to the underpinning studies.

My plea here is for better recognition of the public nature of research and the responsibility of all parties — not just researchers — to use their best endeavours to ensure that research findings are placed in the public domain quickly, with adequate detail and without distortion. The disclosure of research findings will sometimes cause difficulties for policy makers or complicate the implementation of favoured initiatives — it would be remarkable, and worrying, if this did not happen from time to time — but these possibilities must never be used to justify keeping research findings out of public debates, any more than findings should be placed in the public arena without being capable of adequate scrutiny.

Conclusion

To conclude, let me recall the familiar observation that the changes we have seen over the past 50 years are as nothing in comparison with the changes we shall see over the next 50. I am sure that if we all met again in 25 years, let alone 50, the context of schooling and public learning will be very different. How exactly it will be different is impossible to say. But it does seem clear

that a number of forces for change are converging — accelerated access to ICT, continually rising expectations driven by consumer demand, professional aspiration and government reform, and economic and labour market pressures.

Whatever the future shapes of institutional teaching and learning, there will be one constant, namely the need for high-quality research and development. In discharging our responsibilities for building up research capacities at national level and beyond, we have to reconcile the opposing forces of competition and cooperation. It is an inescapable fact that many of us here compete for research funding and qualified staff — a situation that is unlikely to change. But our collective capacity in relation to the job to be done is not so great that we can afford to dissipate any of it in sterile territorialism.

The challenges facing us are to capitalise on our diversity while reducing the incoherence of present organisational arrangements, to invest in professional development despite the pressures on research budgets and deadlines, to insist on the public dimension of the work we carry out and, if these were not challenges enough, to find ways of working collaboratively within a competitive climate. Then perhaps we shall ensure that educational research does make its unique contribution to the enhancement of learning, whether in schools and other institutions or in society at large

References

CONNOR, H., PEARSON, R., COURT, G. and JAGGER, N. (1996). *University Challenge: Student Choices for the 21st Century.* Brighton: Institute for Employment Studies.

FAIRWEATHER, J.S. (Coordinator) (1993). *Teaching, Research and Faculty Rewards: a Summary of the Research Findings of the Faculty Profile Project.* Philadelphia: Pennsylvania State University, National Center on Postsecondary Teaching, Learning and Assessment.

FELDMAN, K.A. (1987). 'Research productivity and scholarly accomplishment of college teachers as related to their instructional effectiveness', *Research in Higher Education*, **25**, 3, 277-98.

GODDARD, J. (1996). 'Importance of being managed', *Times Higher Educ.* Suppl., **1219**, 15 March, 13.

LONERGAN, B.J.F. (1957). *Insight: a Study of Human Understanding.* London: Longmans.

NEUMANN, R. (1996). 'Researching the teaching-research nexus: a critical review', *Australian Journal of Education*, **40**, 1, 5-18.

OECD (1995). *Educational Research and Development: Trends, Issues and Challenges.* Paris: OECD.

STONEY, S.M., JOHNSON, F. and GALLACHER, S. (1995). *Review of Government-funded Educational Research and Development in Europe.* Report sponsored by the European Commission. Slough: NFER.

TAYLOR, W. (1973). 'The organization of educational research in the United Kingdom.' In: TAYLOR, W. (Ed) *Research Perspectives in Education.* London: Routledge & Kegan Paul.

2.

Educational research within national education systems

This section offers case studies of research policy and strategy within four contrasting national educational systems: those of The Netherlands, USA, France, and England and Wales. These inputs provide a wealth of narrative and insight into how educational research policy is determined, funded, structured and implemented. As such, they furnished useful contextual and stimulus material which acted as reference points for later conference presentations and helped to provoke discussion amongst delegates. Country presenters were asked to frame their talk around three themes, each of which is briefly outlined below.

First of all, presenters were asked to reflect on the scope and nature of their country's educational research policy, how national policy and research agendas were established and how these were carried forward into strategy. They were particularly invited to describe the roles of central and local government, government agencies, the universities and independent institutes in funding, managing and conducting educational research and to address such issues as the extent of centralisation, diversification and competition within the system and the balance of responsibility between the different types of organisation.

In the second part of their talk, representatives were called upon to examine the main educational research traditions and paradigms in their country and how these had developed. As part of this, they were asked to focus on the relative emphasis placed on theoretical and basic research, applied research, educational evaluation, action research and practitioner-led research and self-evaluation. One or two presenters also chose to comment on the relationship between advice and inspection and educational research activity.

Presenters were then invited to explore the relationship between educational research and educational reform and, associated with this, the crucial issue of the extent to which research had informed policy-making and practice in recent years. Lastly, they were encouraged to move into forecasting mode and speculate on the dominant educational reforms, research priorities and challenges facing their systems as the millennium approached.

National case studies

Educational research within England and Wales
Ralph Tabberer, NFER

Introduction

In this paper I want to achieve three things. First, it is important to convey the essential scale and nature of the educational research endeavour in England and Wales. Second, it is valuable to concentrate on a distinctive, but not unique, feature of this country's system: research in the marketplace. Third, it is appropriate to examine the place of research in the context of school inspection which, in England and Wales, has been generating a body of evidence for schools and for national policy makers. By doing these three things I hope to help ensure, like my colleagues within this session, that what we are talking about in this conference, in terms of the role of educational research in mature systems, is based on the place that research has now in education, as well as our aspirations for its future.

Diversity and complexity

Seamus Hegarty has set out in his paper above that educational research is a diverse system, and that its role and relationships with other areas of knowledge and endeavour are complex. Nobody should be surprised that we come to this view based on our experience of the system in England and Wales. It has for a long time been conventional wisdom to represent educational research in England and Wales as highly diverse in comparison with that in other countries.

There is not a single funding source but many, diverse funding bodies. There are not a few providers, but a large and diverse body of researchers in different organisational settings. There are many routes into educational research in this country. And there are, of course, different definitions of educational research.

This makes it especially difficult not just to describe our system but also to quantify it. It is not that there are so many sources of funding that we cannot do the arithmetic here, despite recent suggestions from international comparative studies of mathematical performance (Keys, Harris and Fernandes, 1996). Rather, estimates vary widely because they work from different assumptions and take into account different parts of our system.

One of the great problems of diversity is fragmentation and, perhaps, missed

opportunities for major work. It is regrettable, for instance, that there has been no exhaustive observational study on primary or secondary school classroom behaviours since the implementation of the National Curriculum in 1988. There have been many small studies, but nothing to provide a satisfactory contrast with the school and teacher effectiveness studies of Galton and Simon (1980) and Mortimore *et al.* (1988). This might not have been true if there were a clearer national strategy, in step with governmental reform.

I concur with the opening argument to this conference that strategy is important but ask that, in any complex system, a strategy should not be considered and developed without first recognising and then marshalling the key forces at work. Where there is complexity, there is also the challenge to understand the different forces at work in our system and to understand the different patterns of behaviour that emerge. Based on that understanding, it might be more possible to foresee a national research strategy. A national strategy will, for example, have to take into account the established body of government agencies, each with its own delegated responsibilities. Managing a national strategy will not be a case of establishing a singular programme of research; in England and Wales, it will be a case of managing the complexity more effectively.

Evidence of our diversity 1: the various sources and levels of funding

Three government departments make a sizeable contribution to educational research and development in England and Wales: the Department for Education and Employment (which prior to 1995 operated as two distinct funding departments as far as education was concerned), the Welsh Office and the Department of Health (Stoney *et al.*, 1995). In 1992, it is estimated that these departments directly contributed approximately 13 million ECU. But this was a golden year, especially due to employment-sourced funding. In 1993, the sum had reduced to less than 6.5 million ECU. It has not exceeded that sum since.

The Economic and Social Research Council (ESRC) too is responsible for disbursing major government funding for educational research in this country, as part of its funding of academic research in the social sciences. The Council largely sponsors university research, although recognised institutes such as NFER can also apply. Its education funding is not easy to fix, however, because of disciplinary overlaps that occur, but there appears to be more than two million ECU per annum for educational research.

Additionally through the Higher Education Funding Council (HEFCE),

university education departments receive government funds both for teaching and research. The main research allocation to departments is determined by their 'research assessment' by HEFCE. Departments are ranked on a 1 to 5 scale, in terms of excellence. This main research element is approximately 40 million ECU per annum and is distributed directly to higher education institutions based on both the volume and quality of departments' research (Wiliam, 1994).

Additionally, specific national agencies fund research and development related to their direct functions. These agencies (or non-departmental public bodies) have increased in number in recent years and now include agencies for curriculum and assessment (the School Curriculum and Assessment Authority and the Authority for Curriculum and Assessment in Wales), inspection (the Office for Standards in Education and the Office of Her Majesty's Chief Inspector for Wales), and teacher training (the Teacher Training Agency), among others. Again, the actual amounts provided for research and development are not easy to define. The smaller bodies expend up to about 250,000 ECU each per annum. The larger body, SCAA, on the other hand, invests very heavily in research and development related to the development of national assessment instruments, which has a development budget of tens of millions of ECUs.

The local education authorities are also a significant, although often overlooked, source of educational research funding. Their contribution to the NFER's dedicated programme of research is now about 1.2 million ECU per annum. They also commission a wide range of individual R & D studies. Then there are charitable and business sources which contribute further to the overall sum. Lastly, and ever-growing in significance, there is the European Commission and a network of other international sponsors.

All of this leads me to be clear that our total sum for educational research is in excess of the 20 million ECU estimated in 1989 (Bassey, 1994) and in the range of 60 to 80 million ECU. The more account that is taken of funding for research and development in national testing, the higher the figure would become. It would be highly misleading, however, to suggest that this is in any sense a singular and strategically planned national investment in research to support educational policy and practice in England and Wales.

Evidence of diversity 2: the research and development providers

The execution of educational research and development is shared among a large number of governmental, non-governmental agency, independent and

university providers. The number of university departments undertaking educational research exceeds 100 (Bassey, op. cit.). The number of research institutes attached to universities is about 50. There are about 40 non-governmental agencies and independent providers. More than 20 governmental agencies are themselves also providers (Stoney *et al.*, op. cit.).

It is appropriate on the occasion of NFER's Jubilee to note the relatively large stake held in this country by the independent organisations, of which NFER is the longest established educational research institution with an annual turnover exceeding 8 million ECUs.

Research priorities

Despite the diversity of our system, the nature of our educational focus and concern in England and Wales is very similar to that elsewhere. The study undertaken by NFER for CIDREE demonstrated that specific topics had special prominence in the European Union states and Switzerland (Stoney *et al.*, op. cit.). There was considerable consistency in the main research and development priorities, notably for:

- educational management

- curriculum

- technology

- teaching and learning

- educational policy

- assessment

- special education

- preparation for working life.

England and Wales provided no exception. Priorities have been pursued relating to improving pupil achievement and the raising of educational standards; the assessment system and curriculum assessment research were significant; and research into school effectiveness and teaching and learning methods featured high on the list. Research into student transition and into vocational education have remained high priorities too. These appear to be common concerns well beyond Europe.

The fact of such common endeavour appears to be evidence of the global dimension of educational debate. In this country, priorities are not set by one

organisation. They are established differently according to the different funding sources. Government departments and specific national agencies providing contract research establish a precise research agenda, usually closely related to policy concerns. The ESRC identifies themes and focus areas in which proposals have to be framed. University researchers supported through HEFCE and wider sources (such as the educational charities) often have more scope to set their own research agendas. Importantly, of course, there is no single research strategy or policy. Nor is there coordination or clear, planned demarcation between major funding bodies.

Key forces underpinning our diverse system

The following are some of the key forces at work underpinning educational research in England and Wales:

research is partly an academic endeavour, in the tradition of scientific speculation, inquiry and specialisation; research is partly a direct servant of policy and development, providing empirical evidence to guide or evaluate progress; research is partly in the domain of the teaching practitioner, where it has a direct part in developing the practitioner's self-knowledge and in informing their action; research is partly a marketplace where businesses compete (and collaborate) for economic advantage.

Many of those attending this Conference will recognise the first three as features of their own systems, and there has been much discussion in the past about the cultural differences which lie at the root of them. The British Educational Research Association (BERA), formed in 1974, represents all these cultures with its membership in this country of university lecturers, contract researchers and some school teachers. It finds itself regularly hosting debate between the cultures.

It is useful here to dwell a little longer on the fourth key force, however, that of research in the marketplace. It is one of the distinctive (although not unique) features of the England and Wales system. It is a feature of substantial importance to an independent body such as NFER.

It is important to observe of England and Wales, that:

a) there is a marketplace for research, established through the funding gained largely by competitive tenders from government departments and their agencies;

b) there is considerable competition within this marketplace, between organisations of different types; and

c) as with any marketplace, some organisations continue to exist by becoming very focused in their expertise and by exploiting relatively narrow market niches, while others seek to continue by creating a distinctive and more broadly based organisational identity and approach.

In his seminal work on competition and the marketplace, Porter (1980) argued that there were three main strategies with which to pursue a competitive market position: focus, differentiation (i.e. the broad but distinct approach), and 'overall cost leadership'. In the educational research marketplace, it is most common to find the first two: there are organisations like NFER which pursue differentiation and others, such as some of the university departments, which pursue focus as their market strategy.

Interestingly, Porter warned that one of the best predictors of failure in long-term competitive strategy would be an organisation's failure to determine whether it is a focused or differentiated approach which it should pursue. Mistakenly, an organisation may attempt both and fall between them.

The reason I have dwelt on this feature of our system is because, in a diverse system, it is necessary to be clear about the influence of key forces. In considering the development and growth of research in mature education systems, Seamus Hegarty has called for a focus on quality rather than on narrow prescription. In addressing quality, however, the research community has to recognise the very different ways in which its different members will define it.

For example, for us to say that the training of educational researchers is a key dimension of the future is critically important. But training will mean very different things to those of us at this conference. To some it will mean a core part of the organisation's competitive strategy for differentiation: the task is to establish (a) skilled, flexible personnel, (b) common ways of working within the organisation, and (c) skills developed according to people's place in the division of labour. To others, concerned with focus, it may mean skilled specialists. To those whose prime concern is for research and action, it may mean technique which is 'good enough', allied to the core capacity for insightful reflection.

Similarly, it is possible to examine differences in the meaning of dissemination. The purpose of dissemination, after all, is not simply to make research accessible to policy makers and practitioners; it is also a part of a market strategy. For a differentiated organisation, it needs to be part of the promotion of a corporate identity; while, for other more focused bodies, it may have to be part of their competitive strategy for establishing their key areas of specialisation. As a community, we learn from each other by understanding these differences, not by ignoring them.

Only by understanding these differences, and by not trying to impose simple superstructures, will the diverse research community find our areas of common cause in improving quality. Those concerned with developing national strategy will also have to deal with the differences created by the marketplace.

A new challenge: inspection and research

There is a further development in England and Wales, which needs to be taken into account in considering the future of research in mature systems. A relatively new development here has been the emergence of inspection as a large-scale source of evidence concerning schools.

In England and Wales, school inspection has become highly systematised for all schools. Over 100 million ECU per annum is spent on inspection in primary and secondary schools. Its prime function is to provide individual schools with judgement and challenge, leading to their action planning for improvement. But the data from observations and judgements in each school also feed a large database containing judgements about school processes across the whole population of schools in England and Wales. There is a strong dimension of classroom observation at the root of each inspection and there are many judgements made about individual lessons and about detailed school processes.

How does research sit with this new body of evidence? It is a very current question and Porter (op. cit.) might point researchers to inspection as a possible market threat: what he would have called the threat of a 'substitute product or service'. The danger is of those producing belts taking over from those who have been manufacturing braces.

But again, the research community must not react too quickly or too negatively to the emergence of a new basis of evidence for practitioners and policy makers. Ultimately, inspection is not research and research, not inspection. Nor is either to be regarded as the higher activity. They are both part of the

diversity and complexity which exist in the world of evidence, policy and practice. In fact, I strongly suspect that they are excellent complements.

Inspection provides judgements from many inspectors, working to a single, written framework of guidance about what constitutes quality and what constitutes evidence. It is building into a national dataset, which will throw much light on relative strengths and weaknesses in our mature system. But further research will be needed: for instance observational research, classroom interaction research, survey research, case study research, sociological research, longitudinal research and action research to provide deeper insights into the nature and cause of these strengths and weaknesses, as well as educational remedies. Again we must relax rather more about, and recognise, the differences so that we can learn from each other and learn better how to combine these complements.

If my estimate of the financial basis for research is right, at between 60 and 80 million ECU, and inspection runs at over 100 million ECU, it is wise to pause too for a moment and acknowledge the emerging strength of a third large source of evidence in England and Wales: school self-evaluation. Through statutory classroom-focused teacher appraisal and through formal school self-review as part of school development planning and target-setting, it is not unreasonable to estimate that an average primary school in this country will expend ten days a year of staff time on school self-evaluation, and an average secondary school correspondingly more. This might translate into a 60 million ECU per annum activity in itself for England and Wales. The place of research in our mature system has to be seen, therefore, alongside these two other major endeavours, equally concerned with judgement, evidence and the pursuit of understanding.

Conclusions

In England and Wales, therefore, there is diverse educational research and, more than that, diverse sources of evidence which guide policy and practice. Each source of evidence has a role in contributing to policy and practice. The consequence of the analysis in this paper is that the best way to secure the role of research is to concentrate on the core issue of its quality, as Seamus Hegarty has suggested, and not on defining too narrowly a singular position and function.

The challenge is to train, disseminate, establish centres of excellence, establish distinct and dependable ways of doing things. The task is to become compelling in ourselves as a research community because it is through

ensuring that our product is strong that we shall guarantee its future value and its future contribution to education.

It may just be that one reason for policy being preoccupied in this country during the 1980s with what Ball (1990) described as 'the messy realities of influence, pressure, dogma, expediency, conflict, compromise, intransigence, resistance, error, opposition and pragmatism', rather than evidence, was because more rational research was perceived as just not good enough to be of greater help. I suspect too that educational research was perceived as potentially reactionary in an era of radical change. On an optimistic end-note, I would say that I do not think the same could be said, either of the quality or the attitude, of so much educational research in England and Wales today.

References

BALL, S. (1990). *Politics and Policy-Making in Education.* London: Routledge.

BASSEY, M. (1994). 'Educational research in the United Kingdom.' In: CALDERHEAD, J. (Ed). *Educational Research in Europe.* Clevedon: Multilingual Matters.

GALTON, M. and SIMON, B. (1980). *Progress and Performance in the Primary Classroom.* London: Routledge & Kegan Paul.

KEYS, W., HARRIS, S. and FERNANDES, C. (1996). *Third International Mathematics and Science Study. First National Report, Part 1.* Slough: NFER.

MORTIMORE, P., SAMMONS, P., STOLL, L., LEWIS, D. and ECOB, R. (1988). *School Matters.* Somerset: Open Books.

PORTER, M.E. (1980). *Competitive Strategy.* London: Penguin.

STONEY, S.M., JOHNSON, F. and GALLACHER, S. (1995). *Review of Government-Funded Educational Research and Development in Europe.* Slough: CIDREE/NFER.

WALDROP, M.M. (1992). *Complexity. The Emerging Science at the Edge of Order and Chaos.* London: Penguin.

WILIAM, D. (1994). 'I'm sorry, but there's not enough money for a third teaching practice visit'. In: REID, I., CONSTABLE, H. and GRIFFITHS, R. (Eds). *Teacher Education Reform.* London: Paul Chapman.

Acknowledgements

Educational research in the USA
Susan Fuhrman, Graduate School of Education, University of Pennsylvania, USA

The paper is organised into three sections, of which the first outlines the two main characteristics of educational research in the USA. The second section discusses the main concerns raised by American researchers and the third assesses what can be done to effect improvements in the system.

Main characteristics of American educational research

Educational research in the USA is highly dispersed, both in terms of the different levels involved (federal, state, district, faculty) and the range of topics covered. For instance, the American Educational Research Association (AERA) has no fewer than 12 divisions, including one ('educational policy') which was added in 1996.

At federal level, research is managed and coordinated by the Office of Educational Research and Improvement, within the Department of Education. The federal government funds a range of Centres, Laboratories, field-initiated studies, ERIC (Educational Research Information Centre), the special education system, as well as senior research offices. The 1997 federal budget is in the region of $350 million. Because of the dispersed nature of educational research activities and funding mechanisms, however, it is difficult to judge the level of contribution from all sponsoring bodies.

The latest figures available from the National Academy of Sciences (for 1992) suggest the following breakdown:

States and localities — $25 million

University faculty members — $400 million

Charitable foundations — $36 million.

There is some evidence to suggest that funding from such 'third sector' organisations as foundations, associations, technical assistance vendors and individuals (such as Walter Annenberg, one-time US Ambassador to the UK) has grown. This is in contrast to state and federal funding which has decreased.

In 1992, the Academy estimated that funding from *all sources* totalled only 0.25 per cent of the total expenditure on education. This was a 68 per cent decline in funding from the 1970s, with resourcing for educational research

forming a much lower level of total budget and experiencing a steeper decline than for other areas of public life. For example, funding for educational research in 1992 stood at 0.8 per cent of total federal funding on education. This compares with research spending accounting for:

- 3.2 per cent of total expenditure on transportation
- 6.9 per cent of total expenditure on agriculture
- 13.6 per cent of total expenditure on health.

Concerns about the US educational research system

There are three particular concerns in the USA, of which under-funding is a primary issue and is partly related to the question of the politicisation of research.

The political agenda

Not unexpectedly, changing administrations have had changing priorities, and this has led Congress to be sceptical about the choice of topics and areas for governmental R & D action in education, and what the pay-offs would be. This has led to changes taking place in the structure of the federal research agency. At first, there was the National Institute of Education, which was set up to be independent. Worried that research did not have sufficient clout under that structure, the Office of Educational Research and Improvement (OERI) was established in 1985 under the Secretary's leadership. In the latest reorganisation of OERI (in 1994), Congress itself established the major research priorities to be addressed by setting up five separate research institutes. These are:

- Achievement, assessment and instruction
- At-risk youth
- Early childhood
- Post-secondary
- Policy.

Each of these is funded at various levels.

Narrow definition of use

The function of research, according to many policy makers, is to have a direct

impact on educational outcomes, with raised student achievement as a key indicator of success. This is clearly very difficult to achieve. Not only do many non-school factors affect achievement, but also the slow speed of research and the way it diffuses through the system make it much more likely that research will have only an indirect effect. Most of the time its impact will be to increase enlightenment in the most general sense. We have, for instance, never reached more than 20 per cent of schools with any knowledge-based reform, and it is much harder to spread reforms which require changes or improvements in adult learning. However, the federal government has funded much of what we know about learning, about how meaning is constructed, about how best to teach reading (this last has had a tremendous pay-off), but policy makers are often slow to acknowledge the usefulness of such work.

Concerns about quality

Because education is subject to faddism and because educators tend to jump on the latest bandwagon, it is easy to blame research for exaggerated claims of effectiveness. Of course, many innovations are chosen without regard to research, but it is true that research is also partly at fault. Longitudinal studies are often believed to be too expensive and we suffer from a lack of replication of results. The premium is on 'new studies' as competitive funding processes highlight the new and the different. Moreover, as the foundations become more important sources of funding, this danger increases. We also under-utilise rigorous controls and methodological approaches; for example, experimental designs are used less than they should be, although it is true that ethics and politics make such designs difficult. However, we do know about how to randomise within the social service sector and we should use this approach more frequently in our research to promote objectivity.

Improving educational research in the future

Inter-disciplinary approaches

First of all, we should make a greater attempt to link research on cognition and instruction to research on organisation and policy. A major reason that reforms, especially knowledge-based reforms, fail to spread is because our understanding of spread and scale fails to take account of the fact that people must learn about how to implement innovation. Much of what we do to spread reforms is simplistic. We remove identified barriers to dissemination and we support efforts, such as demonstration projects, that assume that everyone is equally motivated to learn or that everyone learns in the same

way. This is, of course, naïve. We must think more deeply about the instructional characteristics of educational policy, according to my colleague, David Cohen. We should ask about policies what we would ask about classroom interventions — how well do they work out in practice for teachers and teaching? From this it follows that in America we need more inter-disciplinary research; unfortunately OERI is currently organised in such a way that separates the key topics.

Secondly, we have to educate the managers and implementers of reforms to ask for evidence of success and to put in place appropriate evaluation mechanisms. We now have a series of whole-school reforms called 'brand name' reforms, because of their famous founders. With Slavin's 'Success-for-All' being an honourable exception, most reforms and interventions suffer from a serious lack of evaluation. There is an active market place for the vendors of reforms, but school people are not asking the right questions about innovations' effectiveness and the extent to which they are fitted to their purpose or context. So research plays a small role, if at all, in some of these reforms, but research suffers when improvement does not result. It takes the blame even though it is insufficiently used.

We therefore need to create more 'honest brokers' in order for more research to be conducted on educational reforms, and we must deal with the incentive structure. For example, foundations with millions to invest in some 'big name' reforms have not done much evaluation. It takes a lot of courage to admit that any investment may not have paid off.

Lastly, we must document the use to which research is put very carefully. We must invest in this effort as well as in the research itself so that we can trace usage, demonstrate effectiveness and improve research utilisation over time.

Context and challenges of educational research and policy in France
Gérard Bonnet, Direction de l'Evaluation et de la Prospective, Paris

Introduction (Level 1)

This presentation will depart somewhat from the main topic of educational research in that it will mostly look at one aspect of it, namely educational evaluation. This is because the institution to which I belong is not primarily a research establishment, but is fully integrated within the Education Ministry.

The Direction de l'Evaluation et de la Prospective (DEP) is the Office for Assessment and Forecasting of the French Ministry of Education (MEN). It is responsible, in conjunction with the various partners involved in education, for collecting and analysing all educational data and conducting the evaluation of the French education system. To achieve this, it maintains and refines the information system at the national and local levels. The relevant information is disseminated through weekly, quarterly and yearly publications in order to promote a better understanding of education for policy makers, practitioners and the general public. The work undertaken by the DEP covers the monitoring of all areas of the education system: cost, financing, organisation, pupils' achievement assessment, school effectiveness, classroom teaching practices, evaluation of innovation, etc. It also prepares short- and medium-term forecasts about trends in pupil and student numbers, teaching staff requirements, etc. Finally, it acts as the official French agency in most major international programmes in these areas.

The Office is therefore ideally placed to assess trends in the French education system and the two points I have chosen to discuss here seem to emerge with prominence from what I have been able to observe. It is my belief that they may also be of some relevance outside France. These points exemplify two challenges which, I think, face researchers and policy makers alike: first, the need for educational evaluation to provide, for the benefit of practitioners, robust explanatory elements about professional practices; and second, the need for policy making to integrate research findings more fully.

The need for a qualitative approach to professional practices through evaluation (Level 1)

To introduce the first point, it is useful to start by recording that, as regards the description of her education system, France has managed to set up evaluation procedures which yield reliable data. These procedures can certainly be refined, including from a conceptual point of view, but most people agree that they provide very reasonable *descriptive* information about what is taking place in education. What is missing at this stage are reliable *explanatory* data regarding the outcome of the education system beyond what is explained by input factors such as socio-economic ones. It seems to me that if we accept that the evaluation of education systems is not an end but a means to an end, i.e. the improvement of the quality of education, then such explanatory factors must be brought to light and used to increase performance.

Moving from there to look at developments in the French education system over the past few years, as measured by the evolution of some specific indicators, it seems that an upper limit has been reached in *quantitative* terms (MEN, 1996a). Quantitative indicators which have been steadily rising over the past decade are now starting to level out. In other words, all the quantitative increases which it was possible to achieve — more pupils and students engaged in education for longer periods, more pupils taking and passing the school leaving examinations, more pupils leaving formal education with either general or vocational qualifications, etc. — all of these have been achieved. Yet there remains a small but worrying percentage of people who are not benefiting from the general improvement of educational opportunity. These are the pupils who fall short of reaching minimum standards, those who go through compulsory education without getting any qualifications at all, those who do not meet the basic requirements in *both* reading and reckoning at particular levels — between six and eight per cent of 11-year-olds are in that predicament (MEN, 1996b). It is as though we are now in front of a blank wall which cannot be broken down using only the habitual and conventional tactics of more teaching, more teachers, smaller classes, changes in the national curriculum, and so forth. Everything possible in these areas has more or less been done, given the prevailing economic constraints. Hence the need for a *qualitative* approach. This implies knowledge of current practices leading to changes in professional practices at all levels of the system, but more specifically at school and classroom level. The first step is therefore to establish with a reasonable degree of confidence what actually goes on in the field.

Professional practices in the classroom (Level 2)

Unfortunately, for historical and cultural reasons, the classroom in France remains very much a black box. No one seems to know just what happens there in terms of teaching practices on a nation-wide scale. This is due to strong resistance on the part of teachers, who feel that any systematic inquiry of this type can only be detrimental to what they see as their professional freedom. Inspections of individual teachers do occur but they are very spaced out in time and mostly geared to specific purposes — the management of teachers' careers, for instance — so that there is no national overview. Steps have recently been taken to commission teams of academics actually to sit in on lessons and analyse what takes place, but only very limited work has been carried out so far since one has to rely on the good will of volunteer teachers. Yet an intriguing picture has started to emerge. It appears that there are huge differences in the amount of time devoted by primary school teachers to a particular subject. For instance, despite the fact that there is a national curriculum, some teachers spend four times as long teaching mathematics as others (MEN, 1994). A similar type of investigation, although different in its methodology, has been conducted to try and match 11-year-olds' achievements in mathematics and French as measured by compulsory standardised testing with teaching practices as described in questionnaires filled in by their teachers in a representative sample of 450 first-form classes (MEN, 1996c, 1996d). Again, a few useful clues have been brought to the surface: among other factors, the positive influence on results of the amount of time spent on geometry in mathematics classes and on written work in French classes. This is clearly an area where there is wide scope for research, first to describe prevailing practices then to identify successful practices, even if transferability remains a problem. But before anything can be undertaken on the necessary scale, teachers must be persuaded that such research will not be used to assess their work and this is not easily done. One way of tackling this problem and establishing confidence has been to organise national testing of pupils' achievements in such a way as to make it a useful teaching tool rather than a control mechanism. Typically such testing, which is mandatory for all pupils at three key stages (8, 11 and 16), is organised at the beginning of the year so as not to antagonise teachers. The aim is also to equip them with the necessary information on their pupils to help them devise their teaching for the coming year. But this is quite different from looking at teaching practices, which unavoidably suggests accountability.

Professional practices in the school (Level 2)

Much the same situation prevails with respect to schools. French schools are not subject to systematic external evaluation through inspection procedures on the scale that is normal policy in some countries such as Britain. To make up for this, standards indicators have been provided to schools to help headteachers understand how their schools are run and to compare that with practices in comparable schools regionally and nationally. The possibility for them to construct further indicators tailored to their specific situation has also been arranged. This is to encourage schools to use this type of self-evaluation to devise and assess their school development plans, which they are required by law to set up in order to promote local initiative. The main area where the standard indicators have led to real external accountability is the publication of three performance indicators (MEN, 1996e) based on pupils' access rate to the last year of upper secondary school and access rate to and results of the final examination (baccalauréat). But this can only go so far and ideally a proper system of full and detailed inspection coupled with the use of relevant indicators should be set up. Again, this goes against the grain and remains to be done on a large scale even if in some instances — assessment of the priority education areas (Zones d'Education Prioritaires — ZEP) — this is beginning to fall into place. As is often the case with such matters, external influences may help things along and the latest initiative of the European Commission through the SOCRATES programme to set up *pilot projects on quality evaluation in school education* may provide a useful framework to initiate the necessary acceptance in France.

This will not come about through the good will of practitioners alone. At the other end of the continuum, policy makers also need to alter their professional practices to take into account the reality of evaluation.

The need for policy makers to integrate evaluation in their professional practices (Level 1)

This brings me to my second — and shorter — point about the need for policy makers to make more systematic use of research and evaluation findings before coming to any decisions. By policy makers I do not mean simply ministers and national or local politicians but also senior officials, who in France often enjoy a pretty free hand in making educational decisions but not infrequently demonstrate a certain amount of disregard for evaluation outcomes. Naturally, there have been instances where policies were built on assessment results, in the same way as there have been cases where policies had to be decided on by ministers solely on political grounds, which is part of

the normal democratic process (Bonnet, 1996). What is less understandable is why some more technical — as opposed to political — educational decisions are made in total and occasionally deliberate ignorance of scientific evidence. Policy makers are quick to point out that most educational research, although fascinating, is unfortunately useless for their purpose. They are not always wrong. All of us have heard this argument and sometimes had cause to go along with it. Yet such criticism can hardly be one-sided.

For a couple of years now we have tried in France, in a small way, to start bridging the gap between the policy making and research communities by holding regular seminars on specific topics of interest to both sides in an attempt to find a common language and compatible working practices. Incredibly, the policy makers have proved to have just as much difficulty as the researchers to adapt. More surprisingly still, they seem unable to state clearly what it is exactly that they would require to help them make informed decisions. Is this just lack of a common vocabulary or proof that they never ask themselves the question? It is after all a sad fact that the number of policy decisions — particularly in the field of educational reform — made on the basis of evaluation findings is disproportionately low in comparison with the number of evaluation reports published by the Ministry itself, let alone by outside researchers. Clearly, all the implications of the shortcomings of policy makers in this area have been neither fully explored nor understood in France. At a time when the Prime Minister has identified evaluation of public policies as one of the basic tasks of central government, it is, to say the least, paradoxical that taking evaluation findings into account is so rarely part of the professional practice of senior officials. This no doubt begs uncomfortable questions regarding their training.

The question of the relationship between policy and research is not new and is addressed from time to time in international circles. The OECD devoted a full conference to the subject in Maastricht in September 1995 under the title *Knowledge Bases for Educational Policies.* On that occasion, some participants aired the idea that what is needed is a new breed of knowledge entrepreneurs to act as intermediaries, as interpreters, between the consumers and the producers of knowledge. This presupposes however that policy makers *are* consumers of knowledge, which may not always be as self-evident as it appears, at least in the French context. Be that as it may, this is an idea that should certainly be looked into in detail and which, it would appear, holds some promise for the future. It is certainly an area which some people in France believe is worth exploring.

The next step in that direction will probably have to be to identify such go-betweens. And this may prove rather more difficult than it seems. For what are the characteristics of such people? The easy way would be to define them as researchers with a political sense. But how many academics or even government scientists are truly conversant with the arcane ways in which administrative structures, let alone politicians, operate? Conversely, the less easy answer would be to rely on officials or former officials with a taste for the groves of academe. But how many among those, assuming they were interested, would possess a sufficient knowledge of research work? One thing is certain, though: the 'knowledge entrepreneurs' would have to possess impeccable professional credentials to be accepted by both sides. Even if this problem were solved, there would remain the need to specify the ways in which these people would operate, to define their status. In short, much has still to be thought out in this connection before this worthy idea becomes operational. And I would not be surprised if once again the impetus had to come from some kind of international cooperation between interested parties in order to overcome national reticence.

Conclusion (Level 1)

To achieve necessary changes in professional practices, one needs to look at what people are doing and, better still, to provide instruments to help people realise what they are doing so that they themselves decide that changes are necessary and that they will implement them. This is what has been described in France as the need to develop a 'culture of evaluation' (Thélot, 1993). It is indeed such a culture which has to be promoted if not only practitioners in the classroom or the school but also policy makers are to come round to the view that whatever they are doing could, and in some cases should, be done better in the interest of pupils. It would appear, however, that this situation may not after all be peculiar to France. I was interested recently by an American paper (Puryear, 1995) where the author laments what he calls 'the widespread absence of an assessment mentality in education'. It therefore appears that my two points, the need for a better understanding of professional practices in the school and the classroom leading to an increase in the quality of teaching and education, on the one hand, and the need for policy makers to be more responsive to evaluation findings, on the other, both have in common the prior development of this 'culture of evaluation' or 'assessment mentality' which many people hope will come about. I have every confidence that the occasion of NFER's jubilee conference will provide an ideal opportunity to discuss how this can best be taken forward.

References

BONNET, G. (1996). 'Effects of evaluation procedures on educational policy decisions in France', *International Journal of Educational Research*, **25**, 3.

MINISTERE DE L'EDUCATION NATIONALE (MEN) (1994). 'Etude exploratoire des pratiques d'enseignement en classe de CE2', *Les Dossiers d'Education et Formations*, 44. Paris: Direction de l'Evaluation et de la Prospective.

MINISTERE DE L'EDUCATION NATIONALE (MEN) (1996a). *L'Etat de l'Ecole*, 6. Paris: Direction de l'Evaluation et de la Prospective.

MINISTERE DE L'EDUCATION NATIONALE (MEN) (1996b). 'Les compétences en lecture, en calcul et en géométrie des élèves à l'entrée au CE2 et en 6è', *Note d'Information*, 96-22. Paris: Direction de l'Evaluation et de la Prospective.

MINISTERE DE L'EDUCATION NATIONALE (MEN) (1996c). 'Pratiques pédagogiques de l'enseignement du français en sixième et progrès des élèves', *Note d'Information*, 96-39. Paris: Direction de l'Evaluation et de la Prospective.

MINISTERE DE L'EDUCATION NATIONALE (MEN) (1996d). 'Pratiques pédagogiques de l'enseignement des mathématiques en sixième et progrès des élèves', *Note d'Information*, 96-44. Paris: Direction de l'Evaluation et de la Prospective.

MINISTERE DE L'EDUCATION NATIONALE (MEN) (1996e). '3 indicateurs de performance des lycées', *Les Dossiers D'Education et Formations,* 66. Paris: Direction de l'Evaluation et de la Prospective.

PURYEAR, J. M. (1995). 'International educational statistics and research: status and problems', *International Journal of Educational Development*, **15**, 1.

THELOT, C. (1993). *L'Evaluation du Système Educatif.* Paris: Nathan.

Educational research in the Netherlands in a nutshell

Arnold A.J. Spee, Ministry of Education, Culture and Science, The Netherlands

Introduction

In this paper, a brief overview is given of educational research in The Netherlands. The paper starts with a description of the organisation of educational research. Educational research in The Netherlands tends to be institutionalised. Universities, university institutions and para-university institutions, educational support institutions, the research council and the Ministry of Education, Culture and Science (further referred to as the Ministry) are all involved in specific aspects of educational research.

Next, the content of the educational research is described. In what fields of education is the research carried out? What type of research is undertaken? What are the themes of educational research? What is the relationship between educational research and other sciences?

In the third part, the process of selecting priorities for educational research is outlined. Educational research is part of the government's research policy in general; so-called 'foresight' commissions are installed for various fields of research. These commissions set priorities for research, based on expected future developments in society. On the basis of the recommendations of these foresight commissions, the government sets the priorities in the biannual Science Budget. The procedure for educational research is no different.

The paper ends with a summary and conclusions.

The organisation of educational research

Yearly 100 million Dfl (roughly the equivalent of 45 million ECU) is spent on educational research and development. This amounts to about 0.4 per cent of the educational budget. Over 90 per cent is financed by the Ministry, and the remainder by other ministries, local government bodies, intermediate organisations, employers and worker organisations. The budget is spent by:

- Universities (35 per cent)
- Educational Support Institutions (35 per cent)
- The Organisation for Educational Research/The Netherlands Organisation for Scientific Research (10 per cent)

• The Ministry of Education, Culture and Science (20 per cent).

This is shown in Figure 1, where the size of the circles reflects the relative budget of the organisations, and thereafter the functions of each are described.

Figure 1. The organisation of educational research
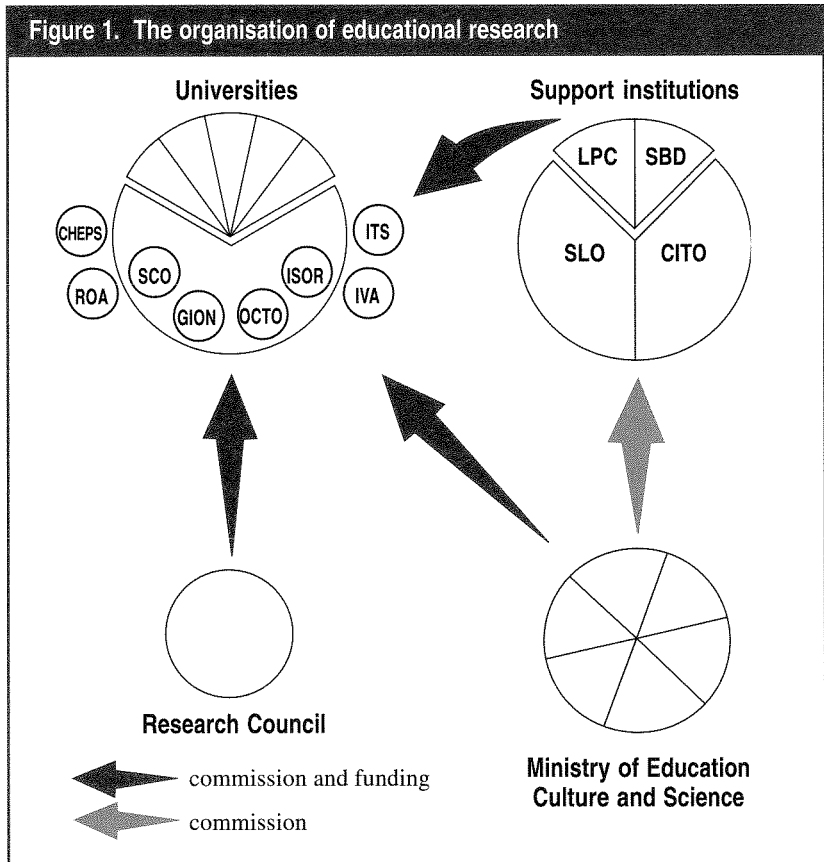
Universities

A Department of Educational Sciences exists in seven universities. The regular research activities are programmed by the universities themselves (the first flow of funds). The second flow of funds is distributed by Research Councils, mainly by the Netherlands Organisation for Scientific Research. The contract research, the so-called third flow of funds, is mainly spent by the Ministry.

The main themes of educational research include:

- optimisation of learning processes;
- learning difficulties;
- social influencing and judgement;
- interactive learning environments;
- development of education and training;
- educational administration and management.

Relatively little attention is paid to teaching methodologies: examples are mathematics (Utrecht), science (Utrecht) and mother tongue (Tilburg). This might be explained by the existence of the National Institute for Curriculum Development, which has a specific role to play in supporting and developing teaching practice.

One inter-university institution for educational research does exist, the Inter-university Centre for Educational Research. Almost all universities with a Department of Educational Sciences participate in this Centre. The main areas of research are: learning and teaching processes, information technology, evaluation of curriculum and school effectiveness and psychometric research.

In four universities the educational research is organised in a separate research institute: the Kohnstamm Institute for Educational Research (SCO, Amsterdam), the Groningen Institute for Educational Research (GION, Groningen), the Research Centre for Applied Research on Education (OCTO, Twente) and the Interdisciplinary Research Institute for the Social Sciences (ISOR, Utrecht). These institutes focus on a specific aspect of educational research.

The para-university institutes are also involved in educational research: the Institute for Applied Social Sciences (ITS, Nijmegen), the Research Centre for Education and the Labour Market (ROA, Maastricht), the Centre for Higher Education Policy Studies (CHEPS, Twente) and the Institute for Social Policy Research and Advice (IVA, Tilburg). These institutes are predominantly financed by contract research.

Apart from the universities and the institutions, educational research is also carried out by a number of commercial and independent organisations.

The educational support institutions

The educational support institutions can be placed in two categories:

- General support services. These include the school advisory services and three national educational advisory centres.

- Special support institutions are the Netherlands Institute for Educational Measurement (CITO) and the National Institute for Curriculum Development (SLO).

The yearly budget for the support institutions is about 200 million Dfl, of which some 30 million Dfl is spent on educational research and development.

Of the *general support institutions*, the School Advisory Services (SBD) form a network of about 60 local and five national agencies, which makes them the largest group of support institutions. The services they provide are mainly targeted at primary schools. Activities are undertaken at the request of schools and range from individual counselling of pupils (e.g. diagnosing learning difficulties) to systems counselling (e.g. defining school regulations). Development activities include policy development, staff development, institutional self-evaluation and the instruction materials development. The School Advisory Services are financed by the municipal authorities and the market at large. The activities of the School Advisory Services are programmed in consultation with the municipal authority, schools and educational organisations.

The three National Educational Advisory Centres (LPC), the Protestant, the Catholic and the non-denominational, represent the interests of particular denominational and non-denominational groups. The work they carry out is roughly the same as that of the School Advisory Services, but mainly targeted at secondary schools. The development of instruction materials is a minor activity of the Centres. The activities of the Centres are programmed by the educational organisations. The Centres have a stronger political influence than the other support institutions.

Of the *specialist agencies*, the National Institute for Curriculum Development (SLO) does not, contrary to what the name may suggest, develop a national curriculum. As the Dutch attach great importance to the 'freedom of education', the Institute is not allowed to develop ready-to-use curricula or teaching packages, only 'models' of curricula and packages. This has placed the Institute in a difficult position, the more so because schools and teachers are not allowed to apply directly to the National Institute, but only through

the formal channels of other educational organisations. It is at the request of these educational organisations that the Institute develops models or alternative models of curricula, school work plans and teaching practices. In addition, the Institute conducts activities at the request of the Ministry, such as the assistance for the implementation of educational innovations.

The task of the National Institute for Educational Measurement (CITO) is to enhance objectivity in the assessment of educational achievement; in other words, to provide support in the field of educational measurement. The Institute develops tests at the request of educational organisations and the Ministry, as well as on its own initiative. The Institute is best known for its primary school-leaving test, commonly referred to as the 'CITO test'. In addition to test development and test administration, the Institute provides various other services, such as the evaluation of school curricula.

The activities of the National Institute for Curriculum Development and the National Institute for Educational Measurement are programmed by the educational organisations (40 per cent), the Ministry (40 per cent) and the institutions themselves (20 per cent). From 1998 onwards, the funds will be distributed by the Ministry on the basis of research proposals submitted by the National Institute for Curriculum Development and the National Institute for Educational Measurement.

The Research Council

Until the end of 1996, a special research council for educational research, the Institute for Educational Research (SVO) existed. SVO fulfilled the following tasks:

- to provide grants for applied educational research at the request of the educational organisations or other (accredited) institutes (40 per cent);

- to advise the Minister of Education, Culture and Science on matters concerning educational innovations and educational reform policies and to help implement educational reforms through educational research institutes (40 per cent);

- to commission fundamental educational research, with a view to furthering the development of education, the development of scientific knowledge and research instruments (20 per cent).

About 80 per cent of the spending of the Institute was comparable with contract research (third flow of funds) and about 20 per cent with pure scientific research (second flow of funds).

At the beginning of the 1990s, the Institute abandoned the distinction between pure and applied research. It was thought that educational research should be more concerned with fundamental dilemmas and problems which are experienced in practice or which are expected to occur in the future. Educational research was programmed in three fields:

- the influence of societal developments on education;
- the quality of different fields of education;
- the evaluation of developments and innovation in education.

From 1997 onwards, the tasks of the Institute for Educational Research have been taken over by the Netherlands Organisation for Scientific Research, the National Educational Advisory Centres and the Ministry. There are several reasons for this decision. Although the Institute for Educational Research had developed a range of standardised procedures with regard to research programming, contracting, assessment and dissemination, this system became so complex that it had unfavourable consequences. Too much time elapsed between the formulation of a question for research and the actual start of the project, not to mention the publication of the outcomes. Furthermore, research proposals are often changed for methodological reasons, which may enhance the technical quality of the research, but which also signify a deviation from the original question to be addressed.

Also, there was a growing tension between the supply and demand side of research. In the 1970s and 1980s the philosophy was that the university institutes needed a certain guarantee of funds, in order to build up their expertise. The governors of the institutes were therefore also the board members of the Institute for Educational Research. So actually, for years the funds of the Institute for Educational Research were divided among the 'big five', the SCO, GION, OCTO, ISOR and ITS. In the meantime, several para-university institutes and commercial institutes came into existence, and for them it was almost impossible to acquire funds from the Institute.

Thus, in order to let the demand factors play a greater role, a large part of the funds was transferred to the Netherlands Organisation for Scientific Research. A Programme Committee was formed, in which half of the members are appointed on the basis of their research experience and the other half on the basis of their social prestige, and a tender procedure for the research grants is being developed. In the transition period from 1997 to 2000, the funds are allocated for 85 per cent, 85 per cent, 68 per cent and 51 per cent in the respective years to the big five educational research institutes, although the former quota for each institute will not be maintained.

The Ministry of Education, Culture and Science

Many departments in the Ministry have their own budget for educational research and related policy research. A large part of the research budget is spent by the Department for Conditions of Employment and Professional Standards to perform research related to the educational labour market. Most of these research projects concern teaching staff in primary and secondary schools. Topics include the causes of absenteeism and early retirement, teacher supply and demand, teaching loads, teacher stress, salary structures and staff funding systems.

The research programme for higher education is developed by the Departments of the Universities and Higher Professional Education, with the advice of the Committee for Policy Research on Higher Education and Scientific Research. The aim of policy-oriented research on higher education is to provide independent, objective information to policy makers about the functioning of the higher education system and the possibilities and effects of controlling that system. Four themes have always formed the basis for the research programme in higher education:

- the dynamism of the higher education system;

- quality indicators and evaluation methods;

- control mechanisms;

- relations between the higher education system and its environment.

The research programme for the vocational education sector is conducted partly by the Department of Vocational and Adult Education, and partly by external institutes:

- The Centre for Education, Training and Work collects quantitative data on adult education. This Centre also conducts research on adult education policies.

- The Max Goote Expertise Centre for Vocational and Adult Education collects, coordinates and distributes expertise concerning vocational and adult education.

- The Centre for Innovation of Vocational Education and Training (CIBB) undertakes both support activities and research-related activities.

- The National Centre for the Study and Development of Adult Education (SVE) conducts research in the field of adult education, particularly

basic adult education. It collects quantitative data on adult education students and conducts survey studies.

The budgets for the Departments for Primary and Secondary Education will increase when they receive a part of the budget that was previously granted to the Institute for Educational Research. Research topics include the functioning of educational establishments, funding of educational institutions, material facilities and administrative–juridical questions. All departments in the Ministry have a coordinator for educational research. These coordinators form a committee to draw up an educational research plan to be submitted to the Minister for approval. Also the Inspectorate carries out educational research projects, on their own initiative or at the request of the Minister.

Coordination of educational research

Out of the budget for educational research, 40 per cent is spent by the universities and the educational support institutes autonomously. The remaining 60 per cent is spent by the Ministry (35 per cent), by the educational organisations (15 per cent) and by the Organisation for Scientific Research (10 per cent).

In order to coordinate the educational research and to prevent overlap, strong relationships have been established between the Ministry, the National Organisation for Scientific Research and the educational organisations. Two members of the Organisation's Programme Committee for Educational Research are also members of the boards of the National Educational Advisory Centres. One person from the Ministry is an observer in the Programme Committee. Twice a year, delegates from the Programme Committee, the National Educational Advisory Centres and the Ministry meet to discuss their respective research plans. In this way, an optimal coordination is ensured.

The content of educational research

Most educational research projects are carried out in primary education (33 per cent) and secondary general education (32 per cent). Research on secondary vocational education accounts for 15 per cent, and that on higher education about 20 per cent, of the educational research undertaken.

Areas and types of educational research

At each level of education the educational research is carried out in all of the main areas of educational science, educational policy and educational practice (see Figure 2).

46

Figure 2. Educational research by area
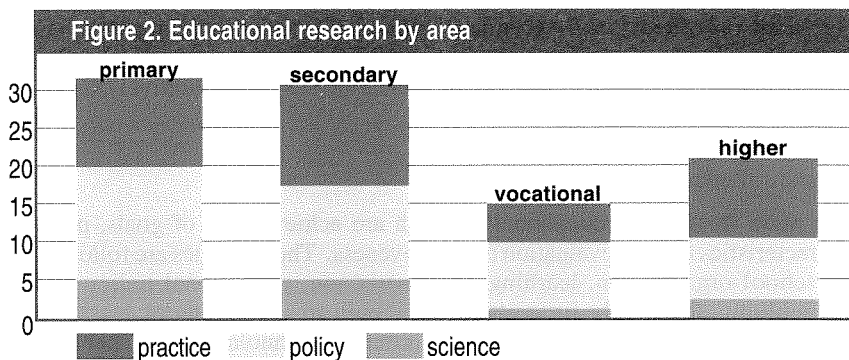
The research on primary and secondary education has a relatively stronger academic orientation than research on other types of education. Research on secondary vocational education is more concerned with advancing the knowledge of educational policy. Research on higher education has a more practical orientation than the research on other types of education.

Apart from a distinction into areas, educational research can be classified according to four types:

- fundamental/strategic research;
- diagnostic research;
- development-oriented research;
- evaluation research (see Figure 3).

Also at each level of education, a different accent is placed on the type of educational research carried out.
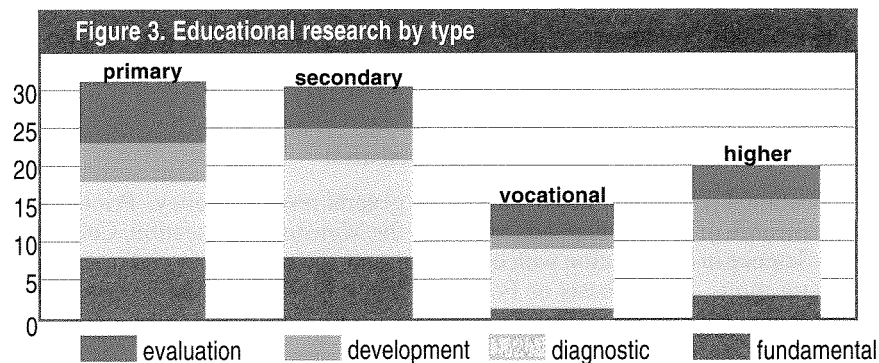
Figure 3. Educational research by type

Research on primary and secondary education is more fundamental than research on levels of education. Research on secondary vocational education is often diagnostic. Research on higher education has a comparatively strong orientation towards development.

Themes of educational research

The main themes of educational research are achievement of goals, pupil characteristics and the evaluation of effectiveness. These themes are followed by school organisation, learning and instruction, educational policy and curriculum. Relatively little attention is paid to conditions of learning, teachers, educational technology and information technology. Forty per cent of the research projects are concerned with a particular school subject. Of these studies, one-third deals with languages, almost one-quarter with mathematics and science and almost one-fifth with vocational skills. Sixty per cent of the research projects are related to social problems. The two most frequently researched problems are work and the labour market and social inequality. Other themes are equal opportunities, health, technological innovation, internationalisation and minorities.

Relationship of educational sciences with other disciplines

Educational problems can be of many different types: pedagogical, didactical, psychological, social, political, etc. Educational research is therefore often carried out not only by educational researchers, but also by, or in combination with, researchers from other disciplines. Apart from educational researchers sociologists and administrative scientists are mainly involved. Psychologists, didacticians, pedagogues, linguists, economists and lawyers play a less important role.

The methods of educational research are mainly case studies and surveys. The analytical methods (models, simulations) and experimental research are less applied.

Selection of priorities for educational research

Educational research policy is part of the general research policy. Priorities for research in The Netherlands are set by the Minister of Education, Culture and Science in consultation with other ministers, the researchers and research organisations, on the basis of Foresight Study Reports. This is shown in Figure 4.

Figure 4. Selection of priorities for research

Foresight Steering Committee
⬇
Foresight Study Report
⬇
Science Budget
⬇
Priorities for research
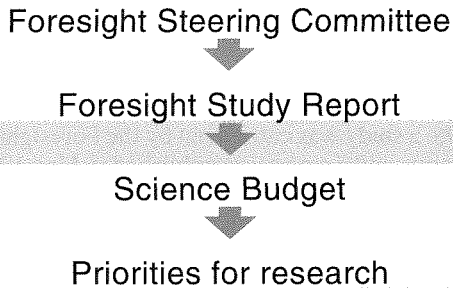
Research policy in general

In 1992, the Minister installed the Foresight Steering Committee (OCV). The task of this Committee is to formulate clear and deliberate options for research, in view of the need for knowledge and know-how in The Netherlands for the next 10 to 15 years. The Steering Committee has carried out 17 Foresight Studies over the past four years. In these studies, not only are the economic and labour market developments taken into account, but also the socio-cultural and environmental aspects. These studies are summarised and published in the report *The Vital Knowledge System: a Perspective of Future Research in The Netherlands*. In this report, clear choices have been made and priorities set for the different research areas.

The government recognised that the priorities set in the report of the Foresight Steering Committee are of prime importance for the future development of science. It took the following actions:

● The report was sent to the universities, the research councils and the research organisations with a request to use the report and the relevant Foresight Studies as a basis for their research strategy for the next ten years.

● In the biannual 'Science Budget 1997', the government set priorities for research, based on the recommendations of the Foresight Steering Committee and asked the research organisations to set targets for research and to concentrate the research funds on these priorities. Because the research organisations are autonomous bodies, a procedure will be developed to reach an agreement on the selection of these priorities.

49

● In the 'Science Budget 1997', the Ministry, together with other ministries involved, indicated for which projects it will provide additional funds to be used as a catalyst for strategic research programmes.

In the Science Budget, educational research competes with other fields of research. One of the 17 Foresight Studies dealt with educational research. From 1994 to 1996, a study was carried out by the Foresight Commission for Educational Research, and a report was published in May 1996. In the report, an analysis is made of future developments relevant for educational research and the recent and expected educational policies and reforms. On the basis of this analysis, priorities are set for educational research over the next ten years.

Future developments in society

The Foresight Commission distinguished five major trends in society and sketched their implications for education.

The first trend is *the globalisation of the economy*. This causes tough and world-wide competition, with the implication that countries that cannot cope with globalisation will not be able to maintain their present standard of living. Countries in western Europe should manufacture products with a high value added and should try to reduce the production costs. Firms should be more flexible and be able to hire or to dismiss personnel at short notice. In the near future, a great part of the labour force will change jobs as well as professions. Strong competition will require high performance from the economy and education. Not only will the government and employers demand high standards in education, but also parents and students. This will be especially evident in higher education, where the private contribution will increase.

Central questions in education will be efficiency and effectiveness. Much attention will be given to efficiency in school careers, the relationship between in-school and out-of-school education and efficiency in school management. Equal attention will be paid to the evaluation of the quality of the educational standards. The professionalisation of the teacher will be of prime importance. The teacher's role will change from an instructor to a counsellor in the learning process. Since knowledge acquired while in school will eventually become obsolete, a system of recurrent education needs to be developed. Only with lifelong learning can the required knowledge be kept up to date.

The second important trend is *the technological development of Communication and Information Technology (CIT)*. This technology has

become part and parcel of the living room already: computer games, tele-banking, Internet, and so on. Most pupils in secondary school have a computer at home. CIT will also play an important role within the school. This faces us with the following questions:

● What are the specific learning effects of these new technologies and can they be more effective and efficient than other learning aids?

● What are the qualities and abilities the teacher must possess in order to make adequate use of these new technologies?

The third trend is *internationalisation*, and in particular integration in the European Union. With a more pan-European labour market, people at work in their own country will come in contact more often with people from other countries. This requires a greater accent on the teaching of languages, especially in secondary education. Europeans who speak a minority language must be able to express themselves in at least one, but preferably more than one, foreign language.

The fourth trend is *the increasing number of people from ethnic minorities*. In the labour market, these minorities are often lagging behind because of their threefold impediments: their low level of education, their different culture and their lack of fluency in Dutch. Education can play a major role in the integration of minority groups in society, but this requires that learning and instruction are geared towards these problems.

The last trend is *increasing individualisation*. In the first instance, this means that education must be tailored to fit the individual person, with inherent differences in instruction if need be. The question is how far this individualisation of school careers, content and method should go. Attention should also be paid to group and social cohesion processes.

Secondly, individualisation will focus the debate on norms and values. In modern society, tradition and shared values have largely disappeared. It becomes the duty of the school to teach the youth to make responsible choices.

Educational reform in the past and in the near future

Educational policy focuses on the problem areas in education that are derived from the trends in society, as mentioned above. In primary and secondary education, the major reforms can be labelled as:

● going to school together;

- combating educational arrears;
- basic secondary education.

In the context of the *going to school together* reform, the number of pupils in special education is growing in primary and secondary education at the expense of regular education. Since 1990, the regular schools have been requested to give special attention to the provision of special aids for pupils that are 'hampered in their development' and to establish relationships with special education institutions. The aim is to give pupils with developmental difficulties more special care in mainstream primary education. The questions to be addressed by educational research are:

- Which children with special educational needs can be realistically educated in regular schools?
- How can the falling of educational standards be prevented in classes with a diversity of pupils?

Both questions are related to the research themes 'learning and instruction' and 'school organisation'.

The policy of *combating educational arrears* is directed to pupils from ethnic and social minorities. Schools with a certain number of pupils from minorities will acquire extra funding for extra personnel. Here two questions have to be raised:

- What changes in the curriculum are necessary in order to combat the educational arrears?
- What organisational conditions must be met in order to make the combating of educational arrears a success?

Both questions are part of the evaluation research that has recently been undertaken.

Basic secondary education was introduced a few years ago. This meant the introduction of a compulsory core curriculum of 15 subjects in the first two or three years of the secondary school, while the streaming of secondary education (pre-vocational education (VBO), junior general secondary education (MAVO), senior general secondary education (HAVO) and pre-university education (VWO)) was left in place. Pupils must reach the same attainment targets, but they may acquire a broader or a deeper knowledge. Discussions are going on now as to whether the attainment targets are too high for pupils in pre-vocational education.

In the later years of secondary education, the so called 'second stage', changes will take place. The curriculum will be distinguished by four profiles: science and health care, science and technology, economics and society and culture and society. Each profile will contain a common section, a compulsory profile section and an optional section. The aim is improve the relationship with higher education. In the second stage of secondary education, the 'study house' concept will be introduced. Increasingly, pupils will have to learn the art of self-study: they will be called upon to develop their own study plans, to work more on their own or in groups and to carry out assignments, projects and the like. The role of the teacher will also change, shifting more and more from instruction to counselling.

These changes are concerned with the organisation of education, with consideration for the heterogeneity of pupils, as well as for motivation, learning capacity and socio-cultural background. For educational research, these questions are related to the themes 'learning and instruction', 'curriculum-content', 'professional teachers' and to the school organisation related to these three areas.

In vocational education, much attention is paid to the required qualifications on the labour market. A new qualification structure is being introduced that consists of:

- qualifications for simple, executable work: assistant training from six months to one year;
- qualifications for non-independent execution of work: basic vocational training from one to three years;
- qualifications for fully independent execution of work: vocational training from two to four years;
- qualifications of fully independent execution of work with a broad working base of specialisation: middle management training from three to four years and specialist training from one to two years.

Setting the priorities for educational research

When looking at the priorities for educational research derived from trends in society and from educational policy, it can be concluded that there is a great deal of overlap. However, some themes are more prominent in relation to the trends in society. These include the integration of communication and information technology (CIT) in the educational process, the teaching of languages and the attention to be paid to norms and values.

53

The Foresight Commission for Educational Research eventually selected two priorities:

- Learning and instruction:

  research themes include principles of design of a powerful learning environment to stimulate the acquisition of knowledge and skills with particular emphasis on the reduction of the inequality of the intellectually and socio-culturally heterogeneous pupil populations;

- Management and organisation of schools and learning:

  research themes deal with the organisation of the instruction process, school autonomy and responsiveness and the mergers of schools; special attention should be paid to conditions of optimal educational outcomes and the flexibility of skills required at the labour market.

Not mentioned as a priority by the Foresight Commission, but considered very important, is the development of communication and information technology in education. The central task for this field of research is to develop the didactic functions of new forms of CIT, such as multimedia, simulation software, telematica and virtual reality.

As stated earlier, the government set priorities for educational research in the 'Science Budget 1997', based on these recommendations of the Foresight Commission for Educational Research. In the Science Budget, the Programme Committee of the Netherlands Organisation for Scientific Research is requested to concentrate educational research in the following areas:

- the relationship between the learning process, the organisation of education and the investment in human capital;

- the integration of new forms of communication and information technology (CIT) in school organisation and the learning process;

- the different forms of adult education, for example lifelong learning.

The utilisation of results of educational research in policy

The government sets the research priorities, but does it make use of the results of the studies?

In 1983, a study revealed several characteristics of the educational research process that severely restricted the potential use of research findings in the policy process. It found that:

- Structural contacts between researchers and policy makers are rare. This is true for each phase in the research process (preliminary phase, implementation and concluding phase). Research supervisory committees are also rare.

- The average scheduled duration of a project is about 18 months, but most projects exceed their time schedule. Phased project planning and structural consultations reduce the risk of projects exceeding their scheduled duration.

- The majority of research projects address educational problems. As a result, policy problems are only partly covered.

- Most research belongs to the one dominant category, the empirical–analytical tradition.

Since this study, progress has been made. A new study of all ministries conducted by the General Audit Office in 1991 showed that the Ministry of Education and Science had invested considerably in evaluation and that it did not compare unfavourably with other ministries in respect of the utilisation of research results.

Some salient examples may be given of the utilisation of research results. One example was the comparative analysis of the salaries of teachers and other civil servants, as a result of which teachers saw their salaries raised. Another example was the report on basic education of the Advisory Council on Government Policy (WRR). This report helped to overcome the political deadlock, which lasted for 20 years, on the organisation of lower secondary education. Less visible, but frequently used, sources are the cohort studies, which monitor the movement of generations of pupils through the educational system. Classic instrumental applications of research findings were the policy decisions concerning the possible extension of vocational training programmes on the basis of labour market analysis. A more recent example was the report on the relationship between class size and pupils' achievement, which led to the proposal to reduce the group size.

The use of research in educational policy remains a complex issue. Educational policy-making seldom consists of simple decisions or choices made from a limited number of alternatives. Major changes generally require new legislation — a long process involving numerous advisory reports, debates and compromises. In some phases of this negotiation process, research may play a part in the analysis of problems and alternative solutions. That is why it is not sufficient merely to list a number of policy recommendations at the

end of each research report. Much more is needed to encourage a particular use of research results in such a complex environment.

Summary and conclusions

Educational research in The Netherlands can be distinguished according to whether it is commissioned educational research (60 per cent) or academic research (40 per cent). Commissioned educational research is oriented towards description and fact finding. The regular university budgets for research are used for the advancement of more theory-oriented academic knowledge.

Educational research is to a large extent institutionalised. This is partly caused by the phenomenon of 'professional compartmentalisation', which has led to the establishment of national institutions for research, development and educational measurement. The Organisation for Educational Research (SVO), the National Institute for Curriculum Development (SLO) and the National Institute for Educational Measurement (CITO) are among the largest institutions of their kind in the world. Furthermore, there are a substantial number of specialised institutes in The Netherlands.

Because of the relatively large size of the budgets for commissioned research, educational research in The Netherlands is strongly market-oriented. Experience in internationally coordinated research programmes shows that Dutch research institutes are very experienced in developing research proposals, including the planning and budgetary aspects, and that they are capable of working very efficiently. They are used to strong competition.

Another consequence of the relatively large flow of funds for contracts is the fairly close link between research programmes and policy programmes. Many national innovations have been accompanied by evaluation research. Cohort studies have been launched in most sectors of education, involving monitoring and analysis of pupils' and students' school careers. Research on educational provisions and curricular content is linked to these longitudinal studies.

Universities in The Netherlands have considerable budgets for educational research (the first flow of funds), so there is no lack of science-driven research. Still, the largest part of their educational research is concerned with problems defined by policy makers and the national school councils.

Is the educational research in The Netherlands different from the educational research carried out in other countries of the European Union? Not so very much, according to an international comparison carried out by the University

Louvain-la-Neuve (1994). On average, most educational research in the European Union is carried out in secondary education, closely followed by primary education, while in higher and vocational education much less research is undertaken. In the European Union, less attention is paid to achievement tests, curriculum, school organisation and school effectiveness than in The Netherlands. The attention paid to learning and instruction and educational policy in The Netherlands is about equal to the average attention in the European Union. In The Netherlands, educational research is carried out relatively frequently by sociologists and psychologists, in comparison with the European Union as a whole. In all countries, economists and lawyers are infrequently involved in educational research.

Priorities in educational research are based on a report of the Foresight Commission of Educational Research, as they are in other fields of research. This report has served as a catalyst for discussion between the representatives from the world of science and the interested parties in the society at large.

The Foresight Commission distinguished five major trends in the society of the future which are related to the development of education: the globalisation of the economy, the technological development of communication and information technology, the higher degree of internationalisation, the growing number of people from ethnic minorities and the increasing individualisation.

The Commission also made an analysis of recent educational reforms and the expected educational policy for the future. The main educational reforms are the integration of children with special needs in regular primary and secondary education, the efforts to provide equal chances for ethnic and social minorities, the introduction of basic secondary education and the four curriculum profiles for the upper stage of secondary education, and the qualification structure in vocational education.

On the basis of these trends and developments, the Committee recommended two priorities in educational research: learning and instruction and the management and organisation of schools and learning. The government accepted and combined these priorities, but added communication and information technology and recurrent education.

The educational research policy-making process involves extensive and elaborate consultative procedures. Although these might damage the efficiency of policy-making, they also create an opportunity for research to influence decision-making. Moreover, policy makers display a growing interest in evaluative research as a means to establish to what extent certain

policies achieve particular goals. This policy framework forms the basis of research policies and contributes to the phenomenon of professional compartmentalisation in the support structure. Research does play a role in educational policy-making and policy evaluation in The Netherlands, but this role is generally not decisive. This may be attributed to the well-known problems of transfer between research and policy and to complexities in policy-making processes and the duration of these processes

References

BIEF (1994). *Inventaire des Recherches en Éducation dans les États de la Communauté Européenne (Inventory of Educational Research in the European Community)*. Louvain-la-Neuve: Université Louvain-la-Neuve.

KLOPROGGE, J., VAN OIJEN, P., ROEMERSMA, F., VAN TILBORG, L., WALRAVEN, G. and WIND, D. (1995). *Educational Research and Development in The Netherlands*. The Hague: Institute for Educational Research in the Netherlands.

MINISTRY OF EDUCATION, CULTURE AND SCIENCE (1996). *The Development of Education: Educational Policy in The Netherlands 1994–1996*. Zoetermeer: Ministry of Education, Culture and Science.

MINISTRY OF EDUCATION, CULTURE AND SCIENCE (1997). *Science Budget 1997*. Zoetermeer: Ministry of Education, Culture and Science.

SCHEERENS, J. and BRANDSMA, J. (1994). *Inventory of Educational Research in The Netherlands*. Enschede: Research Centre for Applied Research on Education.

VAN GENNIP, J. and PELKMANS, A. (1994). *Zwaartepunten in Het Nederlandse Onderwijsonderzoek (Main Points of Educational Research in The Netherlands)*. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.

VERKENNINGSCOMMISSIE ONDERWIJSONDERZOEK (1996). *Nederlands Onderwijsonderzoek Verkend 1996 (Foresight Study of Educational Research 1996)*. Amsterdam: Overleg Commissie Verkenningen.

3

Nature and function of educational research

Barry McGaw, Australian Council for Educational Research

The term 'research' is so widely used these days that it has almost lost its currency. It is used by primary school children to describe information-gathering in a library, by adults in virtually every walk of life to describe any activity in which they are finding out something they did not know already, regardless of how many people might already know it, and by professional 'researchers' as a specialist term to cover a great variety of activities.

For any serious discussion, however, we must claim specialist use of the term 'research' but then make clear what will fit under the term and what subclassifications we might wish to use. The title of this session invites such a discussion.

Characterising educational research

Defining research

Broad definition

The Organisation for Economic Cooperation and Development (OECD) conducts regular surveys of research activities, focused more broadly on research and experimental development, which are defined in the Frascati Manual as:

> creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society and the use of this stock of knowledge to devise new applications (OECD, 1992, p. 29).

The key distinguishing characteristics of research and development are *originality* and *investigation* as a primary objective. Under this definition, the Frascati Manual excludes:

- education and training, except for postgraduate students' research and its supervision;

- scientific and technical information services;

- general purpose data collections;

- policy-related studies such as the analysis and assessment of existing

programmes and the work of legislative commissions of inquiry; and

- testing and standardisation.

Where would national or international programmes to monitor student achievement fit, for example? Are they to be seen as 'general purpose data collections' in whole or in part? Should national monitoring programmes be considered 'research and development' in the initial years, as novel systems are developed and implemented, and then be removed from the category once the programme becomes established? Should the 'horse-race' part of international surveys be excluded and only the exploration of relationships between achievement levels and other variables be included?

Some investigative activities common in education would be excluded. Applying well-established techniques to problems, such as in the diagnostic testing of children with learning disabilities, would be excluded, and indeed this example is explicitly identified in the Frascati Manual as a case for exclusion (p. 34). If the data from such tests were collected as part of a study of the links between other variables and learning disability, then the testing component would constitute research and development.

The OECD definition and its associated exclusions may be of little immediate relevance in a conference on educational research but non-trivial in other contexts. Both national surveys of research and development activities, and the OECD surveys which draw upon them, document the extent of national investment in educational research and development. The OECD definition and its exclusions determine, or influence, what is counted. In addition, the OECD education indicators reports in the *Education at a Glance* series will also now include indicators of national investment in educational research and development established on a similar basis.

Subclassifications

The Frascati Manual defines three main subcategories of research and development activities:

- *basic research:* experimental or theoretical work undertaken primarily to acquire new knowledge, without any particular application or use in view;

- *applied research:* original investigation undertaken to acquire new knowledge but directed primarily towards a specific practical aim or objective; and

- *experimental development:* systematic work, drawing on existing knowledge, directed to produce new products, processes, or services, or to improve existing ones substantially.

An education-specific example is provided by the Frascati Manual (Annex V), distinguishing these three types of research and development in the field of reading:

- *basic research:* study of the reading process, for example an investigation of how human visual systems work to acquire information from symbols such as words, pictures and diagrams;

- *applied research:* study of the reading process for the purpose of developing a new teaching method; and

- *experimental development:* development and testing of a special reading programme.

For the Australian national research and development surveys, the Australian Bureau of Statistics subdivides *basic* research into *pure basic* and *strategic basic* research according to the following definitions:

- *pure basic research:* research directed into specific broad areas in the expectation of useful discoveries; and

- *strategic basic research:* research providing the broad base of knowledge necessary for the solution of recognised practical problems.

Education as a field or discipline

The OECD classifications are generic in the sense of being intended to apply to all fields of research and development. When it comes to education, there is not only the question of how these definitions might be applied but also the question of what research and development should count within the category of *education*.

Education is more appropriately seen as a field than as a discipline. This arises from the diversity of what, on any reasonable grounds, could be claimed to be educational research and development. Diversity is evident in the substantive questions addressed, ranging from learning and teaching to the politics of education; in the scope of the questions, ranging from learners and teachers to systems; and in the methodologies used, quantitative and qualitative. It is also clear that educational research and development is informed by a variety of disciplinary perspectives. What makes a particular research and development activity *educational* is its focus on the field of education.

Taking education to be a field of inquiry and not a discipline is not without problems, however, since the boundary of the field will inevitably be somewhat vague, wherever it is drawn. This can be illustrated by responses to the Australian Bureau of Statistics' surveys of research and development activities.

In those surveys, the Bureau uses not only type of research (pure basic, strategic basic, etc.) but also *Field of Science* and *Socioeconomic Objective*, the latter referring to the area of expected national benefit from the research. In the survey for 1988–89, expenditure on educational research and development (in March 1992 prices) was reported to be $70.3 million for education as a socioeconomic objective but $63.6 million for education as a field of science. The latter figure was lower because it excluded expenditure on educational research and development incurred outside of the field of education by researchers declaring their allegiance to some other field in the social sciences or humanities, usually a discipline, while acknowledging their objective to be to address educational questions.

This kind of distinction is not possible in the OECD reports since OECD includes education in its *Fields of Science and Technology* but not in its *Socioeconomic Objectives*. That is, the OECD classification includes education only in the list of 'disciplines'. Australia's inclusion of education in both subclassifications in its national surveys allows the possibility of work having disciplinary orientation and a commitment to the field of education.

The term *educational research* is best preserved for work in which the central organising feature is a dominant commitment to the field of education. If the research focus is the concerns of a discipline, the work ought to be seen as related to that discipline and so named. The research may be relevant to education but still not be educational research.

There are numerous examples of research relevant to education, on which educational researchers and even policy makers and professional practitioners may draw, without these examples needing to be named as educational research. Within the discipline of psychology, research on learning, on the nature and development of expertise, on human development, and on attitude formation can all fall within the category of *educationally relevant* research without being called *educational* research. Within the discipline of sociology, research on social class and socioeconomic status can be relevant to education and can crucially inform educational research on the social reproductive role of schools. Within the discipline of politics, research on policy making in federal systems, even work on educational policy making, can be relevant to education without being educational research under a definition which

demands that the prime focus be education and not the discipline.

There are many professional researchers in educational research institutions, particularly university faculties of education, who are inclined to honour a disciplinary perspective from which they work as their *raison d'être* as researchers rather than the field of education, even though they may claim their work to be relevant to education. For them, the appropriate course could well be to transfer to the relevant disciplinary department.

Connecting research with policy and practice

Demanding of educational research that it have a central commitment to the field of education does not resolve, in any immediate or straightforward fashion, the manner in which the research endeavour contributes to the field. To do that requires consideration of the manner in which research can link with policy development and professional practice. It also opens a significant epistemological debate.

Differences in perspective

Academic and professional concerns intermingle in educational research. Academic concerns typically range across the purposes, procedures and reporting of educational research. Professional concerns tend to be practical. They draw upon educational research as a body of ideas from which to generate new solutions, not new knowledge. In some cases, policy makers and practitioners seek enlightenment to inform their action; in other cases they seek recipe knowledge to prescribe their action.

Educational research is thus expected to generate different kinds of knowledge, reflecting the different questions with which different groups approach it. While they may all have a connection with the fields of policy and professional practice, there is not the same immediate concern in all cases with implementation, or even identification, of desired changes. The reason is that the sources of their questions vary. Anderson (1986) suggests that the sources of research questions include:

- intellectual curiosity;
- hypotheses arising from theory;
- personal concern about particular problems; and
- requests from policy agencies.

Another way of thinking about the differences between researchers and those professionally engaged in the practice of education, either as policy makers or as teachers, is in terms of their professional perspectives. Key differences are summarised in Table 1.

Table 1: Differences in perspective of researchers and professional practitioners

Researchers	Policy makers and practitioners
willing to work on parts of problem	concerned with complex, applied problems — holistic, context-based
interested in raising new questions	interested in facilitating policy and practice choices
preference for critical analysis and critique	sometimes a preference for legitimating choices already made
professional rewards (promotions criteria) emphasise basic research	professional rewards depend on development and implementation of changes in policy and practice

Differences in research tradition

The epistemological debate in educational research is more fundamental than the differences in perspective of the community of researchers and the communities of policy makers and practitioners. It is to do with the manner in which knowledge is gained and that, in turn, influences how educational research is conceived. Four long-term epistemological tensions are evident in educational research, with it conceived as:

- either a natural or an interpretative science;
- either a laboratory or a practical science;
- either a universal or a local science;
- either serving deliberative or practical ends (McGaw *et al.*, 1997, p. 4).

The implications of these tensions can be illustrated with descriptions of three broad research traditions in education, viz. natural science investigation, interpretative studies, and interventions and evaluations.

Natural science investigation

The natural science tradition has not disappeared entirely. It continues on the assumption of a natural, lawful reality to be discovered, pursuing it through controlled experimentation and the quasi-experimentation afforded by semi-controlled studies capitalising on natural covariation and using statistical adjustments to control for pre-existing differences by casting questions in an 'other things equal, what if?' form.

In the natural science tradition, the experimenter maintains an objective distance from subjects. Interpretations are expected to be verifiable by an independent data analyst and the data gathering itself replicable within sampling error in independent circumstances.

Whether the conclusions are anticipated to be universal or local depends on the extent to which the effects being investigated are understood, or expected, to be free of cultural influences. In his address to the American Psychological Association on the receipt of a Distinguished Scientific Contribution Award, Cronbach (1975) introduced the possibility of local laws as the product of psychological research in the natural science tradition and proposed systematic investigation of interactions of other variables with local conditions. The most notable work that flowed from his proposal was the investigations of aptitude–treatment interactions that he subsequently reviewed (Cronbach and Snow, 1977). A more elaborate method for investigating interactions of context with research findings was provided by meta-analysis (Glass, 1976; Glass, McGaw and Smith, 1981).

Connections with educational practice from this research tradition first proceeded in the belief that experimental, or quasi-experimental, conditions shown to be efficacious could be engineered into programmes or materials to be used by professional practitioners. Greater understanding of the complexities of the domain of professional practice led to more modest expectations of influence, but not an abandonment of the principle of the efficacious experimental treatments serving as a guide to professional practice. There remains an expectation that good research will help to shape planned interventions.

In some cases, implementation itself becomes the subject of experimental investigation. Good research on learning usually leads on to instructional research, which is then seen as providing essential validation of the conclusions of the learning research. On a larger scale and usually in a quasi-experimental mode, evaluation studies investigate the impact of programmes in complex settings.

Interpretative studies

Research in the interpretative tradition reflects new ways of thinking about knowledge and competence in human practice. It sees knowledge as actively constructed by individuals, but not in isolation from the communities in which they develop. That is seen to be true of learners who might be the focus of professional practice and research, but is seen also to be true of researchers and professional practitioners.

Individuals are seen to participate in communities of discourse which shape the behaviour of all participants. The research community has a scholarly discourse within which theories are developed and accepted modes of challenging theories established. The communities of professional practitioners and policy makers also each develop a common language which shapes individuals' interpretations of their own practice and of the possibilities of their work.

Research in the interpretative mode usually seeks to understand not just what people do, but how they derive meaning from what they are doing. Much interpretative research focuses on the nature of discourse in various communities of practice. In education, various studies are now charting the language of classrooms, with particular attention to the ways in which students are initiated into the questions, points of view, expectations of evidence, and forms of verification of different disciplines. The policy development process is being subjected to similar analyses.

Interventions and evaluations

Another body of educational research takes social change as its object of inquiry, and seeks both to evaluate the overall effectiveness of interventions and to determine which aspects of an intervention, singly or in interaction, are responsible for positive or negative effects. In educational research, the interventions are likely to be programmes of teaching or curriculum change, but they may also be broader policies about, for instance, patterns of funding or selection of students.

Intervention and evaluation research calls for careful observation and measurement within either a natural science or interpretative framework. Because the objects of study are complex social interventions, attention must be paid to specifics of context, including historical and cultural factors that may influence the implementation of planned interventions.

The focus may be deliberately designed interventions, sometimes designed

by the researchers themselves. In such cases, control or comparison groups are typically used to estimate the effects of the intervention. Various statistical adjustments may also be used to control for possible confounding or interacting variables.

In other cases, natural variations in educational programmes can provide comparisons of different interventions. Comparisons are possible.between countries, states, schools or even individual teachers. Inferences from these non-experimental studies of natural variation depend on the use of increasingly sophisticated statistical procedures.

Interpreting differences in research traditions

Debates among researchers about the different research traditions are hotly contested. Much of the debate has been advanced as though there were a single answer, with the result being attempts to establish epistemological, and methodological, hegemonies.

Some, such as Lincoln and Guba (1984), point to an historical development in which 'different sets of basic beliefs have guided inquiry in quite different ways' in successive periods from 'prepositivist' to 'positivist' and then to 'postpositivist'. Others, such as Husén (1988), focus on the positivist (natural science investigations) and postpositivist (interpretative). Popkewitz (1984) describes these two as 'empirical–analytic' and 'symbolic' but adds 'critical' as a third. Critical theorists emphasise the 'search for knowledge of the "genuinely human" ... to improve the human condition' (Walker and Evers, 1994, p. 5024) and so build a direct connection between research and practice.

Walker and Evers (1994) suggest that there are three ways in which the competing research traditions can be dealt with. Two accept the diversity, one seeks unity.

One approach to the diversity is to see the traditions as epistemologically different and incommensurable and so mutually incompatible. This approach, which Walker and Evers (1994, p. 5022) call 'oppositional diversity' results in claim and counter-claim about which approach is the valid one. Proponents argue for the validity of their epistemology and then for the utility of their preferred methodology in yielding understanding of educational phenomena and establishing the base for links with policy and practice.

An alternative approach to diversity is still to see the traditions as epistemologically different and incommensurable but complementary. This 'complementary diversity' thesis, according to Walker and Evers, is attractive

67

to many 'in view of various distinct desiderata in educational research, not all of which can be met by any one single paradigm' (p. 5024). From this perspective, no one epistemology need prevail. Each can yield its insights and contribute to a richer whole.

Walker and Evers (1994) believe neither approach accepting diversity is adequate. Instead they argue that 'there is a touchstone for judging the respective merits of different research traditions and bringing them into a productive relationship with one another' (p. 5022). From the perspective of their 'unity thesis', however, it does not follow that all theories are equally reasonable. 'Having identified common ground between theories, their differences are next rigorously set out ... to identify the strengths and weaknesses of each, and to reach a decision on the theory which is strongest under present circumstances' (p. 5029). The criteria used in this coherentist approach are more than empirical adequacy. They include as well 'consistency, simplicity, coherence, comprehensiveness, and fecundity' (Evers and Lakomski, 1994, p. 1773).

Reviews of educational research in Australia

Three reviews of educational research have been conducted in Australia in recent years. One provided a review of the then state of educational research and proposed a variety of policy changes in relation to the research enterprise (McGaw *et al.*, 1992) and a second provided a similar review focused more narrowly on research in vocational education and training (McDonald *et al.*, 1993). The third reviewed the outcomes of educational research projects funded under the Australian Research Council's Large Grants Programme in the period 1989 to 1993 (McGaw *et al.*, 1997).

It is instructive to reflect on the findings of these reviews in terms of the analysis of the nature of educational research and development presented in the preceding sections of this paper. Given the strong links many Australian educational researchers have with counterparts in other countries, it is likely that the Australian experience will be similar to that in other places.

Strengths

By some measures, educational research in Australia was shown in the strategic review to be in good shape. From the major national funding agency for basic research, the Australian Research Council, education attracted 15 per cent of large grant funds allocated to the social sciences in 1992. This

proportion was exceeded only by psychology (40 per cent) and economics (23 per cent), and was well in advance of sociology, (seven per cent) and geography (six per cent). In funding for new projects, education was second after psychology (McGaw *et al.*, 1992).

Funding for educational research, however, was relatively low considering the size of the education sector within higher education and appeared to have been associated with a relatively low rate of applications from education academics — about one per 50 full-time equivalent education academics compared with about one per 25 full-time equivalent non-education academics in the social sciences. The relatively low application rate reflected the history of much of the education sector in higher education — originally in teachers' colleges and then in colleges of advanced education without a research brief or infrastructure funding for research and only in the early 1990s transferred fully into the university sector. This has had some impact on the profession's view of research as well, with the great majority of teachers trained in institutions and by staff without any substantial engagement with research. That is very different from the education environment in fields such as medicine and psychology.

A great deal of the educational research is of high quality. This judgement was confirmed by a systematic analysis of the outcomes of the educational research projects that gained the highest levels of Australian Research Council funding over the five-year period 1989–93. Work from these projects was widely published in international journals and the researchers had formed strong links with leading overseas researchers in the specialties.

Weaknesses

Despite these strengths, there was other evidence of serious problems. The whole educational research enterprise seemed at the time of the strategic review in 1992 to be a fragile resource requiring some reorientation and increased support. Areas of concern evident in submissions to the review included:

- education administrators and practitioners perceiving much educational research to be irrelevant to their concerns;

- educational researchers perceiving a lack of support for, and interest in, their work from government agencies and the higher educational sector;

- there being little effective dissemination of research;

- research activities being poorly coordinated and thinly spread;

- funding being low relative to the size of the education sector (0.35 per cent of education expenditure being allocated to research and development compared with 1.40 per cent — and soon after 2.0 per cent — of health expenditure being allocated to research and development); and

- there being little planning for training researchers.

Researchers seemed generally isolated from one another, working as individuals on their own projects, and certainly isolated from the communities of policy makers and practitioners. That was not true for all, however, as the subsequent review of the outcomes of the large grants revealed. Much of the work on those large grant projects involved coherent teams of collaborators, sometimes in different locations, but working together in a sustained way. A good number of them had also forged effective links with policy makers and practitioners.

Examples of educational research and development

The discussion so far has offered a general description of the nature of educational research and development and a brief reflection on the conclusions of several substantial reviews of educational research in Australia. Three examples of current Australian educational research, which have a strong standing in the research community and well-developed connections with the communities of policy makers and practitioners, are now described.

There are other examples that could have been chosen, of course, but these will illustrate the diversity of approaches that is currently adopted and, to some extent, also illustrate the benefits and limitations of each.

Cognition to learning to instruction

The first example is of research on learning, derived from a strong theoretical model of cognitive processes and leading on to instructional research and successful implementation in workplace training.

Sweller has developed a theoretical model of problem solving and learning which accounts, in terms of cognitive load, for the variable levels of success of different instructional strategies. The model grew out of extensive studies of learning (e.g. Sweller, 1983; Sweller and Levine, 1982).

Subsequent work established that some standard techniques for teaching problem solving in mathematics, such as working numerous examples, can be very ineffective because the cognitive demand of the problem-solving tasks themselves is often so great that insufficient cognitive resources can be directed to learning the structure of the problems and thus acquiring the intended general problem-solving skills. This led on to the identification of more effective teaching strategies that reduce cognitive load and direct cognitive effort to learning rather than to extraneous activities, and then on to the definition of the characteristics that those strategies must have. For example, studying worked examples in algebra can produce faster and more effective learning of problem types and problem-solving strategies than solving sets of problems (Sweller and Cooper, 1985). This is not true for all worked examples and the research provides both a theoretical explanation of why it is not and a delineation of the characteristics of worked examples that will be effective (Ward and Sweller, 1990).

Sweller extended the application of his model of learning by working in industry, developing instructional programmes for staff to improve the rate at which operational skills are acquired. New techniques for formatting instruction, which better manage the cognitive load of the learners, have proved highly effective (Sweller *et al.*, 1990; Chandler and Sweller, 1991).

It is interesting to note that, while Sweller's application of his work to mathematics education was not warmly embraced by some mathematics educators, as illustrated by the reception for a presentation he made to the Mathematics Education Research Group of Australia, his work on industrial training has been speedily taken up. The company in which he and Chandler worked contributed substantially to the funding of the research and development and then collaborated in a successful submission for further research funding from the Australian National Training Authority Research Advisory Committee. That further work has demonstrated that the use of redundant information in multi-media presentations (e.g. presenting the same information as text to be read and in audio presentation) inhibits learning by increasing the cognitive demand of the learning task without benefit.

Sweller's work represents a sustained effort within the tradition of natural science investigations. It has clear and strong connections with aspects of professional practice since it has provided systematic evidence of the efficacy of instructional strategies derived from the underlying theory and the results of prior research.

Longitudinal study of individual progress

The second example is of a longitudinal study, following successive national samples of individuals from upper primary or lower secondary schooling to adult life, with annual documentation of education and labour market participation. It is work in the tradition of quantitative sociology, fitting linear models to data to explore the links between a variety of educational and social background factors and subsequent educational performance and workforce participation.

Using samples of individuals born in 1961, 1965 and 1970, Williams and his colleagues documented changes in patterns of participation in the final years of secondary schooling and in higher education of groups making the transition to higher education at the beginning, in the middle and at the end of the 1980s (Williams, 1987; Williams *et al.*, 1993).

This work showed the broadening social base of the student cohort continuing in education as participation rates rose, but documented the continuing strong influence of socio-economic factors on completion of secondary education and participation in higher education. In these respects, it provided direct evidence on the effectiveness of policies of the Commonwealth government intended to increase equity in participation in post-compulsory education, and the Commonwealth drew heavily on the work in its own private and public evaluations of its policies (Commonwealth Department of Education, 1987).

This research also demonstrated that private schooling has an influence on school completion and access to higher education which is not explicable in terms of the relatively select nature of the students enrolled in private schools, that is, not explicable in terms of social characteristics or prior educational achievements of the individual students but rather attributable to the schools themselves. Those conclusions generated an ongoing debate about the influence of school type on subsequent success.

More recent work with this longitudinal database, using additional national samples of individuals born in the years up to 1980, has provided new evidence about increased equity in the Australian education systems. This work shows that family socio-economic status has had a declining influence over a 20-year period on students' performance levels in reading comprehension and numeracy (Marks and Ainley, 1996).

These *Longitudinal Surveys of Australian Youth* have not had any direct connection with professional practice in schools. Rather, the research has

had a clear connection with policy makers, and a direction which is not one-way. Policy makers have defined the research questions with which to come to the data and, increasingly, are playing a role in shaping the annual data collection in anticipation of their own subsequent policy questions. The new federal Minister is regularly commissioning analyses of the database to investigate policy options that he is exploring.

The database provides information on links between a variety of background measures, a small number of proxies for measures of aspects of the schooling process, and a variety of outcome measures. It provides little information on individuals' own interpretations of their experiences or rationales for their decisions.

Linking researchers with practitioners

The third example is of action research in which professional practitioners and academic researchers are engaged as collaborators. It is a project sponsored by government and teacher's unions and funded through a national professional development programme. It is based on a view that professional development of teachers will be enhanced if they are enabled to reflect critically on their own professional practice.

Action research aims to help practitioners investigate the connections between their own theories of education and their own day-to-day educational practices; and to integrate the research act into the educational setting so that it can play a direct and immediate role in the improvement of practice (Carr and Kemmis, 1986).

In action research, practitioners are encouraged to treat their own educational ideas and theories, their own work and work practices, and their own work settings as objects for analysis and critique. On the basis of careful reflection, they may uncover theoretical ideas which turn out to be unjustified, ways in which practices shaped by habit or tradition have become irrelevant, and how the settings in which they work place obstacles in the way of attaining educational goals.

The particular project funded under the national professional development programme operated under the descriptive title, *Innovative Links between Universities and Schools for Teacher Professional Development*. It has provided the opportunity for academic researchers in 14 universities to work in partnership with practising teachers on a whole-school basis in around 100 schools (Yeatman and Sachs, 1995).

73

Sachs describes ways in which the research interests and perspectives of the university and school-based personnel are complementary and mutually supportive (Sachs, 1996). She points out that teachers are concerned primarily with understanding and improving practice and so have relatively context-specific interests. What they gain from the researchers is cross-contextual research information which enables them to situate their reflective practice in a wider context.

Academic analysts not only become directly acquainted with what it means to engage in continuous improvement in teaching and learning in a specific context, but become aware of what teachers regard as important and relevant and why. This helps to ensure that, when academic communication is oriented toward practitioners, it is informed about the needs and requirements of practitioners and practice settings.

Academic research, on the other hand, is concerned with validity and generalisation and so is essentially cross-contextual, whether at a micro or macro level. While this research is not focused on *doing* and *improving* it is concerned with understanding the processes of action and improvement. Exposure to the day-to-day realities of school life through collaboration with professional practitioners seeking to reflect critically on their own practice in order to improve it enriched the academic researchers' own understanding of practice and tested the relevance of their cross-contextual information and theories.

Participants reported, through a formative evaluation, that the relationship between teachers and academic associates was mutually beneficial when it was an equal relationship and best when the participants viewed each other as learners in the project (Yeatman and Sachs, 1995). With this approach, the issue of dissemination is entirely redefined. There is no remote audience for research with whom effective communication must be established. There is collaboration on what the research questions are and on how to address them. There is divergence on the scope and level of generality of conclusion sought but neither is seeking an interpretation for the other to use.

Strengthening educational research and development

The analysis of the nature of educational research and development, the reviews of the educational research in Australia and the three varied examples all point to ways in which the research enterprise might be strengthened. This will need to involve changes to the research enterprise itself and changes

to the linkages between the research community on the one hand and the communities of policy makers and professional practitioners on the other.

Changes in research effort

Maintaining eclecticism

One important message is the need to maintain an eclectic approach. Epistemological or methodological hegemony will limit and stultify the research and development enterprise. The balance may well be shifting from 'natural science' to 'interpretative' models of research but there continues to be significant research of both a basic and applied kind in the natural science tradition which informs our understanding of educational processes and which can introduce and validate new practices. Sweller's work is an example.

The review of the outcomes of the work of the 20 educational research teams funded between 1989 and 1993 under the Australian Research Council's Large Grant programme found research fitting all three of the models outlined earlier, viz. 'natural science', 'interpretative' and 'interventions and evaluations', as shown in Table 2 (McGaw *et al.*, 1997).

Table 2: Distribution of funding by research model for research team

Team	Natural science	Interpretative	Interv & evaluation	Total
1	215,000			215,000
2			210,000	210,000
3	82,500		117,500	200,000
4	95,000		95,000	190,000
5		85,000	85,000	170,000
6	160,000			160,000
7	150,000			150,000
8	75,000	75,000		150,000
9	145,000			145,000
10	140,000			140,000
11		140,000		140,000
12	67,500	67,500		135,000
13	125,000			125,000
14		110,000		110,000
15	105,000			105,000
16		100,000		100,000
17	100,000			100,000
18	90,000			90,000
19		80,000		80,000
20	80,000			80,000

The classification in Table 2 shows that the dominant model in the work of the research teams was the natural science model, with almost 60 per cent of all funding going to projects using this model. Research in the interpretative tradition attracted almost 25 per cent and intervention studies less than 20 per cent.

This may, however, be an underestimate of the overall influence of the interpretative model in educational research. Several of the research teams reviewed shifted over the period from 1989 to 1993 toward more interpretative frameworks. Given general trends in the social sciences, it is likely that increasing numbers of the best education scholars may make similar moves, perhaps bringing to interpretative research more use of quantitative methods than has been the case in the past.

Concentrating effort

A second, somewhat controversial, strategy would be to concentrate the research effort to some extent. This would permit more sustained work by those most likely to deliver but would, of course, be at the expense of others who might have hoped for support for at least some aspects of their work.

In circumstances of limited resources for research, concentrating funding on fewer researchers rather than spreading it thinly would involve picking 'winners' in whose work to invest. It would be important then to have a rigorous evaluation of the outcomes of the funding so that resources can be redirected over time if performance does not justify continued support.

Some research funding programmes appear to offer renewal of funding on the basis of the quality of each round of applications for further funding, with little evidence of the results of previous funding other than self-report in the next-round application.

Building multidisciplinary teams

The requirements of eclecticism and concentration of effort can be brought together in the development and funding of multidisciplinary teams to work on educational issues. Enriched understandings can come from the variety of substantive and methodological perspectives a multidisciplinary team can offer. Application of the criteria of a 'coherence theory of justification' (Evers and Lakomski, 1994, p. 1773) will ensure that simple criteria such as empirical adequacy do not dominate in the formulations of the research findings.

76

There was some multidisciplinarity in the 20 teams whose work was reviewed in Australia. A team reviewing the development of youth policy over several decades consisted of a historian, a political scientist with research interests in social class and international social movements and a researcher with a background in youth work and youth policy. A team concerned with mathematics learning consisted of a university mathematician and teacher and a cognitive psychologist. Other teams were more homogeneous, however, concentrating expertise and bringing the benefit of broader experience and a capacity to share in some ways more deeply than a multidisciplinary team.

Improving links with policy making and practice

If the unique feature of educational research is its engagement with a field of policy and professional practice, then work needs to be done to build and sustain good links between the research community and the communities of policy making and professional practice.

Setting key agenda issues collaboratively

One way to work at this is to have the research agenda set collaboratively. This can be done for the work of an individual team, as it is currently being done for the *Longitudinal Surveys of Australian Youth*, or it could be done in nominating some priority areas within a national research enterprise or even in setting a national research agenda.

A good example of the limited imposition of national priorities in the context of a research funding programme which could accommodate other non-thematic work is provided by the Australian National Training Authority's national, competitive research funding programme set up following the recommendations of the review of research on vocational education and training (McDonald *et al.*, 1993). The priorities have been established by a group representative of all three communities. Substantial funding has been provided to establish a few centres focused on particular priority areas, some large projects have been funded in other priority areas and a variety of large and small projects have been funded in the priority areas as well as outside them.

Establishing personnel exchanges

Apart from active collaboration from different home bases, as in the Innovative Links Project bringing together school and university personnel to work on shared projects while maintaining their institutional affiliations, personnel

exchange programmes can build links in new ways between the separate communities.

Building incentives for engagement with policy and practice

Perhaps the most important thing to do in pursuit of collaboration is to build incentives which encourage it. The Australian National Training Authority Research Advisory Committee gives priority in funding to project teams in which there will be genuine collaboration among team members from the different communities. That Committee also ensured that its funding programme was accepted as one of the competitive programmes from which winning of grants brings additional research infrastructure funding to universities from the university funding agencies.

More generally, work needs to be done to reduce the impact of the different interest and reward systems identified in Table 1. Unless they are reoriented in some way, the interests of the different communities will not intersect and, worse, may even be in conflict.

References

ANDERSON, D. (1986). 'Where do the questions come from? For whom are the answers intended? An exploration of research and policy', *Australian Educational Researcher*, **13**, 5-25.

CARR, W. and KEMMIS, S. (1986). *Becoming Critical: Education, Knowledge and Action Research*. London: Falmer.

CHANDLER, P. and SWELLER, J. (1991). 'Cognitive load theory and the format of instruction', *Cognition and Instruction*, **8**, 293-332.

COMMONWEALTH DEPARTMENT OF EDUCATION (Ed) (1987). *Support for Staying on at School: the Role of Student Assistance in Postcompulsory Secondary Education and Technical and Further Education*. Canberra: Australian Government Publishing Service.

CRONBACH, L.J. (1975). 'Beyond the two disciplines of scientific psychology', *American Psychologist*, **30**, 116-27.

CRONBACH, L.J. and SNOW, R.E. (1977). *Aptitudes and Instructional Methods: a Handbook for Research on Interactions.* New York: Irvington.

EVERS, C.W. and LAKOMSKI, G. (1994). 'Educational administration: ethical and philosophical issues.' In: HUSÉN, T., and POSTLETHWAITE, T.N. (Eds) *International Encyclopedia of Education.* Second edn. Oxford: Pergamon, pp. 1769-1775.

GLASS, G.V. (1976). 'Primary, secondary and meta-analysis of research', *Educational Researcher*, **5**, 3-8.

GLASS, G. V., McGAW, B., and SMITH, M. L. (1981). *Meta-analysis of Social Research.* Beverly Hills, California and London: Sage.

HUSÉN, T. (1988). 'Research paradigms in education'. In KEEVES, J.P. (Ed) *Educational Research, Methodology and Measurement: an International Handbook.* Oxford: Pergamon.

LINCOLN, Y.S. and GUBA, E.G. (1984). *Naturalistic Inquiry.* Beverly Hills, California: Sage.

McDONALD, R., HAYTON, G., GONCZI, A. and HAGAR, P. (1993). *No Small Change: Proposals for a Research and Development Strategy for Vocational Education and Training in Australia.* Canberra: Vocational Education, Employment and Training Advisory Committee.

McGAW, B., BOUD, D., POOLE, M., WARRY, R. and MCKENZIE, P. (1992). *Educational Research in Australia.* Canberra: National Board of Employment, Education and Training.

McGAW, B., HAMILTON, D., KEEVES, J. and RESNICK, L. (1997). *Education: Evaluation of the Australian Research Council Large Grant Funding Program 1989–1993.* Canberra: Department of Employment, Education, Training and Youth Affairs.

MARKS, G. and AINLEY, J. (1996). *Reading Comprehension and Numeracy among Junior Secondary School Students in Australia.* Melbourne, Vic.: Australian Council for Educational Research.

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD). (1992). *Proposed Standard Practice for Surveys of Research and Experimental Development (Frascati Manual).* Paris: OECD.

POPKEWITZ, T. (1984). *Paradigm and Ideology in Educational Research.* London: Falmer.

SACHS, J. (1996). *Renewing Teacher Professionalism: What Role for the University?* Alexander Mackie Lecture, University of Sydney.

SWELLER, J. (1983). 'Control mechanisms in problem solving', *Memory and Cognition*, **11**, 32-40.

'Cognitive load as a factor in the structuring of technical material', *Journal of Experimental Psychology: General*, **119**, 176-92.

SWELLER, J. and COOPER, G. (1985). 'The use of worked examples as a substitute for problem solving in learning algebra', *Cognition and Instruction*, **2**, 59-80.

SWELLER, J. and LEVINE, M. (1982). 'Effects of goal specificity on means–end analysis and learning', *Journal of Experimental Psychology: Learning, Memory and Cognition*, **8**, 463-74.

WALKER, J.C. and EVERS, C.W. (1994). 'Research in education: Epistemological issues.' In: HUSÉN, T. and POSTLETHWAITE, T.N. (Eds) *International Encyclopedia of Education*. Second edn. Oxford: Pergamon, pp. 5021-31.

WARD, M. and SWELLER, J. (1990). 'Structuring effective worked examples', *Cognition and Instruction*, **7**, 1-39.

WILLIAMS, T. (1987). *Participation in Education*. Hawthorn, Vic.: Australian Council for Educational Research.

WILLIAMS, T., LONG, M., CARPENTER, P. and HAYDEN, M. (1993). *Higher Education in the 80s*. Hawthorn, Vic.: Australian Council for Educational Research.

YEATMAN, A. and SACHS, J. (1995). *Making the Links: a Formative Evaluation of the First Year of the Innovative Links between Universities and Schools for Teacher Professional Development*. Perth, Western Australia: Murdoch University.

Respondents' comments

Sally Brown, University of Stirling

Barry McGaw has offered us a thought-provoking paper which looks at the nature and function of educational research. He has provided an interesting and wide-ranging map of the domain and since there is no way that I could address all the things he has to say, there are just three broad issues on which I will attempt to add a little more to the debate. The first of these additions suggests that although McGaw's paper makes it clear that research is an ambitious and demanding enterprise, he does not alert us to the current and fundamental challenges which his rational 'research traditions' are experiencing, or to the failure of actual studies to conform to these classical traditions, or to the debilitating discomfort researchers experience as they become aware that they cannot fulfil the traditions' criteria.

For example, there is a rather stark omission from this paper if it purports to be a *contemporary* statement about the nature of, and approaches to, educational research. No reference is made to the influence of the postmodernist thinking on the debate, epistemological or methodological, about different research traditions. It seems odd that major initiatives like that of the OECD, on which much of the early part of the text is based, seem unaware of the blast of deconstruction which postmodernist questioning has landed on the kinds of truth claims pursued by the research traditions in education cited in the paper. As Eagleton (1996) has said:

> Postmodernity is a style of thought which is suspicious of classical notions of truth, reason, identity and objectivity, of the idea of universal progress or emancipation, of single frameworks, grand narratives or ultimate grounds of explanation. Against these Enlightenment norms, it sees the world as contingent, ungrounded, diverse, unstable, indeterminate, a set of disunified cultures or interpretations which breed a degree of scepticism about the objectivity to truth, history and norms, the givenness of nature and the coherence of identities (pvii).

One may be unwilling, of course, to accept what postmodernism stands for and be unconvinced by what can be seen as its caricatures of knowledge and its relativism (the last of these is a charge which is usually denied by its followers). Indeed, many of us are still committed to enlightenment, rationality, ideas of progress, seeking logical explanations for educational phenomena and so on. What we cannot do, however, is simply ignore the challenge postmodernism presents to us to reflect on the discontinuities in

our own conceptions of what counts as educational research.

McGaw has given us an account of the research enterprise which provides rather grand frameworks of research traditions and prestigious debates about how the different traditions and notions of truth relate to each other. In describing the OECD work, he talks about 'defining research' and classifying the characteristics of research, and goes on to distinguish 'educational research' from 'research relevant to education'; the latter is identified with the grand narratives of other disciplines (psychology, sociology, natural science, history). The clear implication of the paper is the systematic establishment, through various approaches, of 'true' representations of reality through this activity we call research.

There is, however, something of a mismatch between this disciplined image of research and the kind of practice in which most of us engage. If we regard the ultimate aim of research as being in some sense to improve education (as McGaw appears to do), then our immediate priority is unlikely to be one of establishing a 'true' representation of reality. Our first concern will be to think of research knowledge as a practical matter of learning how to cope with specific aspects of the educational world and of addressing our everyday purposes of practice and policy. The uncertainties and diversities of these activities lead us to design research studies which will, at best, dip into McGaw's research traditions and more generally display a hybrid appearance which focuses on some research problem or question rather than on conforming to a classical paradigm. The Australian examples which McGaw describes in his paper illustrate this point rather well.

It could be argued that the discrepancies between McGaw's models for research and research practice are recognised in his call for eclecticism as a way of strengthening research. If the discrepancies are not addressed directly, however, the grand narratives' discourse can have a considerable impact on the confidence of new researchers. There are dangers that they may be paralysed into inactivity because they have not yet mastered all the characteristic methods or theoretical frameworks of the traditions from natural science or interpretive approaches or whatever. The other possibility is that they endure the private guilt which arises from an awareness that their own research strategy is less than perfect and the generalisable claims they make would not stand up to close scrutiny. Many sit on the edge of methodological or theoretical empires conscious that either they have no passport to join or their passport, if the truth were known, is invalid. They may, of course, gravitate to the research traditions of other disciplines and engage in 'research relevant to education' rather than 'educational research'; such research rarely

influences thinking among the community of educators, however, because it so seldom addresses the questions which are seen as useful by that community.

The main point that I have been making is to draw attention to the need to move beyond the framework outlining formal research paradigms and definitions. The examples which McGaw has cited from Australian educational research are concerned to some extent with classical 'objectivity' and they draw on various theories and research traditions. Their main characteristic, however, is their dependence on pragmatic decisions based on what looks as if it might be useful in a given social context. Whether research will be useful in such contexts will depend on the extent to which there are the conditions in place for it to have an impact on educational thinking, policy or practice. That is the second issue on which I wish to add something to what Barry McGaw has said.

Dissemination and impact of research

McGaw's paper suggests that those engaged in policy development or professional practice tend to have concerns which are practical, sometimes seek enlightenment to inform their actions, sometimes want recipe knowledge to prescribe their action and sometimes prefer knowledge which legitimates choices which are already made. I would argue that research can rarely, if ever, offer knowledge of the 'recipe for action' kind. The notion of an instrumental function for a specific piece of research, in the sense of having a direct and immediate impact on policy or practice, is usually misguided. This is not to say that there can be no impact, but there has to be a prior enlightenment stage.

There are several things that need to be said about the enlightenment process. The first is concerned with the kind of knowledge that educational research is able to provide. It can help our understanding about how things are in education (a mapping of the field) and why they are the way they are. It can articulate the ways in which they might be different and alternative actions or decisions which might be taken to achieve change. It can elaborate on the implications of making choices among those alternatives. What it cannot do is tell policy makers or practitioners what they should decide or what they should do. At a general level, the context in which the research has been carried out inevitably is in some ways similar to and in some ways different from the context in which the decision about policy or practice is to be made. That decision will depend, therefore, on a wide variety of factors as well as research knowledge — personal ideology or preference, availability of resources, competing policies and practices, readiness of others to cooperate

and a specific judgement about the similarity of the context to that of the research.

The contextual differences are minimised, of course, if the research is collaborative between researchers and policy makers or practitioners (the 'collaboration' might be within one person), and especially if an action research approach is adopted, as in Barry McGaw's third example of educational research and development. The 'other factors', however, are still around and will still influence decisions.

To take a rather obvious example, action research has told us a great deal about how we could provide for students with special needs to maximise their access to, and achievements in, education. What determines decisions is not the research knowledge, however, but rather the available resources and entrenched attitudes. A second example, outwith the action research mode, relates to setting and streaming (homogeneous ability grouping) in Scottish education during the last year of the Conservative government (1996–97). The government commissioned a review of research from around the world to be carried out by the Scottish Council for Research in Education (Harlen and Malcolm, 1997). In marked contrast to the findings and messages which came from that review, strong pressure was put upon schools to reconsider their practices of mixed ability (heterogeneous) grouping in favour of grouping by ability (Scottish Office Education and Industry Department, 1996). The excuse that little research of this kind had been done in Scotland was offered as a reason for ignoring the work of the government's own commissioned study; it seems much more likely that other factors (political ideology? distrust of the researchers? the Minister's personal preference?) underpinned the decision.

I am arguing that if the aim is for educational research to have an impact on practice, its ideas and findings have to change educators' *understandings*, and that will not be achieved overnight. It is only when the ideas enter into the commonsense discourse of communities' policy makers and practitioners that impact is achieved. There are three important implications in this for researchers.

First, they have to accept that it is only in cases where researchers are themselves the decision maker that there will be any memory of commonsense ideas having arisen from research. In general, it will take five to ten years for such ideas to take hold and no one will remember, recognise or celebrate their research origins. Researchers have to be patient and accept that their rewards will come from intrinsic satisfaction that eventually their ideas have

been taken up rather than from any public acclaim.

Secondly, it seems unlikely that research findings will make any inroads into professional thinking if they emerge from isolated studies with no sense of coherent programmes. The pattern of idiosyncratic inquiry which has characterised so much research, especially empirical work, has had a major part to play in limiting the impact of research on educational policy and practice. Researchers, but also funding bodies, have to look to the establishment of cumulative studies and replications of work if they are to provide convincing evidence and to collaborate in ways that discourage poor research which contributes neither to theory nor the practice of education.

Thirdly, researchers have to consider much more thoughtfully *how* they can influence understanding in the educational community. If we take the case of impact on teachers, elsewhere, Donald McIntyre and I have placed great importance on their professional craft knowledge (Brown and McIntyre, 1993). By that we mean

that part of their professional knowledge which teachers acquire primarily through their practical experience in the classroom rather than their formal training, which guides their day-to-day actions in classrooms, which is for the most part not articulated in words and which is brought to bear spontaneously, routinely and sometimes unconsciously on their teaching (Brown and McIntyre, 1993, p.17).

The point to be made here is that it is on this professional craft knowledge that research ideas or findings have to make their mark. To achieve that, teachers have to be prepared to listen, to recognise that they can make a difference in their own classrooms, to understand that what research has to offer is educationally superior to what is already in place and to be in a position to judge the 'costs' to themselves of the proposed changes.

Traditional patterns of inspection, in-service, dissemination of research and curriculum innovation, however, have too often rested on a 'deficit model' of teaching apparently aimed at overcoming teachers' weaknesses. Not surprisingly teachers in these circumstances have become defensive and have rarely been in the market for new external ideas which they may see as impractical and at odds with their preferred ways of working. Researchers have to be seen to be taking account of, and valuing, how teachers make sense of their own work and to be creating the conditions under which both the costs and benefits of taking on research ideas and incorporating them into individuals' professional craft knowledge are addressed. The general

constructivist notion of starting where students are, rather than where one would like them to be, is just as important in contexts where research dissemination has to take account of how teachers construe the world of education rather than of the constructs which we would like them to have. A rather similar story could be told about the impact on policy makers.

This section has focused on matters of the impact of current research on policy and practice. In my final area of comment on Barry McGaw's paper I am going to draw attention to two concerns for research in the future.

The future: scope and theory

In the first section of this paper, I suggested that an overview of educational research has to face the intellectual controversy and uncertainty of postmodernism. In addition, such an overview should address the place of educational research in the contemporary world with its changing ideas of democracy, and plethora of social, political, economic and cultural upheavals. The very clear emphasis which educational research has targeted on 'learning', for example, must now encompass an area far beyond formal educational settings to include learning in the home, the workplace, the peer group and the community. And the substantive focus of research on learning may be on matters such as environmental pollution, human aggression, community health, discrimination, prejudice, social justice or economic regeneration as well as the traditional textbook topics. In Scotland, for example, the Scottish Civic Assembly is looking for a Parliament which is *not* a replication of the Westminister model. The essence of the Assembly's approach is the need for the citizenry, as well as elected parliamentarians, to have a central involvement in the formulation of policy. What has educational research to offer in the way of ideas about that?

A much more extended version of this kind of argument is to be found in Ranson (1996). This article, 'The future of education research: learning at the centre', was the outcome of a series of seminars, supported by the Economic and Social Research Council in 1994–95 which brought together many of those most active in research in the United Kingdom. A particular point which was repeatedly referred to in these seminars was the place of theory in inquiries into education.

Barry McGaw, in offering three illustrative examples of Australian research, referred to one as being 'derived from a strong theoretical model' and another (action research) as encouraging practitioners' own theories. In general, however, much of educational research has been weak on theory, and many

researchers have been undecided about what should count as theory. There has been little in the way of intellectual adventures which have challenged the powerful orthodoxies. Yet theories are essential if we are to produce convincing explanations of educational matters, introduce new ideas or concepts and tie them to practice, challenge existing policies and practices, analyse power bases or promote imaginative development. The interaction with theory developments in other social science disciplines has been limited (though with some notable exceptions such as work on the environment, national identity, communications).

A good example of an area which has been self-critical of the paucity of its dependence on theory is school effectiveness and school improvement. The early school effectiveness research was large-scale quantitative work which has been unkindly described as suffering from empirical blight. Very clearly such studies needed to be complemented by qualitative case studies and phenomenological approaches. The addiction to the inspection of variables selected as important by policy makers and researchers had to be broken so that adequate efforts could be made to comprehend how those involved (teachers, parents, students) made sense of their school's world. Only by doing that could researchers come to understand what was going on in schools and how improvements could be brought about. At last, the effectiveness research has begun to collaborate with the school improvement studies and they are developing links with much broader ideas, especially theories of culture from a range of other fields.

In conclusion, Barry McGaw's stress on the need for 'changes in research effort' and 'improving links with policy making and practice' is something with which I would concur. However, my emphasis for a strategy to strengthen educational research and development would be to encourage researchers to be less timid and more inventive about how to assess the impact of their work, more comprehensive and creative in their conceptions of the relevance of educational research and much more attentive to theory, including theory developed in other fields. But most of all we have to be clear that this is a field which is not static. To say that it is in transition is perhaps to mislead: the changes are not transitory phenomena; they reflect something which is as dynamic as education itself.

References

BROWN, S. and McINTYRE, D. (1993). *Making Sense of Teaching*. Buckingham: Open University Press.

EAGLETON, T. (1996). *The Illusions of Postmodernism*. Oxford: Blackwell.

HARLEN, W. and MALCOLM, H. (1997). *Setting and Streaming: a Research Review*. Edinburgh: SCRE.

RANSON, S. (1996). 'The future of education research: learning at the centre', *British Educational Research Journal*, **22**, 5, 523-35.

SCOTTISH OFFICE EDUCATION AND INDUSTRY DEPARTMENT (1996). *Achievement for All: a Report on Selection within Schools*. Edinburgh: Scottish Office/HMSO.

Neville Postlethwaite, University of Hamburg

I would like to congratulate the NFER on its fiftieth birthday and for having brought together such an élite group of educational researchers from across the world. Floreat NFER! Second, I congratulate Barry McGaw on his erudite and insightful paper. In response, I prefer, with your indulgence, to split my response into two parts. Part 1 is a trip down memory lane to show what it was like at the NFER when I was working there some 34 years ago. In Part 2 I would like to expand on one point from Barry McGaw's presentation in terms of some of my own practical experience in working with ministries of education in the area of policy-related research.

First experiences

I began work as an educational researcher at this august institution — the NFER — in 1961. I had been a schoolteacher and after four years of teaching had come to the conclusion that teaching was not for me for a lifetime. While I was a teacher I had undertaken a piece of research on au pair girls. It would probably fall into the category of 'social psychological research'. I was lucky in that I could go to the London School of Economics and ask Oppenheim for advice whenever I wished about questionnaire construction. I learned all sorts of things about measurement and scaling and computing, but I did not learn about probability sampling as the *Economist* was quick to point out when they reviewed the publication in 1962 — no standard errors of sampling! When the report was published, I was on TV and the radio, and the pressure groups for enacting legislation about au pair girls discovered me. They asked me detailed questions about the research results. Legislation ensued. This is probably the only time in a lifetime of research that legislation ensued directly from any research that I have undertaken or had to do with. And it was based on research from a judgement sample! Do legislators grasp at straws?

I had scaled several questions to form a scale of 'happiness with time in England'. So, the NFER asked me to join their test development division. I told them that I knew nothing about educational testing but they assured me that this did not matter. I think that I gained the impression that if you've scaled once you can scale anything! So off I went to 79 Wimpole Street, just down the road from the Barretts — so to say. My boss was Douglas Pidgeon and the Director of the NFER was Bill Wall. I remember that on my first day, eager to please as I was, I wrote 200 intelligence test items and I would like to point out that 150 survived the trialling exercise!

Then one Friday, Wall came down the corridor and asked me to write a proposal to do a piece of research. He wanted the whole proposal on the coming Monday. So, I wrote a 20-page document that weekend and handed it in on the Monday to Wall's secretary. She was meant to type it and then I assumed that I would discuss it with Wall since there were several parts of which I was not sure at all. On Thursday I popped into Vanessa's office to ask if she had finished typing it. Not only had she typed it but Wall was, at that very moment, at the Ministry of Education discussing the proposal. He returned later that day with a grant of £56,000 to do the research. This sum was exactly 56 times my annual salary. And so were research proposals written in those days.

I should add that by the end of 1962 I had met the famous Gilbert Peaker, the one who 'invented' 'Design Effect' in sampling before Leslie Kish did (see the *Royal Statistical Journal* of 1952) and was subjected to intensive instruction in the principles of sampling. This also involved my going up what appeared to be huge mountains in Cumberland with Gilbert ascending the slopes at a very fast pace and saying 'Now, boy, put the following matrix into your head and we shall take the determinants while walking up'! This instruction continued over the next nearly 20 years as Peaker and I worked together.

I was actually rather poor at that time and the salary was such that I had difficulties making ends meet. So, I went away at the end of 1962 to foreign parts and joined IEA. I have never been back to England to work since then so what I am about to say may have little relevance to the English scene but my remarks are based on experiences that I have had in many countries where I have worked.

Reactions to paper

Some general remarks

I agree with all that Barry McGaw has said. Education is not a discipline but a field of inquiry. It uses the empirical and hermeneutic approaches. It draws heavily on social science research methods.

Although experiments are in theory desirable to answer many questions in educational research, they are in practice very difficult to carry out with strict randomisation and sufficient cases in the treatment and control groups. When I first read the recent report of the Tennessee 'class size' experiment I was much impressed. But, soon after that I saw Sig Prais' review where he had

gone to a lot of detective work to find out how the groups changed over time and at that point I no longer had confidence in the results. Indeed, all other research would appear to show that between about 15 and 40 pupils per class there is no effect on achievement, and even under 15 pupils per class achievement only rises on condition that the teachers individualise instruction. This, however, does not prevent political parties from promising smaller classes — probably because teachers like smaller classes!

On the other hand, sample surveys and even school censuses rely on what Kish called 'natural variation' and sometimes there is no variation or some variation depending on one or two outliers. I am at present struggling with some multivariate analyses and try as I may I cannot explain more than 20 per cent of the variance — but many of the independent variables have low means and very little variance!

Ideologies are often more important than research findings (or lack of research findings) when it comes to making policy. I have been hearing of the Victorian school system reform where schools have almost total power — control of their own budgets, the hiring and firing of teachers, and to a great extent the interpretation of a syllabus. The decision to do this seems to have been taken on ideological grounds, but I am sure that Barry will correct me if I am wrong. But I wonder what the effect will be on the standard deviation of achievement, especially between schools. Will schools become so different that it will be virtually impossible for parents to move from one part of the state to another, especially since the economic performance of a state often depends on job mobility? Why was some small-scale work not undertaken to try and identify possible consequences of the action before it was undertaken?

Much small-scale research can produce good ideas — the identification of a problem in the schools, the production of new measures, and so on. But to what extent can the results of such research be generalised to the whole school system or to several subject areas at any one cycle of education? This brings us back to probability samples!

Ministries' more direct needs

If I am working in a ministry of education, I am responsible for the allocation of resources to schools or at least to regions. I have a constrained budget. I will be loath to allocate resources on the basis of research that deals with only one or two grade levels in one subject area. I need to know the effect of 'manipulable' variables on achievement in a particular school type or school cycle and for as many subject areas as possible. And I want big effects, not small ones.

I do not particularly care which form of multivariate analysis has been used as long as I know the limitations of the approach used. John Keeves, another Australian whom Barry knows well, has recently become wedded to HLM, but a few years ago he analysed one data set using OLS, LISREL, PLS (between pupils and between schools), and HLM at two levels. I was much relieved when I plotted several path coefficients from the different analyses to see that the results were almost identical. Those factors having big effects were the same. There were some minor deviations for the smaller effects, but I was not interested in the smaller effects.

I also want to know about variation in the system. Most governments are interested in equity. I am most interested to know the differences between regions in terms of both inputs to schools and outcomes. At the same time I want to know which of the inputs (and processes that I can control) have most impact on achievement now. I realise that any outcome is usually the result of many factors but I do not have sufficient money to deal with all factors now. I need to know those having the biggest effect. But, I also need to know the cost implications. I know that it is very difficult to get accurate income and expenditure data from schools and even sometimes from regions, but I do need some idea of the magnitude of the costs. And it is not helpful when researchers tell me that costs are not their business.

Barry McGaw has hinted at much more cooperation, and since I have been doing just that since I ceased my IEA responsibilities I should like to comment on some of my experiences.

Working with ministries

Ministry constraints and first meeting with ministry personnel Ministries have two problems not often appreciated by researchers. The first is that they are like a hydra-headed monster (to quote C. E. Beeby, another Antipodean) in that there are several different departments and they often do not know what each is doing. There is the buildings department, the supplies department, the teacher training department, the standards and control department, the curriculum department, the examinations department, and so on. They often do not talk with each other. It is sometimes difficult to know to whom to speak. In general, I have found it desirable to get either the Minister or the Permanent Secretary (or both if they are at loggerheads) to bring together the heads of departments or divisions in order to have a half day's discussion on what it is that they want from a particular research project. When there is unclarity this gives the researchers the opportunity to try and clarify their thinking. The second problem they have is that they spend most

of their lives dealing with today's crisis and they do not have time to read research very much or to think about the major issues where research could provide some relevant information. The half day's meeting can give them some time to think or at least become motivated enough to say that they will discuss the problems of issue identification and that the researchers should return for further discussion in a couple of days' time. The researchers need to handle these meetings with great sensitivity and listen carefully.

It should be added that there are some ministries where the higher echelons have not yet recognised what research and evaluation can do for them. They have failed to identify issues where research is needed. I was recently undertaking an evaluation of the research and evaluation of a ministry where fairly good young data processors were to be found in the ministry expending a lot of energy undertaking meaningless tasks even though they had databases that could have been used immediately for elucidating some policy questions. In these sorts of situations it is up to the researchers to begin to educate the people in the higher echelons — a very delicate task.

If this meeting goes well the rest is *relatively* straightforward.

Produce dummy tables I am a big believer in dummy tables. First of all, it is important to formulate a series of research questions to cover the gamut of the questions that resulted from the ministry meeting. They should of course then be discussed with the ministry people. These questions are best formulated in lay terms, i.e. not as hypotheses as researchers are often wont to do. There should be one or more dummy tables to answer the question(s) posed. The advantages of dummy tables are that they show the measures needed, the forms of analysis required, and they help to avoid the collection of data on variables that will never be analysed. (An aside: in my experience it is very difficult to get any major effect of individual teacher variables on achievement, but when school constructs are formed for all of the relevant teachers in the school, big effects are often found even at a disaggregated level of analysis. Only twice have I been in the lucky position of having been able to measure teacher subject matter knowledge, but in every case this measure has had an enormous effect on pupil achievement.)

Careful construction of measures It goes without saying that the measures should be constructed very carefully with appropriate trialling. There is a tendency for some researchers to think that because they speak their mother tongue well that they can construct questionnaires. Such people should not be allowed to construct questionnaires. Often a lot of time is devoted to test construction because there is a known technology. There is little technology

93

for questionnaire construction and the secret is pilot, pilot, pilot. Questionnaire construction for international instruments is even more complicated.

Careful definition of the target population and good probability sampling The required standard errors of sampling should be well thought through and then applied. (An aside: if equity is important in the research then every effort should be made to have a good estimate of the within-grade, within-school variance. I have observed that several researchers say that it is too difficult administratively to test a random sample of kids drawn from all classes in the grade. In my experience this is not so. Yes, effort is needed but it can be done. The danger of not doing this is that the variance within schools is less than it should be and then the intra-class correlation for estimating the between-school variance is artificially high.)

Good data collection Of course missing data can be imputed but it is an enormous job to do it properly. It is much better to expend more effort at the data collection stage by having the administrators ensure that there are no missing data in the factual questionnaire data. There is no control over the data collection for the descriptive and attitude measures or for the test data. However, it is possible to build in all sorts of checks during the data collection phase to minimise missing data.

Careful data entry, cleaning, weighting, and merging This is easy to say but the number of times that things go wrong is surprising. Let me add that in my experience the sort of people who are good at data entry and cleaning are a special breed of people and need to be nurtured. If you have a good team for this kind of work and allow it to disintegrate it is virtually impossible to rebuild.

Data analyses for completing the dummy tables Once the data are clean and the data analyses have been pre-planned, then the production of the various analyses should be easy going. Each estimate should also have an accompanying standard error of measurement. Now that it is relatively easy to calculate standard errors of sampling I rarely read research reports that do not include them because I have the feeling that the researcher does not know what he or she is doing! At the same time, there are researchers who have taken a two-stage cluster sample but report the standard errors produced automatically by SPSS or SAS, and these are of course the wrong standard errors!

Writing the report As we all know, reports should be written in simple language that can be understood by all. When working with ministries it is essential

that the policy suggestions emanating from the research are stated. This is an area where many good researchers fail. There seem to be few people who can turn the research results into suggestions for policy and action. The action also needs to be specific. This is where the researcher needs the help of the ministry people again. They know the educational legislation and what is or is not possible. It is often useful to report the policy suggestions using the following categories:

Cost/Time	Short term	Medium term	Long term
Low cost			
Medium cost			
High cost			

Review by ministry personnel These policy suggestions are then reviewed in detail by the ministry personnel together with the researchers. Again, in my experience this is where the ministry people have excellent perspectives in order to make the policy suggestions workable. A one-day workshop is usually sufficient time for this exercise.

Amending the publication The policy suggestions are then amended and the research report produced.

Ken Ross (another Australian) and I undertook one such exercise where the whole job was accomplished in 11 weeks from the Minister and division heads telling us what they wanted to our producing a three-volume report for discussion at the ministry. The probability sample included 150 schools and 20 pupils per school drawn from across all classes within the target grade. This was a bit of a 'tour de force' and as soon as a number of countries wished to undertake a similar exercise collaboratively the time frame went up to two years. This is but one example of how to work collaboratively with a ministry, but Ken and I have found that it works.

4.

The psychometric enterprise — or the assessment aspiration

Chris Whetton, NFER

This conference is to be a celebration of 50 years of the work of the Foundation. As with all fiftieth birthdays, celebration should be somewhat tempered by reflections on what has been, what is, and what can be made of the future. I shall follow this temporal structure but see this not simply as a catalogue of what has changed in testing and assessment but as a reflection of the educational system of the times.

The theme of this talk was given to me as 'The psychometric enterprise'. However, I added the subtitle of 'The assessment aspiration'. It has been asserted that the last ten years or so has seen a paradigm shift from testing to assessment; from psychometrics to an educational measurement (Gipps, 1994). The psychometric model has its roots in psychological measurement and as such it tends to assume characteristics of the mind. It concentrates on comparisons between people, generally using a norm-referenced approach, and emphasises the importance of standardisation and reliability. In contrast, educational assessment is less well defined. It has as its core the idea that assessment should be closely related to, and support, learning. It is considered to be a dynamic rather than static process. The terms 'authentic assessment' and 'performance assessment' cover part of the universe, emphasising that higher-order skills and cognitive processes are more important than testing knowledge and memory of facts. These are the types of processes frequently aspired to by modern advocates of educational assessment. Further, they are recommended as a means of influencing the curriculum and raising standards (e.g. Resnick and Resnick, 1992).

I propose to concentrate on developments in England and Wales, not for parochial reasons but because this has been the main concern of the Foundation's work and it is this that I will use to illustrate what I believe is a general set of trends both in Britain and internationally. Much of this work has been on the assessment of primary-age children, but there are parallels with the main assessment enterprise in Britain, school leaving certification. GCE O- and A-levels, CSEs and finally GCSE could be shown to have followed the same broad pattern, although always with the additional main purpose of providing a certification process for entry to the workplace or higher education. I will not, though, attempt to deal with international

comparisons. The Foundation has served as the English arm of the IEA and organised many of its studies but these are well described elsewhere (e.g. Husen and Postlethwaite, 1996).

The production of, and research into, educational tests have been a constant feature of the Foundation's work over its 50 years. Indeed, one of the original purposes in establishing the Foundation was to provide an organisation to develop tests to meet the needs of local educational authorities (LEAs).

These 50 years may be characterised as far as assessment is concerned as falling into three overlapping eras, each reflecting the prevailing educational ethos of their time.[1] These are not discrete and have no firm start and end dates, rather they are changes of emphasis with legacies from earlier eras continuing to influence directly and indirectly. The first era, which ran from the late 1940s through to the mid 1970s, had selection as its underlying purpose and model. At that time, secondary education in Britain was characterised by selection of children at the age of 11 to attend academic grammar schools. Those children not selected continued their education in secondary modern schools. The style of test used for this selective process emphasised the standardised, norm-referenced reliable measurement of children. These were high-stakes tests on which a great deal of a pupil's future could depend. This type of testing continues today in a minority of LEAs and schools where selection has been retained. In fact, recent emphases on selection by schools may mean a reintroduction of, or an increase in, this form of testing.

The second era was that of monitoring, running from the mid 1970s to about 1990. This encompassed both national monitoring and LEAs conducting their own testing programmes for monitoring purposes. The nature of the tests became more open-ended, exploring pupils' capabilities on a survey basis, including some novel performance-based tasks. This was unthreatening low-stakes assessment for the benefit of the educational system with little or no impact on the individuals being tested.

Finally, from around 1990 there have been two, sometimes contradictory, themes. Much of the work on testing has been in response to the market-based reforms of recent governments. These reforms have encompassed local management of schools, a reduction in the powers and role of LEAs, the introduction of a National Curriculum, a standards-based assessment regime, publication of results and regular external inspections of schools with the resulting reports published. The philosophy behind these is that of competition causing standards to rise (Lawton, 1996). Hence, one major purpose of testing became for accountability. In contrast during this same era, for many in

education, the emphases have been on the integration of assessment processes and outcomes into the curriculum and teaching. This movement too has at its heart the belief that assessment processes can lead curriculum change and that this is essential for raising standards. The two models are based on differing sets of values and beliefs. The 1990s have been characterised by conflict over these two sets of values which has occurred in many countries (Madaus, 1996). However, it has perhaps been in Britain that the conflict has been fiercest. A truce is currently in place, but it may be perhaps one of those truces which allow rearming rather than leading to a peace. A successful resolution of the contrasting approaches embodied by these two models has yet to be achieved. Later in this paper, I will elaborate on this conflict with special reference to the testing of seven-year-olds in England and Wales.

I should first like to expand on these three ages of assessment. The first era, that of standardised tests for selection, arose from the need for such tests with the increasing numbers of grammar schools in the 1940s and 1950s. By the early years of the 1950s, NFER tests were available in Mathematics, English, 'Verbal Ability' and 'Non-Verbal Ability'. These were trialled and standardised in conjunction with local education authorities. A similar rival exercise flourished in the North of England and Scotland, based on tests produced at the Godfrey Thomson Unit in Edinburgh. These were known as Moray House Tests.

The underlying assumption of this use of tests, and indeed of the selective process of the time, was a psychometric one (Simon, 1996). The work of British psychologists in the 1930s had been influential in promoting a view of intelligence as a relatively fixed, innate capacity which determined a ceiling on children's attainments (Sutherland, 1984). The two most eminent British psychometricians of the time, Sir Cyril Burt and Sir Godfrey Thomson, were both Vice Presidents of the Foundation by 1953. Hence, the tests of the time were ability based rather than curriculum focused. They were often, to our modern eyes, very culturally biased, being by today's standards extremely verbose. They were toward the objective end of the spectrum, with tightly defined marking keys, items generally requiring a one-letter, number or word response. The marker reliability for the tests would be likely to be extremely high, although with lower requirements for reporting no information is given on this in the manuals or evaluations of the time. Thus, the prevailing imperative was for selection devices which were reliable. This is not to say that validity was not regarded as important. Almost the first research project established at the newly founded NFER was on the accuracy of the selection process.

These early researches culminated in a major piece of work by Yates and Pidgeon (1957) which looked at the validity of the selection procedures, which tests gave the best predictions, and how the school's own gradings could be incorporated into the selection process. It is a tribute to the robustness of these findings that 30 years later LEAs which had remained selective were still using the administrative procedures recommended by this research.

It is interesting to note that even in what was essentially an actuarial enterprise of prediction, what we have since come to term consequential validity had some importance. There were advocates of the use of English essay examinations in the selection process both because of criticisms of the 'backwash' effects of objective tests in narrowing the curriculum and lowering expectations of pupils and because of increases in predictive power to later examinations which they were said to provide.

Interestingly, in the light of current efforts to integrate teachers' National Curriculum assessments of their pupils with test results, in many LEAs, schools' ratings were incorporated into the rank ordering of the pupils. It was left to schools to determine how they undertook the assessment and arrived at their judgements. However, these were then scaled, using the test results for the school, before the combined test and schools' ratings were used in the selection process.

A second strand of concern at that time was the effects of coaching on test results and much work was devoted to defining the parameters of the effectiveness of coaching. NFER studies were by Watts, Pidgeon and Yates (1952), Yates (1953) and Yates and Pidgeon (1957). Other results are summarised in Vernon (1960) or Jensen (1980). These studies distinguished between practice — that is taking the same or similar tests beforehand — and coaching which is instruction in test taking and analysing and answering items. They demonstrated that coaching to improve scores on intelligence tests is ineffective without practice and that coaching is most effective for naïve test takers but can lead to gains of four or five points of standardised score. Practice alone can lead to similar gains but there are diminishing returns from repeated practice, although slight gains occur up to five repetitions. Again these types of study were important in that they informed LEA procedures and regulations to schools on coaching pupils. Some adopted rules which did not allow coaching or practice within the primary schools. Others attempted to be fair to all pupils by prescribing fixed amounts of coaching and practice testing.

Since the tests were required to select among children fairly, the effects of age had to be removed and the familiar standardised score table was a

necessary part of all tests. The procedure for establishing these tables was formalised by Lawley (1950) and remained in use until the late 1980s when it was replaced by a more sophisticated technique, devised by an NFER statistician (Schagen, 1990). The use of such tables has so entered the psyche of the teaching profession that, 40 years on, they have demanded such tables for National Curriculum tests, which as shall be seen later were intended to be interpreted in a criterion-referenced fashion.

Indeed, the importance of this era for a discussion of modern assessment is that its imperatives, which were necessary for selection, shaped the nature of the tests as having to be objective, reliable, often ability-based (with an underlying notion of children having a fixed potential), with a norm-referenced philosophy and an age-adjustment to give standardised scores. Although selection began to be rejected in the 1960s and 1970s as an educational solution, the nature of assessment associated with it did not start to be challenged until the beginnings of the criterion-referenced movement in the US (Glaser, 1963) began to cross the Atlantic or the views of Wood (1986) and Nuttall (1987), described below, gained currency.

Back at the NFER, alongside the development of tests for selection, other series of mathematics, English and, especially, reading tests were produced for publication. These spanned the primary years, giving teachers and schools measures of the progress of their pupils against national norms. They were the beginnings of the use of tests for monitoring, for screening for pupils to receive special or remedial treatment or even, to a very limited extent, for diagnostic purposes. The first marketings of these tests led to a test agency within NFER, then to a separate NFER Publishing Company and finally, in 1981, to the NFER-NELSON Publishing Company, a joint venture with the International Thompson organisation. From 1981, all the commercial activities of marketing the tests have been undertaken by NFER-NELSON, while the development of, and research on, the tests remains within NFER. The use by LEAs and schools of NFER tests continues to the present. They are used for screening pupils, monitoring classes and year groups, transfer information, checks on progress and, most recently, for value-added analyses of school effectiveness.

It must be said that what characterised the users of tests in schools in the 1970s and early 1980s was a lack of clarity of why they were using them. Typically, a test would be administered, the results of the class would be recorded and filed but only used to show to individual parents as a defence against criticism. Research by Gipps *et al.* (1983) showed that, although tests of reading and mathematics were widely used, understanding of the

meaning of the scores or of issues such as reliability and validity was very limited.

In the 1970s, there were two signals of the start of the next era of assessment work in Britain and at NFER: the monitoring era. These were the great debate in education initiated by Prime Minister Callaghan's Ruskin College speech (1976) and the widespread introduction of comprehensive schools. As the need for tests for selection declined, the annual cycle of production of tests in verbal reasoning, mathematics and English ceased. It was replaced by a much more customer-led notion of testing in which LEAs were able to select the questions to suit their purposes and local curriculum. They did this from *item banks,* which are sets of categorised, trialled and calibrated questions of known difficulty and performance. LEA officers were able to be guided in selecting appropriate questions which could form a unique test available to just one user. This concept of item banking was originally expounded by an NFER researcher in an article in *Nature* (Choppin, 1968), although surprisingly the term had been around for over ten years before this, being used in the NFER's Annual Report of 1956. For a time, the Foundation was at the forefront of utilising a new statistical methodology, Item Response Theory, which was just becoming available through the development of computing. Originally, this was done through the use of Rasch scaling of the items but its use for this purpose was criticised and ultimately discontinued. I will elaborate on this below, but first must make a slight detour to discuss two theoretical developments which occurred at around this time.

During the last 30 years the two pillars of test theory have become Item Response Theory and validity theory. They are completely different, with differing starting points, but share the attribute of being impenetrable to all but the very dedicated.

The development and then use of Item Response Theory, largely in the USA, have allowed statistical solutions to a great many measurement problems (Lord, 1980). Effectively, such models attach a number of parameters to each question in a test which describe the outcome when a person who has a given level of ability attempts to answer it. These parameters are relatively invariant and can therefore be used when the question is used in another test. This gives great advantages in attacking several persistent problems in test development. It allows the most efficient selection of items for a test, it allows tests to be linked together or equated and it allows individualised, tailored testing. However, it also has limitations. It enshrines the notion of a test as a collection of independent items which do not interact together, limiting approaches to test construction and question format. It also assumes

unidimensionality — that is that a single statistical attribute is being measured. This may not reflect the reality of educational data.

Views of validity have also altered over the last 50 years. Initial theories clearly separated three types of validity: content, criterion and construct. Each of these was appropriate in its own place. Content validity was supreme in measures of curriculum-based attainment, criterion validity for selective purposes and construct validity where relationships between variables could be fitted into a theoretical framework.

In the 1970s, Cronbach began to stress that validity must be seen in relation to fitness for purpose (e.g. Cronbach, 1970, 1971). The reason for using tests is to take decisions and there must be evidence that the test adds to the accuracy of these. Hence, reference must be made not to the validity of a test but to the validity of the inferences drawn from it. The three types of validity had counterparts in being types of validation evidence. Gradually, validity became unified and regarded as a single unitary concept. As Cronbach (1984) put it: 'The end goal of validation is explanation and understanding. Therefore ... all validation is construct validation.'

Messick (1980) accepted this unified view, but added the important notion that the appropriateness, meaningfulness and usefulness of the inferences depend as well on the social consequences of the testing. Hence, social values and social consequences cannot be ignored in considerations of validity. This integrated the notion of consequential validity to the other three aspects but within a unified framework. Messick (1989) also went on to separate evidence and consequences as one facet of validity with test interpretation and test use as another facet. This leads to his famous two-way classification of the conceptualisation of validity (Table 1).

Table 1: Facets of validity

	Test Interpretation	Test Use
Evidential Basis	Construct validity	Construct validity + relevance/utility
Consequential Basis	Value implications	Social consequences

What is being emphasised here is that in evaluating validity evidence, the values involved in test interpretation and the social consequences of the use of the test are central. This becomes the legitimisation of those who would seek to change educational practice through changes in assessment practice.

Much of this theory is very difficult to incorporate into day-to-day test development. However, it can come into its own in evaluations of testing systems and in providing a framework for understanding the changes in assessment practice.

Returning now to the historical narrative: throughout its early life, at irregular intervals, the Foundation had conducted national surveys of reading standards (reviewed in Start and Wells, 1972). These had generally used a repetition of the same short sentence completion tests, but by the 1970s these two tests were becoming antiquated. There had also been surveys of mathematical attainment, but by the late 1970s there was considered to be a need for a much more widespread monitoring of many aspects of the curriculum. The government set up the Assessment of Performance Unit (APU) in 1975. Its terms of reference were '*to promote the development of assessing and monitoring the achievement of children at school and to seek to identify the incidence of underachievement*'. It is interesting to note the emphasis on measuring levels of attainment but the absence of either a proposal or mechanisms for raising standards.

Project teams were set up in mathematics, language, science, foreign languages and design and technology. Three of these teams were based at the NFER and, between 1979 and 1989, these undertook six surveys in mathematics, six in language and three in foreign languages. A review of the work of the APU is given by Foxman, Hutchison and Bloomfield (1991) and an evaluation of its early years by Gipps and Goldstein (1983).

A major success of the teams at NFER was to broaden the notion of testing from written tests alone to open-ended assessment tasks undertaken individually or in small groups with visiting assessors scoring the process. Hence, alongside written tests, aspects such as practical mathematics and speaking and listening in English could be assessed. The teams' statisticians also demonstrated that, through the use of complex survey designs using many linked tests, a complete picture of national performance in a subject could be derived, giving much richer and more complex information than possible from a single test. This was achieved through the use of samples of pupils, rather than the complete population.

At NFER, the initial means of scaling the items, linking them and the tests and for reporting on performance was through one-parameter Item Response Theory (known as Rasch: see, for example, Wright and Stone, 1979). This was also to be used in making comparisons over time. The use of Rasch was criticised in various forums, principally by Harvey Goldstein (Goldstein and Blinkhorn, 1977; Goldstein, 1979; 1980; 1983). The criticisms centred around a number of concerns. First, that the use of a statistical model to select items would distort the curriculum content of the tests and therefore the monitoring process. For example, it is well known that mathematics items involving spatial concepts often do not fit together with other mathematics items, but these questions should not be excluded from the assessment. Second, where tests are too easy or too difficult, biased estimates of ability might be obtained. Third, that the underlying logistic model might not be the most parsimonious. Other distributions might be used. To practising test developers, all these did not seem to be problems — merely examples of what would happen if someone was applying inappropriate rigid rules to what is generally a judgemental process. However, the central criticism concerned the assumption of unidimensionality in the data. It was argued that this assumption was not sustainable in attainment tests and that since the tests had more than one dimension, estimates using one dimension were biased. (This is true of all assessment procedures which arrive at a single score to summarise attainment, including classical test theory.) More importantly, over time, the procedures for replacing items which have become less relevant could lead to the dimensionality of the set of items changing over time, so that the Rasch model, or any other item response model, would not be applicable. As a consequence of these criticisms, the use of Rasch for linking data between years was abandoned, although it continued to be used for analysis within years.

The APU Science team utilised a different analysis method, Generalisability Theory, as a method of making comparisons over time. This did not suffer from the same criticism but, none the less, the project had considerable difficulty in addressing questions of change over time (Foxman *et al.*, 1991). The Foreign Languages team also used an alternative to Rasch, Efficient Estimates methods (Goldstein and James, 1983).

I have dwelt on this story for several reasons. First, it helps to explain why the monitoring approach began to be regarded with suspicion. Much professional advice was saying that it was logically impossible to report on changes in standards with any precision (e.g. Nuttall, 1986). Alongside this, modern psychometric methods were seen to have failed. The use of Item Response Theory was certainly discontinued within NFER even outside of

the APU projects. Since that time, development and use of Item Response Theory have been much greater in the USA, for example at ETS, in Australia and in Europe. A much more robust view of the utility of Item Response Theory compared with its theoretical shortcomings seems to have prevailed.

However, in Britain, these attitudes had something of an influence on decisions to do away with national monitoring at the time of introducing the National Curriculum and on the nature of National Curriculum assessment as an almost anti-psychometric judgement-based system. The argument was advanced that the blanket testing of every child could serve the monitoring function of gauging changes in standards.

Having said that, the models of assessment developed by the APU teams and the information they secured on the performance of children in the curriculum had a major influence on views of assessment. Foxman *et al.* (1991) set out the key notions to which all subject teams were committed:

- the centrality of *communication*, through language, graphs, diagrams or symbols
- the purposefulness of *interpreting and solving problems* — challenging pupils to work purposefully toward a goal, the path to which is not precisely defined
- the role of *practical situations* so that assessment takes place in context and in relation to specific tasks.

Foxman *et al.* (1985) set out a taxonomy for classification of mathematics tests and items. This is shown in Table 2 and indicates the wide range of modes of assessment envisaged in terms of presentation, method of working, response, type of interaction and timing.

Table 2: Taxonomy of assessment modes

Presentation	Mode of Working	Response	Interaction	Time Constraints
Printed	Written	Written	Pupil–examiner	Timed
Aural by tape	Oral	Oral	Pupil–teacher	Untimed
Aural by assessor	Keyed Practical	VDU Construction	Pupil–pupil	Extended
VDU	Calculator			

From Foxman et al. *(1985)*

106

One of the main innovations was the use of assessors who conducted individual or small group tasks. The chief purpose of this was to observe the process of the child attempting to answer a question in order to determine the methods used and the exploratory procedures adopted. In fact, these processes were the principal object of the mathematics assessment, rather than any sort of enumeration of the percentage answering correctly. This is not to say that there were not written tests. However, even these were designed increasingly to include open-ended short response items which lent themselves to an analysis of the errors made.

Similar innovations occurred in the APU Science surveys, which adopted a view of science as an experimental subject concerned with problem solving. Again, practical and written tests were used, in both cases with a considerable emphasis on exploring the processes children use to arrive at answers.

In the Language surveys, previous narrow conceptions of reading which had utilised only sentence completion tests were replaced with measures which regarded reading as a purposeful activity with an emphasis on putting reading materials to practical use. They eschewed short unrelated passages as stimulus material in measures of comprehension. Instead reading booklets were produced which included works of reference, works of literature and everyday reading materials. Questions required written responses but also the completion of tables of information or labelling of diagrams. To assess writing, four tasks were generally required: a short piece of writing, a more extended piece, a text-editing task and questions about writing preferences. The tasks varied in the freedom given to pupils to define subject matter and in the form of the writing required.

These forms of assessment contrast markedly with those used for selection in the previous era. The emphasis had swung away from manageability and reliability toward validity, away from objective short answers to open-ended lengthy tasks in which the process was as important as the answer. But, crucially, they had also swung away from time-limited, group-administered tests toward individual or small group assessment tasks. These were affordable in surveys using light sampling but the lessons being drawn, perhaps implicitly and certainly incorrectly, were that they could be appropriate even if all pupils in a cohort were to be tested.

The forms of assessment also paralleled what was being written about educational testing. Wood (1986) presented a contrast between the requirements of psychological measurement and those of educational measurement. Psychological measurement is attempting to produce pure

measures. It thrives on unidimensionality and its statistical procedures emphasise internal consistency and reliability. To do this, variance in the measures is required and individual differences are paramount. In contrast, measurements of attainment are complex and messy. They are multidimensional, resist reliability, are responsive to the child's situation and the teaching received and, in an ideal world, produce little variance. Wood then proposed a set of criteria for educational measurement. It:

- deals with the individual's achievement relative to himself rather than to others

- seeks to test for competence rather than for intelligence

- takes place in relatively uncontrolled conditions and so does not produce 'well-behaved' data

- looks for 'best' rather than 'typical' performances

- is most effective when rules and regulations characteristic of standardized testing are relaxed

- embodies a constructive outlook on assessment where the aim is to help rather than sentence the individual.

These themes picked up the trend of the times. They were being implemented to some extent in the British external examination system and were largely in tune with the values of the APU.

Nuttall (1987) took these ideas forward, suggesting that the requirements of validity were that best performance must be demonstrated and, in order to do this, the task and conditions of assessment should be such that:

- tasks are concrete and within the experience of the individual

- tasks are presented clearly

- tasks are perceived as relevant to the current concerns of the student

- conditions are not unduly threatening, something that is helped by a good relationship between the assessor and the student.

He concluded with a ringing endorsement of this approach.

> With the benefit of hindsight, it seems strange that so much effort should have been put into the development and validation of general paper-and pencil tests, when everything points to their artificiality, their remoteness from the nature of any normal job and their unelaborate

conditions of administration. It seems likely that considerations of utility and reliability have prevailed over considerations of validity. The signs are now that validity is claiming its rightful pre-eminent position, and that the careful specification of the universe of generalization is helping to stimulate improved conditions of assessment and more thought about evoking an individual's best performance (Nuttall, 1987, p.155).

All of this chimed in with the unified theory of validity, and its consequential components, as outlined earlier. The prevailing view then was of the primacy of validity within a new form of educational measurement — the assessment era was about to dawn.

The influences of the APU types of assessment and the academic writings on assessment were apparent in the views of the group which in 1988 proposed a structure for National Curriculum assessment and in the work of the groups devising the curriculum itself. The underlying government notion was that results would be used to compare schools and LEAs as a means to raising standards. However, in contrast, perhaps fuelled by the APU experience, teachers' desires were for a broad range of assessments and for useful diagnostic information.

Since this is an international conference, I should perhaps first give a short account of developments in England and Wales in National Curriculum assessment from 1988 to the present. The Education Reform Act specified that there should be an assessment system with testing at or about the end of four key stages, which were: five to seven years; eight to 11; 12 to 14; and 15 to 16. The form of this testing, the uses of the results and all other details were left open to be decided by the Secretary of State for Education. Accounts of this era are now beginning to be published. These give the story from the viewpoints of various participants in the events. Among them are: the politician's story (Baker, 1993); the chief executive's story (Graham and Tytler, 1993); the Council member's story (Daugherty, 1995); the union official's story (Barber, 1996) and finally the test developer's story (Sainsbury, 1996). In addition there were evaluation projects examining different aspects of the process (eg. Shorrocks *et al.*, 1992; Whetton *et al.*, 1992; Ruddock and Tomlins, 1993; Gipps *et al.*, 1995).

An initial blueprint was provided by a group known by the acronym TGAT. This stood for Task Group on Assessment and Testing and was composed largely of educationalists. The group was chaired by Professor Paul Black and consisted of two directors of research or assessment organisations (one

of whom was the Director of the NFER, Clare Burstall), two headteachers, a chief education officer, an ex-chief HMI for Primary Education, an economic researcher, an emeritus professor of electronics and a personnel director of an engineering firm. The terms of reference of the group were to advise on:

... the practical considerations which should govern all assessment including testing of attainment at age approximately 7, 11, 14 and 16, with a national curriculum; including the marking scale or scales and kinds of assessment including testing to be used, the need to differentiate so that the assessment can promote learning across a range of abilities, the relative role of informative and of diagnostic assessment, the uses to which the results of assessment should be put, the moderation requirements needed to secure credibility for assessments, and the publication and other services needed to support the system — with a view to securing assessment and testing arrangements which are simple to administer, understandable by all in and outside the education service, cost-effective, and supportive of learning in schools.

To do all this they were given around five months.

The aim of the assessment arrangements proposed by TGAT was to provide an assessment system to meet five distinct purposes. These were that it should be:

- formative — providing information on where a pupil is, enabling teachers to plan the next stages

- summative — providing overall information on the achievement of pupils

- evaluative — providing aggregated information on classes and schools to assess curriculum issues, as well as the functioning of teachers and schools

- informative — providing information to parents about their own children and general information about the school

- for professional development — giving teachers greater sophistication in assessment, recording and monitoring so that they can evaluate their own work.

To meet these objectives, TGAT proposed an innovative assessment system (GB. DES and WO, 1988). Attainment was to be measured on a continuous scale of ten levels, covering the entire five to 16 age range. This ten-level scale was to be criterion-referenced with each level defined by a set of criteria.

Pupils were to attain a level by demonstrating the performance set out in these criteria. This differed fundamentally from the more traditional norm-referenced test model derived from the selective model where pupils are ranked in order in comparison with others of their age.

The full TGAT scheme had many elements and ramifications. It was hailed as a success because it appeared to meet both the government's desire for a system based on accountability and the teachers' desire for a system based on professional judgement and diagnostic assessment. Ultimately, the detailed implementation proved that this synthesis was impossible to achieve. However, one important part of the TGAT proposals was that teachers should assess their pupils against the same criteria as the National Curriculum tests. There should then be a moderation process in which the two sets of results were examined together and differences resolved. Teacher assessment against the National Curriculum criteria has remained part of the assessment system, but its purpose in the system has not always been clearly apparent and has varied from time to time.

The TGAT model as implemented was such that for each level of each attainment target there were between one and ten statements of attainment which defined it and which required assessment in some way. There were also many attainment targets for each subject. As a consequence, across a subject, there were a large number of statements. The attainment targets in each subject and the statements of attainment defining each level were determined by separate working groups of experts in each subject. Thus the criteria did not represent empirical findings about pupils' attainments, but rather the judgements of the working groups comprising experts and interested parties. They were targets in the sense of what ought to be achieved rather than being defined in any normative sense.

In 1989, when the National Curriculum was beginning to be introduced, the Foundation was again heavily involved. The NFER was one of the agencies commissioned to produce standard assessment tasks for seven-year-olds (known at the time as SATs),[2] following the blueprints given in the TGAT report. These SATs were large-scale, cross-curricular tasks and broke new ground in assessment as the development team attempted to meet several different imperatives: for validity and classroom authenticity; for reliability and rigour; for wide coverage; for detailed diagnostic information; and for manageability. Ultimately, all these needs could not be satisfied in a single instrument or system and in later years there was a return to more standard forms of tests. However, some of the innovative aspects, for example the approach to reading, remain a firm feature of the key stage 1 assessment

regime. In addition to developing the key stage 1 assessments, the NFER undertook important evaluations of them each year, helping to provide a means forward in resolving the tensions of the new system (Whetton *et al.*, 1992; Sainsbury *et al.*, 1992).

The assessment of children at the end of key stage 1, that is seven-year-olds, was the first part of the assessment regime to be put into place. This was largely an accident since this was the key stage which was shortest and therefore required end of key stage assessments first. In many ways this was unfortunate since the assessment of children at this age was the most unknown area and also the most problematic. Established assessments of children at this age used individual administration or were relatively informal, but for the first time, these were having to be applied to a compulsory universal testing regime. The SATs initially involved three subjects (mathematics, English and science), covering a sample of nine attainment targets leading to over 100 statements of attainment which had to be assessed.

The history of the development illustrates well the tensions between: authenticity in the assessment as representing the curriculum (and how children respond to it); the requirements of accountability and reliability; and manageability in terms both of teachers' time and children's time (Sainsbury, 1996).

The initial attempt to meet these tensions was the 1991 SATs. These tasks were designed to reflect quality work in infant classrooms and featured assessments of investigations taking place with small groups of children. Reading was assessed through an individual interview in which a passage from one of 27 specified books had to be used. For the assessment of the attainment targets, each statement of attainment was interpreted in the context of the specific activity, giving a basis for teachers to make their judgements of children. In order to reduce the number of assessments needing to be made, a crude form of tailored testing operated. When the first assessment took place in primary schools across the country, there was a great outcry about the assessment load. The work involved class teachers for too long — an average of around 44 hours (Whetton *et al.*, 1992) whereas the specification had been for 30 hours of teacher time. Children's education was disrupted for far too long. The reliability of the SATs was criticised since many of the tasks involved individualistic interactions with children, and because teachers found it difficult to assess children working collaboratively. The SATs had attempted to meet some of the principles of Wood's educational assessment by providing a means whereby all children could show their best. However, in reacting to this, value judgements of fairness by the differing parties then

intruded. The underlying principle of fairness adopted was one in which children were treated differently, for example, they were allowed to read a book which interested them in order to show their best. This can be contrasted with a view of fairness which believes that this must be achieved through everyone being treated the same. In the reading assessment, this view was articulated most closely by educational psychologists (e.g. Pumfrey and Elliott, 1991), who were extremely critical of the approach to testing — contrasting it with (to them, an obviously better) psychometric approach in which the same set of single words are read aloud or written sentences completed.

However, to balance these negatives, there were positive outcomes which were largely associated with the effects on the curriculum and on teachers' knowledge and skills in assessment. For the first time ever, scientific and mathematical investigations were taking place in every infant classroom and teachers were involved in making assessments of children's work in these areas. LEA advisers claimed that they had striven for years to achieve this and had not succeeded. Compulsory authentic assessment had an immediate effect. But for many, the gains in beneficial effects on the curriculum were outweighed by the manageability and reliability arguments.

As an unforeseen side-effect, the SAT had also challenged the rhetoric of what was called good primary practice. This was the view that teachers could manage a classroomful of children working in small groups so that all were actively engaged in meaningful and productive work. The SAT experience showed that some teachers could do this and these were often the ones who had little difficulty with the assessment process, but for many other teachers the process put such a strain on their working practices that the methods broke down. Inadvertently, it was perhaps these experiences which began the process of questioning the realities of primary classroom teaching practice. They also indicate the difficulty of embedding authentic assessments in classroom practice when the practice itself is variable and differentially successful.

Over the five years since 1991, the balance between validity, manageability and reliability has gradually (and at times rapidly) swung toward manageability and reliability. This became inevitable once the teachers' unions, which organised a boycott of the assessments in 1993, had decided that the paramount issue was manageability. The use of the results for accountability purposes — the publication of school results — also pushed in the direction of reliability through standardisation. The advocates of curriculum authenticity could not stand against these arguments. The trend has therefore been toward a reduction in the number of subjects tested (science is no longer included) and in the

aspects involved (the process attainment targets are excluded). However, some authentic aspects remain. Despite persistent attempts since 1992 to change the reading assessment to a written test, the individual interview using an actual book remains. It is perhaps the one example where the arguments of curriculum authenticity have stood against manageability. They have done this because the teachers themselves are prepared to undertake the individual assessments, which fit with their views of the reading process for young children.

In response to widespread criticism, the curriculum was reviewed and reduced in size and structure in 1993. The attainment targets themselves have altered in structure from the statements of attainment to a more broadly based hierarchy of level descriptions which are intended for the teacher to find a best fit to the child's attainments. It is a model which bases assessment on direct global judgement rather than on separate elements which must then be aggregated. As such it is suited to teacher assessment, but it remains the case that test developers have generally had to explicitly redefine elements in order to arrive at questions for the National Curriculum tests.

An interesting question, and one to which I do not know the full answer, is why the US experience of criterion referencing was not heeded at the outset of the National Curriculum. Some aspects were, in that the dangers of minimum competency tests were avoided with the progressive level-based system. However, the problems of large numbers of criterion-referenced statements were ignored. The US gurus had discovered the dangers of over-prescription (Popham, 1992). However, this did not seem to influence either TGAT or the subject working groups. By the time test developers came on the scene, the statements of attainment were established facts and the direction given to agencies was that these must be individually assessed. It was asserted that this was the view of government lawyers — the statements were the defining instrument and must be directly addressed. (As an aside, when circumstances altered with a teachers' boycott, the advice of government lawyers changed and tests assessing attainment targets as a whole were suddenly legally permissible.)

It is a feature of the English system of the development of a national curriculum and assessment system that the process has been essentially private (Ruddock, 1995). By this I mean that the structure and detail are generally decided by officers of national bodies or by appointed committees who must, of necessity, utilise underlying constructs and principles. There are then consultation processes. However, the constructs embedded in the curriculum and assessment structures and their intellectual justification and precursors are

seldom made explicit but must be inferred from the resulting document.

This is as true of the TGAT report, which seems to have been heavily influenced by the graded test movement (Pennycuick and Murphy, 1988), as of the latest revision of the curriculum, which seems to have adopted the model of achievement standards as fuzzy concepts, originated by Sadler (1987). (See Sizmur and Sainsbury (1997) for an elaboration of this.)

It is important to mention teacher assessment here since this still forms, in theory, a large part of the assessment system. The original TGAT model was that the SATs should moderate teacher assessment, making it comparable between schools. However, this was never a real possibility given the nature of the attainment targets as produced. After a muddled intermediate status, teacher assessment is now to be regarded as having an equal but separate status to the test results. The outcomes of teacher assessment are to be published alongside those from National Curriculum tests. Advocates of the primacy of teacher assessment have seen it as being both more valid and more reliable than the tests. However, most evaluations show that there is a long way to go before this hope can be justified (e.g. Sizmur *et al.*, 1994).

The interaction of the testing regime and its effects on curriculum and assessment practice is obviously a complex one. In many ways the early years of the National Curriculum assessment must be seen as a failure. The aspirations of manageable authentic assessment were not met and there have been constant changes both to the curriculum and its assessment. We are now in the midst of a five-year moratorium on changes in the curriculum. I have two comments to offer here. First a statement from the government in 1989, which has largely been forgotten.

> Experience of the National Curriculum and its component parts in operation will inevitably require changes to be made — to attainment targets, programmes of study and assessment arrangements (GB. DES, 1989).

It was always to be expected that there would be changes in the assessment arrangements, but perhaps the inevitability of changes of direction as a part of the management of change should have been more overt.

Second, a reflective view at some distance based on evaluations of these early assessments.

> To conclude, we believe we have offered evidence that, as a result of the national curriculum and assessment programme, teachers have

redirected the focus of their teaching and this has been reflected in improved national assessment results in the 'basic skills'. Greater care in planning, close observation of children and a more detailed understanding of individual progress impacting on teaching were reported by over half our headteachers, as well as a lasting effect on collaboration and discussion by a smaller number. Many of our Key Stage 1 teachers have moved away from intuitive approaches to assessment towards more systematic, evidence-based techniques. The SATs have acted as a training device, and group moderation has broken down barriers. There is a clear picture of enhanced understanding and practice in assessment for the Y2 teachers, and heads and LEAs are putting these skills to good use further up the system. All this has been achieved, however, at a cost to teachers' lives and ways of working. Finally, we believe our evidence shows that the improvements in practice, both in teaching and assessing, would not have resulted from the introduction of traditional standardized tests alone, but depended on a wider approach with moderated teacher assessment at its core (Gipps *et al.*, 1995).

More recently, Gipps has also made the following bullish pronouncement about assessment in England:

We have some of the most sophisticated assessment practice in the world: this assessment is capable of supporting learning, it operates across a broad curriculum, and can be sufficiently reliable for comparability purposes; what is more it is supported by teachers who are probably the best teacher-assessors, markers and examiners in the world (Gipps, 1996).

If this is true of teachers of young children then, given their levels of understanding before 1989, the SATs had a great deal to do with this change. Although they are gone, they may have left a positive legacy, but we have yet to see whether this is a legacy that is frittered away.

What then can be discerned about the changes over the last 50 years? There have certainly been improvements in the technical adequacy of tests in a psychometric sense. The statistical underpinnings in terms of examination of reliability, item characteristics, and item bias have developed out of all recognition and such matters are now routinely examined during test development. The choice of items to make up tests has become much more efficient. The range of item types is greater and the tests themselves much fairer in that considerable pains are taken to make the tests accessible and

unthreatening. The range of methods of presenting questions and the modes of response are much broader. All of this has taken place within the psychometric tradition.

Alongside this, the new assessment tradition has developed. Advocacy of authentic assessment and performance assessment is widespread. The goal of embedding assessment into the curriculum is accepted as legitimate. The forms of assessment are much more varied. Does this constitute a paradigm change, as suggested by Gipps (1994)? It is perhaps the nature of a paradigm shift that those being shifted do not recognise what is happening to them. Indeed, they are supposed to die off, leaving the new paradigm to the next generation. However, as a practitioner working in test development it did not feel like a shift, more a succession of different imperatives interacting and informing each other. As you have seen, that succession is the way I view the changes, but I do accept that we have witnessed a broadening in the nature of testing methods and the uses of testing from a psychometric model to a situation with more overt attempts at the integration of assessment practice and the curriculum. We may not have a new paradigm but we do have a growing understanding of the whole assessment enterprise and its diversity.

What we should also infer is that some of the imperatives of the psychometric age remain with us in our current concerns. In particular, where there is an emphasis on accountability, reliability in the sense of repeatable, standard measures remains important. Where there is an emphasis on embedding assessment into the curriculum, it cannot be at the expense of reducing the time given to the teaching process. Utility and manageability remain important issues even when we strive for the assessment aspiration.

Nothing, however, especially in assessment, stands still. The two major current concerns now in England are what are termed 'value added' and 'baseline'. These are related in that they derive from notions of progression. Assessment and its analyses are moving on from being static measures at one time point to considerations of change in pupils and institutions.

The National Curriculum structure enshrined the notion of progression — building it into its definitions of levels of attainment. However, it has also become much more prevalent as an evaluation question. The original proposals for publication of results of schools — the so-called league tables — took no account of background information on the pupils or schools. This could include social and environmental factors, but a major determinant is obviously the previous attainments of the pupils. Measures of what has become known as 'value added' are now frequently made (GB. DFE, 1995).

117

There is no agreed definition of what this term, borrowed from economics, actually implies in an educational context. In some studies, the initial measures are effectively pupil background variables: gender, social class measures, length of schooling, type of school, etc. Such analyses are equalising schools in terms of their intakes in order to gauge their differential effectiveness. In other studies, previous attainment is used as the input variable, and these, perhaps, can be more justifiably termed value added. Both these types of studies can most usefully be seen as contributing to school improvement. To do so, they must somehow be differential and diagnostic, pointing schools in specific directions perhaps in relation to particular subjects or teachers.

There do, though, remain problems with the value added approach. There is resistance from some to the complexity of the full and proper multilevel analyses. However, it is not necessarily the case that the full statistical model needs to be understood, rather that there should be an understanding of the results, including their dependability, so that they can be used effectively. More troubling are the findings that such school effectiveness measures are not consistent over time, differing from year to year (Nuttall *et al.*, 1989). This might suggest that the underlying processes are too variable to actually contribute to school improvement. Although there are a number of projects and services in this area (for example, QUASE at NFER, YELLIS at Durham University, etc.), the translation of the numerical enterprise into a researched demonstration of an impact on school effectiveness remains to be accomplished. Taking a parallel with the eras of assessment I have been describing, the studies are still at the normative stage. A school's standing can only be in relation to the remainder in the analysis. Value added studies cannot detect school improvement where it occurs throughout the sample or population.

A consequence of the concentration on value added measures has been the movement of the input measures lower and lower down the age range. For secondary schools, the input measures are of children aged 11, and this can be done fairly reasonably through formal written tests. To gauge the effectiveness of junior schools, the input measures are of seven-year-olds. As we have seen, this enhances the need for these assessments to be more standard and more reliable — possibly at the risk of their validity and suitability for young children. There are currently moves to extend this value added ladder one rung further down, so that the effectiveness of key stage 1 schools may be judged. This has been given the shorthand term 'baseline assessment'. This means that children arriving in school who are less than five years old must be assessed. As before, to be useful, these assessments

will have to tend toward a standard and reliable format. This is not an easily managed task for a large number of very young children, particularly if the information is to be diagnostically useful, which is the intention stated to teachers. A similar conflict to that over the assessment of seven-year-olds may well be in prospect.

As I have said, my view of the past 50 years is perhaps one of movement from simplicity and single purposes for tests to one of complexity and multiple purposes for assessment. How then could assessment change in the future?

First, I do not think that this is a question which is capable of an answer without guessing at the new imperatives for education in the next century. Just as the selective era determined one form of assessment, the monitoring era another and the accountability era several, the priorities and structures of education in the next century will determine the next era. Just as the past models were layers on top of each other, we will have further overlays but with many of the same structures and imperatives operating underneath.

In public, rather than expert, debates about the needs of society, two sets of values are often espoused. The first is the basic curriculum notion. Advocates believe that primary-age children should be taught a very limited set of skills in reading, writing and numeracy — defined narrowly as arithmetic skills. In this model there is no necessity for, and even a prohibition of, the use of modern technology. Calculators are thought to detract from numeracy and the use of computers for writing is anathema. In this view, a solid grounding in the basics is then sufficient to embark on the specialisms of a secondary education. The assessment structures associated with this are seen to be traditional ones. Pupils are assessed as individuals, using paper-based techniques, and the emphasis is on reliability.

The second set of values are those often advocated by employers. These emphasise social and personal skills such as team working, communication skills, punctuality and appearance. The assessment of these implies much more open tasks and procedures. The issue of the assessment of an individual within the dynamics of a group situation has not been conclusively researched. Some of the APU work addressed this, but only for explanatory purposes. If such processes are to be the valued outcomes of education, then the assessments must address these appropriately.

This then leads to the whole area of computers and information technology in the curriculum and hence in assessment. There will soon be no economic reasons why computers should not be available to every pupil in every lesson.

These would not be the most up to date model — more likely the technology of ten years before — but even so they would have considerable capabilities. However, the limiting factor is likely to be the value judgements placed on their use as part of teaching and learning.

An area awaiting considerable innovation is the use of computers and, more generally, an exploitation of their multimedia and communication capabilities. The major use of computers in educational testing thus far has been, first, in administrative support and, second, in the simple translation of paper-based tasks to the screen. Some such translations are impressive, for example in adaptive testing.

Adaptive testing involves holding a large number of questions in the computer so that each candidate receives a different selection of them. The questions are chosen using algorithms which match their difficulty to the person's measured ability. The scores are all placed on to a common scale using statistical techniques, usually Item Response Theory. This involves savings in time since each person has a shorter test, tailored to their own ability levels. For the same reason, the accuracy of the measurements can be increased.

Thus in computerised testing, the emphasis has been on increasing reliability and manageability. The computer has not yet come into its own in terms of validity. In order for it do so, it will have to be the case that its potential must be accepted as appropriate and regarded just as any other tool. If this were so, then the full capabilities could begin to be exploited. Pupils could enter essays using the keyboard and edit these themselves within the testing period. The tests themselves could be transmitted to markers by e-mail, speeding up this process. (This is already happening in the USA.) There is the prospect of software marking essays either partially or wholly. However, even these possibilities are simply the translation of a paper-based format into a computer. Beyond this, information and situations to set the context of the assessment could be presented in a variety of ways, through software on spreadsheets, and databases within the computer, or through video or multimedia simulations. The use of such information could be interactive with a new form of tailored testing in which situations unfold according to the responses made during the course of the test. At present such tests have been developed for a variety of psychometric purposes, from early diagnoses of Alzheimer's disease to the selection of trainees to be aeroplane pilots. In the educational sphere, there have been demonstrations of the use of video rather than text as a stimulus to questions. However, such developments often founder because of the need to retain consistency with the past and to offer only the lowest common denominator because of issues of access by pupils.

The next steps would be to use the modern communications capacity of the computer. If the skills we value are the location of information, its synthesis, and its effective communication, what better test than to allow access to the Internet and use of it during the test? These types of possibility pose traditional assessment problems. First, the marking scheme for such tasks may be difficult to devise and even more difficult to scale fairly since every run through of a task may be different. In such circumstances, neither the assumptions of classical test theory or Item Response Theory theory are met. A whole new statistical basis may be required — or the acceptance that none is possible. However, much more fundamentally, such assessments can only have credence if they are valid representations of the educational processes occurring at the time. All of what I have described is technically possible but there is no valid point in developing such instruments if they do not match the ethos of the curriculum and educational practice. If the curriculum encompasses such possibilities and the public and politicians accept them as a proper activity in education, then assessment practices can utilise them. The alternative is that technology will be rejected as inappropriate or damaging, in which case the place of computers in educational assessment will remain the faster operation of previous, outmoded technologies that do not occur anywhere in society except in schools.

At the other end of the scale from computerised testing are the individual judgements of teacher assessment. The recent past has seen the political advance of teacher assessment. In Britain, the notion of teacher assessment integrated into the curriculum for teaching purposes has become a commonplace. However, teachers' understanding of the process and their ability to manage the process have not matched this. Essentially what teachers are doing is acting as an expert system involved in decision making. To do this, they are utilising their experience as well as, to some extent, the structures of the National Curriculum. Underpinning this is a model of learning which is both sequential and behaviouristic on the one hand and constructivist on the other.

But these models are implicit in what teachers are doing rather than in any formal understanding of the processes. They do not have an understanding of the meaning and place of reliability and validity in what they are doing. But, neither has the assessment profession given them that understanding, or any tools to include it in their teaching processes.

The lesson which I draw from both the future use of technology and the future of teacher assessment is that both must move away from the explicit atomised view of educational processes, towards something more akin to

expert systems. They must involve tasks which are different for each individual, both because they have differing starting points reflecting the individual's interests and because they allow different paths to be followed during the task. The tasks themselves must be open-ended but be reflections of what is valued in education — and valued not just by researchers, or by teachers, but by parents, employers and politicians also.

The National Curriculum experience illustrates the difficult dilemma between producing assessments which are designed to lead change and implementing those which reflect the majority of current practice. The former is more painful but can ultimately be the most beneficial. But it can only work when there is a general consensus, or at least a majority in favour, and that consensus must extend over the different interested parties.

There is little new in any of this. It parallels the views of Nuttall and Wood on the primacy of validity and the need for educational assessment, rather than psychometrics. However, I think what the last decade has shown is that consequential validity in the Messick sense has been regarded too narrowly. The assumption was that assessment researchers could define the form of assessment because they believed its consequences to be educationally desirable. However, other education professionals, parents, the press and politicians may resist this. It seems to me that in fact the values of a wider society must be considered. The real aspiration for assessment should be that a more general consensus on its worth can be achieved. But, regardless of this, just as in each previous age of assessment the testing and assessment system has paralleled the educational structures and imperatives, so the next age of assessment can only be embedded in the education system the various interested parties will allow to happen.

---

[1] This part of the discussion is an expanded version of the ideas in one section of Hegarty (1996).

[2] The TGAT report (GB. DES and WO, 1988) originated the term 'standard assessment tasks', which was soon abbreviated to 'SATs'. From 1992 onwards, this abbreviation was no longer used in official documents. According to Daugherty (1995), this was because of ministerial dislike of educational jargon and because the agency responsible for the Scholastic Aptitude Test in the USA threatened legal action over the copyright of the acronym 'SAT'. It remains in current use by teachers and in many articles and other reports. The official term is now 'standard task' or, increasingly, '(standard) tasks and tests'. This article will make use of the term 'SATs', however, when speaking in an historical context, as that is how published sources referred to them.

References

BAKER, K. (1993). *The Turbulent Years*. London: Faber & Faber.

BARBER, M. (1996). *The National Curriculum: a Study in Policy.* Keele: Keele University Press.

CHOPPIN, B. (1968). 'An item bank using sample-free calibration', *Nature*, **219**, 870-2.

CRONBACH, L.J. (1970). *Essentials of Psychological Testing*. Third edn. New York: Harper & Row.

CRONBACH, L.J. (1971). 'Test validation.' In: THORNDIKE, R.L. (Ed) *Educational Measurement*. Second edn. Washington, DC: American Council on Education.

CRONBACH, L.J. (1984). *Essentials of Psychological Testing*. Fourth edn. New York: Harper & Row.

DAUGHERTY, R. (1995). *National Curriculum Assessment: a Review of Policy 1987-1994*. London: Falmer Press.

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

FOXMAN, D., RUDDOCK, G., JOFFE, L., MASON, K., MITCHELL, P. and SEXTON, B. (1985). *A Review of Monitoring in Mathematics 1978–82: Parts 1 and 2*. London: DES, APU.

GIPPS, C. (1996). *Assessment for the Millennium: Form, Function and Feedback* (based on an Inaugural Lecture delivered at the University of London Institute of Education, on 6 June 1996). London: University of London Institute of Education.

GIPPS, C.V. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London: Falmer Press.

GIPPS, C., BROWN, M., McCALLUM, B. and McALISTER, S. (1995). *Intuition or Evidence? Teachers and National Assessment of Seven-year-olds*. Buckingham: Open University Press.

GIPPS, C. and GOLDSTEIN, H. (1983). *Monitoring Children: an Evaluation of the Assessment of Performance Unit*. London: Heinemann.

GIPPS, C., STEADMAN, S., STIERER, B. and BLACKSTONE, T. (1983). *Testing Children: Standardised Testing in Schools and LEAs*. London: Heinemann.

GLASER, R. (1963). 'Instructional technology and the measurement of learning outcomes: some questions', *American Psychologist*, **18**, 4, 519-21.

GOLDSTEIN, H. (1979). 'Consequences of using the Rasch model for educational assessment', *British Educational Research Journal*, **5**, 2, 211-20.

GOLDSTEIN, H. (1980). 'Dimensionality, bias, independence and measurement scale problems in latent trait test score models', *British Journal of Mathematical and Statistical Psychology*, **33**, 2, 234-46.

GOLDSTEIN, H. (1983). 'Measuring changes in education attainment over time: problems and possibilities', *Journal of Educational Measurement*, **20**, 4, 369-77.

GOLDSTEIN, H. and BLINKHORN, S. (1977). 'Monitoring educational standards — an inappropriate model', *Bulletin of the British Psychological Society*, **30**, 309-11.

GOLDSTEIN, H. and JAMES, A.N. (1983). 'Efficient estimation for a multiple matrix sample design', *British Journal of Mathematical and Statistical Psychology*, **36**, 2, 167-74.

GRAHAM, D. and TYTLER, D. (1993). *A Lesson for Us All*. London: Routledge.

GREAT BRITAIN. DEPARTMENT FOR EDUCATION (1995). *Value Added in Education*. London: DFE.

GREAT BRITAIN. DEPARTMENT OF EDUCATION AND SCIENCE (1989). *National Curriculum: From Policy to Practice*. London: DES.

GREAT BRITAIN. DEPARTMENT OF EDUCATION AND SCIENCE and WELSH OFFICE (1988). *National Curriculum Task Group on Assessment and Testing: a Report*. London: DES.

HEGARTY, S. (1996). 'A half-century of educational research', *Educational Research*, **38**, 3, 243-57.

HUSEN, T. and POSTLETHWAITE, T.N. (1996). 'A brief history of the International Association for the Evaluation of Educational Achievement (IEA)', *Assessment in Education*, **3**, 2, 129-41.

JENSEN, A.R. (1980). 'External sources of bias.' In: JENSEN, A.R. *Bias in Mental Testing*. London: Methuen.

LAWLEY, D.N. (1950). 'A method of standardizing group tests', *British Journal of Psychology* (Statistical Section), **3**, 86-9.

LAWTON, D. (1996). *Beyond the National Curriculum: Teacher Professionalism & Empowerment*. London: Hodder & Stoughton.

LORD, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

MADAUS, G. (1996). 'International patterns of assessment: policy change.' Symposium at Annual Meeting of the American Educational Research Association, New York, 12 April.

MESSICK, S. (1980). 'Test validity and the ethics of assessment', *American Psychologist*, **35**, 11, 1012-27.

MESSICK, S. (1989). 'Validity.' In: LINN, R.L. (Ed) *Educational Measurement.* Third edn. New York, NY: American Council on Education and Macmillan Publishing Co.

NUTTALL, D.L. (1986). 'Problems in the measurement of change.' In: NUTTALL, D.L. (Ed) *Assessing Educational Achievement.* London: Falmer Press.

NUTTALL, D.L. (1987). 'The validity of assessments', *European Journal of the Psychology of Education*, **II**, 2, 109-18.

NUTTALL, D.L., GOLDSTEIN, H., PROSSER, R. and RASBASH, J. (1989). 'Differential school effectiveness', *International Journal of Educational Research*, **13**, 7, 769-76.

PENNYCUICK, D. and MURPHY, R. (1988). *The Impact of Graded Tests.* London: Falmer Press.

POPHAM, W.J. (1992). 'A tale of two test-specification strategies', *Educational Measurement: Issue and Practice*, **11**, 2, 16-17, 22.

PUMFREY, P. and ELLIOTT, C. (1991). 'A house of cards?' *Times Educ. Suppl.*, **3905**, 3 May, 12.

RESNICK, L.B. and RESNICK, D.P. (1992). 'Assessing the thinking curriculum: new tools for educational reform.' In: GIFFORD, B. and O'CONNOR, M. (Eds) *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction.* London: Kluwer Academic Publishers.

RUDDOCK, G. (1995). 'England and Wales.' In: ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT *Performance Standards in Education: In Search of Quality.* Paris: OECD.

RUDDOCK, G. and TOMLINS, B. (1993). *Evaluation of National Curriculum Assessment at Key Stage 3: Report on the 1992 National Pilot Assessment of Mathematics and Science.* London: SEAC.

SADLER, D. (1987). 'Specifying and promulgating achievement standards', *Oxford Review of Education*, **13**, 2, 191-209.

SAINSBURY, M. (Ed) (1996). *SATs the Inside Story: the Development of the First National Assessments for Seven-year olds, 1989–1995.* Slough: NFER.

SAINSBURY, M., WHETTON, C., ASHBY, J., SCHAGEN, I. and SIZMUR, S. (1992). *National Curriculum Assessment at Key Stage 1 in the Core Subjects 1992 Evaluation. A Report by the NFER/BGC Consortium.* London: SEAC.

SCHAGEN, I.P. (1990). 'A method for the age standardization of test scores', *Applied Psychological Measurement*, **14**, 4, 387-93.

SHORROCKS, D., DANIELS, S., FROBISHER, L., NELSON, N., WATERSON, A. and BELL, J. (1992). *The Evaluation of National Curriculum Assessment at Key Stage 1 (ENCA 1 Project): Final Report.* London: SEAC.

SIMON, B. (1996). 'IQ redivivus, or the return of selection', *Forum*, **38**, 1, 4-7.

SIZMUR, S. and SAINSBURY, M. (1997). 'Criterion referencing and the meaning of National Curriculum assessment', *British Journal of Educational Studies*, **45**, 2.

SIZMUR, S., SAINSBURY, M., ASHBY, J. and HARGREAVES, E. with JONES, E. (1994). *Teacher Assessment 1994 Key Stage 1 Evaluation (Mathematics and Science). Final Report.* London: SCAA.

START, K.B. and WELLS, B.K. (1972). *The Trend of Reading Standards.* Slough: NFER.

SUTHERLAND, G. (1984). *Ability, Merit and Measurement: Mental Testing and English Education 1880–1940.* Oxford: Oxford University Press.

VERNON, P.E. (1960). *Intelligence and Attainment Tests.* London: University of London Press.

WATTS, A.F., PIDGEON, D.A. and YATES, A. (1952). *Secondary School Entrance Examinations.* London: Newnes.

WHETTON, C., SAINSBURY, M., HOPKINS, S., ASHBY, J., CHRISTOPHERS, U., CLARKE, J., HEATH, M., JONES, G., MASON, K., PUNCHER, J., SCHAGEN, I. and WILSON, J. (1992). *National Curriculum Assessment at Key Stage 1: 1991 Evaluation. A Report on the Working of the Standard Assessment Task by the NFER/BGC Consortium.* London: SEAC.

WOOD, R. (1986). 'The agenda for educational measurement.' In: NUTTALL, D.L. *Assessing Educational Achievement.* London: Falmer Press.

WRIGHT, B.D. and STONE, M.H. (1979). *Best Test Design.* Chicago, IL: MESA Press.

YATES, A. (1953). 'Symposium on the effects of coaching and practice in intelligence tests: an analysis of some recent investigations', *British Journal of Educational Psychology*, **23**, 3, 147-54.

YATES, A. and PIDGEON, D.A. (1957). *Admission to Grammar Schools.* London: Newnes.

Respondents' comments

Albert E. Beaton, Boston College

It is an honour for me to have this opportunity to participate in the Jubilee honouring the fiftieth anniversary of the National Foundation for Educational Research. Of course, I have long been aware of important work that NFER has done, but I had not realised what a huge contribution it has made to the field of educational testing and measurement until I read the paper of Chris Whetton. Chris amassed and integrated NFER's many accomplishments in a thoughtful fashion. NFER has a right to be proud of its past and should be confident of its future.

Chris Whetton's paper is very well written and interesting. Reading it brought back many memories and informed me about many things that I did not know. For example, I did not know that two superstars in my graduate school days, Godfrey Thompson and Cyril Burt, were at one time Vice Presidents of the NFER. However, I will not repeat here what Whetton has written so well; instead, I will discuss some of the implications of recent NFER work.

I was especially interested in the recent NFER work on the standard assessment tasks, also known as the SATs. I was introduced to the SATs through a videotape that was shown at the Educational Testing Service, where I was employed at the time. The SATs were a very ambitious project that seemed to break many boundaries that had inhibited educational testing and promised much more powerful uses and interpretations of test data. I must admit that I also wondered what sorts of data analyses and inferences would support the new tests.

Over the next years, very mixed messages about the SATs crossed the Atlantic. At first, we heard that the teachers loved the tests, then the teachers' unions had boycotted their administration. Apparently, there are now a number of reports that evaluate the results according to different perspectives. Whetton's paper lists at least eight such reports. I will have to read them all to form my own opinion; apparently, there is not, and never will be, a single agreed-upon consensus on the value of the SATs.

It seems to me that the SATs did have some important effects, but too much was expected of them. Policy makers, educational administrators, teachers and parents ask too much from a single test, and perhaps we professionals promise too much. I am reminded of men's stores in the United States that sell stretch socks, proclaiming that one size fits all. Perhaps, but those socks fit many men very poorly, especially those with big feet, like mine. Tests that try to do too much will be unable to do everything well.

Tests may be used for a number of purposes: diagnosis, assignment to track or group, feedback to teachers, accountability, selection, promotion and so forth. Tests that are designed for one of these purposes may be useless for others. Let us look at just one factor — the time of year when the students are tested — although many other factors are important in the design of a testing system. Assignment to track or group is done at the beginning of the school year; diagnosis is done throughout the entire year with quick feedback necessary; selection and promotion are done at the end of the year. Accountability is best done at the end of the year, but the 'value added' approach is done best with highly reliable tests at the beginning and the end of the year.

Adding the requirement that our tests be fast, cheap and good also presents a problem. I have concluded that we can have only two of these three desires. Without a technological breakthrough, a testing system can be fast and cheap, but will not be good; it can be fast and good, but will not be cheap; and it can be cheap and good, but will not be fast. And it takes some skill to develop systems that have at least two of these desires. It should be clear that, for these and other reasons, a single test cannot satisfy all of these purposes. We need to be clear about what a test can be expected to do and what it cannot do.

With the advent of authentic and alternative assessment, there is a blurring between pedagogy and measurement and between teaching and assessment. Assessment has always been an important part of teaching, certainly since Socrates. Some modern assessment tools are quite old. In secondary school, a teacher of mine had students' names written on file cards which he sorted before asking individual students questions. Little did this teacher know that he had invented BIB (Balanced Incomplete Block) spiralling. This teacher used this method both to evaluate individual students' knowledge and to decide what should be taught next.

Whetton makes clear that the SAT experience did improve pedagogy and also helped introduce a new national curriculum quickly and efficiently. These are clearly good pedagogical goals, but are they measurement?

The discussion of validity and reliability by Wood and Nuttall, cited by Whetton, bothered me. Wood and Nuttall seem to treat reliability and validity as if they were bipolar, as if having one meant being sparse on the other. The rhetoric of the authentic testing movement argues that testing tasks be 'real' tasks and that the tasks' purposes be transparently clear. To me, this is face validity, i.e. the tests appear to measure what they purport to measure. Face

128

validity is clearly a desirable feature of a test, but it is generally agreed that face validity is not validity at all. Clearly, it is not consequential validity in the sense of Cronbach and Messick.

It is a basic concept of educational measurement that a test cannot be valid if it is not reliable; this is not an either/or situation. In fact, the reliability coefficient is the maximum value that any validity coefficient can attain. A test cannot predict, explain or relate to something else if it does not measure its own construct consistently. To produce valid inferences from tests, the tests must be reliable. The challenge to educational measurement is to produce tests that are both reliable and valid for their intended purposes.

The new types of tests that are being proposed and developed may need new psychometric methods for validation. Perhaps our present methods of reliability and validity coefficients are inadequate. New psychometric methods may arise as the new test methods and purposes are more clearly articulated. The basics of measurement will still be the same: the measure of a unit, whether it is a student, classroom or school district, must be reasonably stable and be clearly related to the uses for which the measure is intended. A new psychometrics will build on the old, not throw it out.

As the NFER and other testing organisations move ahead to meet new testing challenges, I am sure that the technology of testing will change. More emphasis has to be placed on the purposes of a testing system and then technology must be tailored to those purposes. An example of this is the National Assessment of Educational Progress in the United States and the Third International Mathematics and Science Study, which achieve the target of broad subject-matter coverage by ignoring unnecessary individual student scores. Surely, the future will involve more advanced mathematical models such as item response models and generalisability theory, as Whetton notes, and these models will evolve as the aims of particular tests become explicit. The evolution of computer-based testing will also extend the opportunities for testing while presenting challenges to our analysis and interpretation of its results.

The history of the NFER over the last 50 years has been outstanding. As well as being a particularly creative organisation, it has also amassed a strong technical staff capable of tackling the challenges that are just beyond the horizon. It seems to me that the future of the NFER is strong and it will continue to be a major contributor to educational measurement throughout the world.

Caroline Gipps, University of London Institute of Education

I am delighted to have been invited to respond to Chris Whetton's paper. I began my career as a researcher at the NFER in 1970 as a research assistant on a test development project. For the last 16 years I have worked at the London Institute of Education on issues around test use and impact, and teachers' practice in assessment.

First, let me commend the paper for its scope and clarity: it is a thorough and well-crafted account of the development of testing in England and Wales in which the NFER has played such a significant part. I shall divide my comments into three parts: the past, the present and the future.

The past

Until I read Chris Whetton's paper I had quite forgotten how seminal and forward looking was the work of Yates and Pidgeon (1957) on the 11+. They were concerned with validity by way of accuracy and consequential validity, and the design of the testing programme to avoid negative curriculum backwash. They also commented on the effects of coaching on performance, which led to the suggestion, revolutionary for the time, that one might allow regulated amounts of coaching and practice before taking such tests.

The lesson of this for me is that we should, as researchers, be spending more time reading other people's work. Not just current work across the disciplines as Susan Fuhrman (above) indicates, but also past work in our own field, what I would call the classics, for what those can teach us. Some time ago I had cause to reread Joan Barker Lunn's book (1970) on her research into streaming. It is a classic for a number of reasons including its research design: it included both quantitative and qualitative data gathering in both the natural science and interpretative paradigms within the one study — an approach that currently some researchers in the field think is novel.

Chris Whetton's paper describes in some detail the history of use, critique and development of the Rasch Model and Item Response Theory. There were many battle scars at the time, although I myself did not get engaged in either the battle or the war. It is true that many, even those with a quantitative background like myself, found the quantitative aspects of this debate impenetrable; my discussions with Harvey Goldstein, Bob Wood and Desmond Nuttall, however, led me to focus more on the logic underlying the model. Is the logic which underlies the statistical model supportable? For if it is not, then no amount of sophisticated statistical technique and refining of

the model can make it acceptable. I believe that the more impenetrable the quantitative aspect of the model, the more important it is to evaluate the underlying logic. My colleagues working in mathematics education at the Institute argue that with the growing use of computers there is a requirement for individuals to understand more rather than less mathematics, so that they can more readily evaluate or judge the worth of the complex statistical and computing models which they are able to use. What I will say about that period of test development is that the work carried out by the APU teams was indeed ahead of its time and laid the groundwork for the development of performance-type assessments in GCSE and National Curriculum assessment in this country and also for performance assessment in the USA.

The present

Chris Whetton's paper describes the development and underlying philosophy of the National Curriculum assessment programme in some detail, appropriately so since the NFER has been one of the key players in this test development and evaluation. The National Curriculum assessment programme is a major element of the current seismic shift which we are experiencing in this country in curriculum and assessment practice.

Our own research (ERSC ref R-000-2344-38; Brown, Gipps and McCallum) has found tremendous development in primary teachers' assessment skills and understanding. This must modify, somewhat, Chris Whetton's view that teachers do not understand reliability and validity: in our discussions with primary teachers, in depth and over a number of years, we found that teachers do indeed understand the issues of reliability and validity — although they may not use those terms, of course — and that in this they are very different from 15 years ago (Gipps et al., 1995). We also found teachers addressing new areas of the curriculum at key stage 1, with increased levels of collegiality, better planning and progression reported by headteachers. This, we would say, is as a result of key stage 1 teachers being involved in a fairly low stakes assessment programme with an emphasis on good quality assessment tasks and reported teacher assessment, underpinned by consensus/group moderation. We found more changes to teachers' teaching and classroom organisation at key stage 2, including increased test preparation, practice and revision. This was in the context of a high stakes programme (with league tables of schools promised in the first full year of assessment) and an emphasis on paper-and-pencil standardised tests rather than assessment tasks. Of course this is exactly what one would have predicted given the differences between key stage 1 and key stage 2 in the content of the assessment programme and the level of 'stakes' involved; it was a natural experiment in looking at the impact of two

different assessment programmes within the same school system. We also found that Year 6 primary teachers indicated that they had changed the amount of time they spent in didactic whole class teaching; they increased the amount of time spent teaching like this because of the amount of curriculum ground they had to cover.

Teacher assessment is a crucially important part of any assessment programme, I would argue. Those of us who advocate it do *not* claim that it is inherently more reliable than standardised or other forms of testing. Quite the opposite, one has to work to make teacher assessment reliable. The reason why one argues for having TA within an assessment programme is twofold: to give teachers a stake in the assessment programme and to maintain in teaching those aspects of the curriculum that cannot be tested by paper-and-pencil tests.

The future

In his paper, Chris Whetton refers to what I and others have called a paradigm shift in assessment, from psychometrics to educational assessment. Whetton's view is that the changes and developments which he describes do not represent a paradigm shift. There is, however, a real tension between the emphasis on the psychometric and measurement model, which underpins current work, and the final scenario which Whetton spells out, of open-ended, variable tasks to be used by teachers as in expert systems. The model underlying current test and assessment design will find it very difficult to cope with the design and evaluation of such new types of assessment. However, resolving this tension is the key task for assessment researchers; we have reached this point because developments in cognition and learning are telling us that assessments should map more directly on to the skills we wish to develop, while national trends for accountability demand assessments that are highly reliable. New models of assessment need new methods and philosophies for evaluating them. That is the challenge for the next few years, and I am sure that Chris Whetton and his colleagues at the NFER will be at the forefront of these developments.

References

BARKER LUNN, J. C. (1970). *Streaming in the Primary School.* Hove, Sussex: King, Thorne & Stace Ltd.

GIPPS, C., BROWN, M., McCALLUM, B. and McALISTER, S. (1995). *Intuition or Evidence? Teachers and National Assessment of Seven-year-olds.* Buckingham: Open University Press.

YATES, A. and PIDGEON, D. A. (1957). *Admission to Grammar Schools.* London: Newnes.

5.

Educational research and educational reform
Wynne Harlen, Scottish Council for Research in Education

Introduction

Identifying the links between two such complex activities as educational research and educational reform is bound to be far from straightforward. Indeed the more we look at, on the one hand, the multiple variables that have to be taken into account in drawing conclusions from educational research and, on the other hand, the plethora of influences besides research on decisions about educational change, the more tenuous the links appear to be. Yet, as Huberman put it, since we (researchers) 'were ultimately put on this earth to bring our findings to educational professionals' (Huberman, 1994), we must not see the task as hopeless; we have to pursue ways of strengthening the links, otherwise our work has no function except for researchers to inform each other. There is room for optimism that a better use of research is possible through studying the situation in all its complexity and, as I hope to show, learning from other professions which strive to increase the impact of research on practice.

In this paper I look first at the links between research and its use by practitioners and policy makers and at the constraints and influences on both researchers and 'users' of research. From this emerges support for the view that we should look across all relevant research rather than at single studies in seeking conclusions to guide decisions or to extend understanding. The second part of the paper then focuses on ways of reviewing research and synthesising findings, drawing on current work in developing guidelines for medical practice to suggest how we might more effectively use educational research in decisions about change in education.

Research and reform at different levels

Both educational research and educational reform or change exist in different forms at different levels. Educational reform exists as planned or potential action at the national or local system level, in documents, criteria, exemplification, reports, etc. It only exists, of course, as real reform, real change, in the teaching and learning at the levels of institution and classroom. Similarly, educational research can operate at the system, local, institutional and classroom levels. If we try to make all the links there are between the research at each level and change at each level, we will lose sight of the main

issues by trying to include each and every variant. So the intention here is to confine attention to the research that can influence policy makers at local and national levels and practitioners at school and class levels.

Linking researchers to users

The classical model linking researchers to users sees these as two separate and distinct elements connected as in Figure 1 where the arrow represents the communication of research knowledge through publication, conferences and seminars and more directly through teaching.

Figure 1

Several difficulties with this model are immediately apparent. One is the identification of those involved as 'researchers' and as 'users'. Users typically are thought of as teachers, administrators, policy advisers and so on, yet in many cases they are also, at times, researchers. Conversely, researchers are frequently 'users' of research, in particular when they are reviewing or synthesising research. So the model is only useful in so far as the identity of researchers and users can change between different situations. A second difficulty is the absence in the model of any way of indicating the selection of research to be communicated that fits the needs of the users. A further, related problem is the absence of any way of indicating how the users' needs can influence the research that is conducted in the first place. Finally, it neglects all the constraints on researchers and users that determine how, or whether, in any particular instance, research will be used. When changes are made there are invariably losers and gainers; there will be those who oppose change *per se,* those whose work will be, as they see it, adversely affected. Thus 'use' of research has to be seen within the context of numerous other influences on those who are, at any one time, the researchers and the users.

Figure 2 represents some improvements to the model based on these points by the addition of the arrow from users to researchers and the indication of the constraints and influences. The components of the model are now briefly discussed.

Figure 2
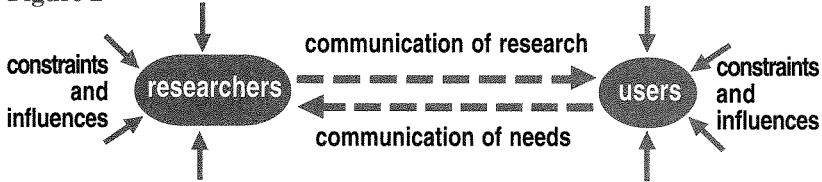
The researchers: constraints and influences

Educational research is far from being a uniform 'thing'. Each study can be located within a variety of physical, social, philosophical, political and ethical contexts which affect the way in which it gathers information and indeed the nature of the information it gathers. For example, the *philosophical* context will include the beliefs of the researchers in relation to:

- the nature of reality (varying from belief in a single reality to belief in multiple realities)

- the researchers' relationship to what is being studied (varying from belief in objectivity to belief in interactivity)

- the nature of truth (from belief in the possibility of context-free generalisations to belief that evidence is a series of individual cases)

- causality (from belief in cause–effect relationships to belief in interpretation as identification of tentative patterns and associations)

- values (from belief in a values-free inquiry process to belief that all aspects of an inquiry are influenced by values)

- modernist or post-modernist stance (from acceptance that certain relationships and structures can be taken for granted to belief that the taken-for-granted should be challenged).

The *political* context will influence the availability of funding for research on particular topics, just as it will influence the use that is made of the research that is carried out. But its influence goes further, to favouring certain methodologies over others — and therefore some researchers over others. There is a difficulty in maintaining rigorous independence in these circumstances, particularly when the research is closely supervised. The option of researchers not undertaking work which might compromise independence is always there, but in reality it is rarely chosen because of other pressures — to bring in research funding, to increase publications and

137

to maintain the employment of researchers on short-term contracts. Most of these pressures are the results of policies to judge the quality of research by indicators of quantity rather than quality.

Then there is the *ethical* context, particularly evident when sensitive information is being sought, as is necessary in certain studies of, for example, the impact of drug education on drug-taking habits. These will limit how information obtained in confidence can be reported. More generally, research raises issues about the ownership and control of data, about the obligations to those from whom information is collected as well as to those who fund the research and about the rights of researchers to publish. National research associations, such as the American, British and Scottish Educational Research Associations (AERA, BERA and SERA) have developed guidelines to help their researchers consider these matters (e.g. BERA, 1992).

These are not the only, nor the most obvious, influences. There clearly are constraints imposed by funding, by time, by researcher expertise available at the time required and by the willingness of those researched to give their time and effort for the sake of research. The matter of time limitation on the researchers deserves further attention, for it is connected with a point made below about the research agenda being set by the policy makers. The latter's need for information has to be within their ever-diminishing timescale for decision making, which is dictated more by political than by educational considerations. Researchers are frequently set impossible timetables for considering the issues, planning the research and preparing the documentation required to tender for a research contract. Maybe this seems reasonable to those funders who regard researchers as technicians, able to respond at any time to what they want them to provide, just as a motor mechanic would respond to a request to service a car. The quality of research in such cases is in jeopardy, and it is a considerable tribute to those researchers who can and do respond to these requests that the quality has been maintained and the profession of educational research not brought into disrepute. However, there is a cost: too little time for reflection, for building on existing content and methodological knowledge, for communicating (other than to funders) and for training new research staff.

What all this means, of course, is that unless a research project is unusually large and adequately funded, and contains several studies within it, it is unlikely on its own to be a firm basis for decision making about policy or practice and least of all for adding to our broader understanding of issues in education. Any one study of this kind, that is, costing less than about £150K,

needs to be considered in the light of other studies, ones which perhaps take different philosophical stances, situated in different contexts, so that any recommendations from it are not biased by the particularities of a single study. Thus the answer to the question often used to taunt education researchers, 'Tell me just one research study which has changed anything', is that this should never be the case. The matter of synthesising evidence from studies is thus a crucial one for using research and one to which I will come in the second part of this paper.

Constraints on the users

Policy makers as users

It has to be recognised that policy decisions affecting education have to be made, whether or not there is research; they do not depend on it. So research will inevitably be, if anything, only one factor in the decision making. Political and economic factors necessarily loom large. For example, the overwhelming evidence of the benefit of pre-school education has not had the effect on policy and practice that it would appear to deserve. Competing concerns on behalf of other sectors in education — ones with perhaps more political prestige — win out in the competition for restricted resources. The benefits of pre-school education are long-term, whilst policy makers have a preference for changes that can be effected on a shorter timescale. The overarching political agenda also influences all decision making; policy makers, after all, have to consider their own futures. Research evidence has to be put beside the advice of policy analysts, professional administrators and inspectors, and beside the advice of teachers' unions, professional associations, parents, representatives of industry, to name just a few. Moreover, as McGlynn has commented, in the capacity of a policy maker and user of research in Scotland:

> It is not uncommon for research to provide decision makers with apparently conflicting evidence or, as least, with evidence about different aspects of a particular educational issue which does not lead straightforwardly to an obvious policy conclusion. Fact and analysis can make a decision maker's life difficult: 'paralysis by analysis' can be a reality. As a result other considerations come into play; some are inclined to go back to intuition, personal judgement and/or anecdotal evidence... (McGlynn, 1991).

Practitioners as users

Practitioners as users of research face different dilemmas. Whereas policy makers need to use research in making decisions which have general applicability, practitioners are concerned with applying research to specific and unique situations. Their dilemmas are perhaps best conveyed through an example. In September 1996 (that is, before the publication of the TIMSS results), I was in conversation with the principal teacher of mathematics in a large Scottish secondary comprehensive school. He was concerned about the school's poor performance in mathematics in comparison with national norms, based on the number of passes at upper grades in the examination that all students take at the age of 16. His aim was, therefore, to increase the performance of the higher achieving pupils and he was planning to introduce setting throughout the school, to separate the better achieving pupils and ensure that they achieved more of their potential. (Setting means grouping children for lessons in a particular subject on the basis of their achievement in that subject.) As he asked my opinion as a researcher on this action, I told him that research showed that such actions had been found to be detrimental to the lower achieving pupils and that, overall, there was no improvement in achievement through setting, whilst there was clear evidence of damaging social effects of this practice (e.g. Boaler, in press). Clearly this posed a serious dilemma for the teacher. It was hard for him to reconcile the necessity of pleasing the 'powers that be' by increasing the performance of high achievers with the research evidence. But there were other variables to take into account. He recognised that the mathematics teachers in his schools were not presently coping well with the wide range of achievement in mixed-ability classes. In the circumstances the research evidence could do nothing to change his conviction that setting would enable him to achieve the necessary improvement in higher-level achievement. All that it could do was to alert him to the likely effects of setting on the lower achievers.

Not long afterwards, this teacher's decision was given formal backing in a report by Her Majesty's Inspectors, advising the Secretary of State to introduce a policy of setting by achievement in lower secondary and upper primary school classes (SOEID, 1996). The committee issuing this advice had also had access to a review of relevant research (Harlen and Malcolm, 1997), but it gave greater weight in arriving at its decision to inspectors' opinions that mixed-ability teaching places too great a burden on teachers. The authors of the report chose to advise setting as the course of action rather than to find ways to relieve this burden and so protect the lower achievers — a longer-term and perhaps initially more expensive way to address the problem.

To return to the pressure on teachers as users, this example illustrates the difficulties of applying general research findings to specific circumstances. The particular teachers in the school, the background of the pupils attending the school, their performance on entry, the form of the timetable — all these and many other factors make a school unique. Research can help to focus attention on aspects to be considered, indeed can provide a mirror in which a school can look at itself, but it cannot replace the need to collect the evidence for this self-study.

Communication between researchers and users

The communication of research needs

In relation to policy makers we are long past the time when it could be said that researchers had their feet off the ground and ignored these users' needs. Indeed in the UK, for the past decade users' views of what research is needed have increasingly dominated the allocation of direct research funding. This is not only true of government agencies and non-governmental funders of research but also of the research councils and to some extent of the charitable foundations whose funds are largely allocated to research in particular areas reflecting current policies and concerns to advance national productivity. The effect is that researchers feel that their work is confined to responding to others' agenda with no opportunity to propose what they perceive as important and valuable. In these circumstances 'fundamental' research, such as would inform an understanding of how learning takes place at different points in our lives, or of the cumulative effect of learning and of how it may be brought about through means other than the conventional teacher–student relationship, are being neglected.

At the practitioner level, the matter is rather different. It is hard for their needs to be communicated and addressed by funded research. In this context, it is relevant to point to the difference between England & Wales and Scotland. Each year the Research and Intelligence Unit (RIU) of the Scottish Office Education and Industry Department (SOEID) collects from all bodies with interests in education — from teachers' and parents' organisations as well as from local authorities and industry — suggestions for areas which are seen as priorities for research. Although this gives practitioners a voice in suggesting the general areas for research, it does not touch upon the very specific needs of individual schools and teachers, of course, and possibly never can do. It may be unrealistic to suppose that outsiders to an institution can solve practitioners' problems (Brown, 1994) when these are situated in very specific

circumstances. What is often needed in terms of research can best be done by practitioners themselves. However, in doing this, since it is unreasonable to expect teachers to add research to a full teaching job, they should be able to call upon help from outsiders, be given some dedicated time and have access to research in appropriate forms. In turn this means that their needs must be taken into account when funding priorities are decided and the Scottish example is a step in the right direction.

The communication of research knowledge

It is not the function of this paper to consider the complex matters of dissemination and the diverse ways in which research knowledge can be brought to different audiences. The point to be emphasised here is one that underlines the advantages of communicating research knowledge derived from a review of several studies rather than from each alone.

Effects in education are the result of the operation of multiple and interconnected variables. Even in closely focused research commissioned by funders to provide information about a particular issue or the effect of alternative practices or to evaluate an innovation, a clear-cut outcome is unlikely. We have only to think of studies of school effectiveness, of the provision of specialists in primary schools or the evaluation of new curricular guidelines to realise that the best designed study is likely to show 'on the one hand ...' and 'on the other hand...'. Single studies — meaning research carried out by a particular group of researchers, addressing particular questions in a particular place at a particular time — have several weaknesses as a basis for guiding educational change. These arise from the constraints and influences which affect a study, mentioned earlier, which inevitably restrict or bias the evidence that it provides, as well as from the complex nature of the situations which are researched. If this is the case in relation to research into well-defined topics, how much more true it is likely to be in relation to broader policy issues. Would it be preferable to present the findings of a study as hypotheses and at the same time to show the extent to which each hypothesis is supported by relevant previous studies? Information from different studies, in different contexts, should help to support what emerges 'on the one hand' and 'on the other'. These are the issues taken up in the next part of this paper.

Research reviews and educational change

Roles of reviews

There are many roles for reviews or syntheses of research beyond the conventional survey which forms an early (and easily skipped) chapter in a thesis or a research report. In relation to educational change, some of these roles are:

- Bringing together the range of available options for change and setting out what is known about the consequences of adopting each. In other words, this is taking advantage of what several people have described as the fate of research, which is to go into the general atmosphere and change the climate of decision making. Reviews help to make sure that what is in the 'atmosphere' at any time is captured so that it does indeed enter 'into the thinking of policy makers and/or practitioners' (Tizard, 1975).

- Combining different kinds of evidence that are needed to provide a firm basis for action, in particular to combine surveys (which can tell about how an innovation is judged by those implementing it but not about how to improve it) with close studies of the processes and interactions accompanying its use (which help to understand what happens in some circumstances but which may not be directly applicable in other circumstances).

- Distinguishing the relevant studies from the less relevant, the good quality research from the poor quality, thus ensuring that decisions are based on the best evidence available.

- Providing guidance as to how dependable the available evidence is in relation to the area of proposed change (which may be novel and consequently not well researched).

- Presenting findings of research on the effect of actions taken to bring about change in one situation in the context of studies of similar actions in different situations and of different actions in similar situations.

This list includes roles not just at the point of deciding options for change before they are made, but at the point of communicating findings in a way that addresses the needs of the decision makers. The implications of a single study generally have to be hedged around with reservations, but could be made firmer if taken as the basis of hypotheses for which evidence from other studies is presented. Taking this seriously would mean giving greater

emphases to the parts of a project, normally at the beginning and end, which are based on reflection and for which time is often particularly limited. These parts of studies would require more extensive work than they are given at present. If they were to become expected (and funded) parts of research, it would inevitably mean less funding available for other parts of projects, or fewer projects. Perhaps more use could be made of existing work, for there is much empirical evidence that is never fully analysed and properly examined; it would not necessarily be a bad thing if more funding were to be allocated to careful and critical reviews of research.

Such a move might well be an advantage whether or not we are considering research concerned with educational reform, although this is the context in which it is discussed here. The shrinking of the globe by information technology is an important reason for taking an international view of possibilities and of the likely effects of change. Politicians and their advisers already do this through the visits they make to other countries to compare practices (for example in relation to the National Curriculum and changes in school inspection procedures). Clearly, others' practices *are* seen as relevant and the more ideas are pooled across the globe the more there will be in common in these practices and the greater the importance of taking note of the research in other countries. At present there is a tendency for selective adoption of practices from other countries without the full knowledge of the circumstances and conditions which support them. Careful reading and review of research conducted in these other countries would add to understanding of how and why particular policies have been adopted. It might also be possible to restrain the naïve conclusions drawn from international surveys of educational achievement — or is this asking too much? It is neither acceptable to dismiss research carried out in other countries as irrelevant because of cultural differences or to seize upon evidence of what is possible in other countries ignoring cultural differences. Careful reviews of research should be able to provide guidance as to what is relevant to a particular decision and how much the findings can be relied upon.

Different facets of research studies

Any research study has several facets, or ways in which it adds to knowledge. As well as the particular topic or subject matter, there is always a wider set of issues to which it relates. For example, studies as different as the evaluation of the implementation of new curriculum guidelines and the evaluation of the pilot introduction of devolved school management are both about the process of change in educational institutions. The study of distance learning

packs for dentists and of the approaches to training playgroup leaders are very different in topic but are both about the education of adults. Beyond this, there is the methodology employed. An investigation of young pupils' attitudes to truanting and a study of adults' attitudes to genetic engineering may seem to have little in common but both have included the use of stories as a means of gaining access to attitudes. Thus the potential contribution to understanding that a study can make can be at least threefold:

to methodology: even an apparently routine use of a method will add something to the understanding of the type of information which can be gained in a particular set of circumstances by a survey, a study or a small number of cases or a single case or a combination of methods

to a generic theme: for example, the process of change, the development of attitudes, the learning of adults, the use of qualitative criteria

to the specific subject or topic: the implementation of the particular curriculum guidelines, the further education of dentists, the understanding of truancy and so on.

We are adept at reviewing and bringing together studies with the same subject matter and we may review studies employing particular methods, especially in the context of research training, but there is a general neglect of reviews of the generic aspect of studies. This is an important omission, for it is at this level that research can be applied widely to new situations and subject matter. Those who have identified generic links between research in different situations and concerning subject matter have advanced understanding of issues that have widespread relevance. An example is Michael Fullan's 'Eight basic lessons of the new paradigm of change' (Fullan, 1995):

- You can't mandate what matters
  (The more complex the change the less you can force it)

- Change is a journey not a blueprint
  (Change is non-linear, loaded with uncertainty and excitement and sometimes perverse)

- Problems are our friends
  (Problems are inevitable and you can't learn without them)

- Vision and strategic planning come later
  (Premature visions and planning blind)

- Individualism and collectivism must have equal power
  (There are no one-sided solutions to isolation and groupthink)

- Neither centralisation nor decentralisation works
  (Both top-down and bottom-up strategies are necessary)

- Connection with the wider environment is critical for success
  (The best organisations learn externally as well as internally)

- Every person is a change agent
  (Change is too important to leave to the experts, personal mind set and master is the ultimate protection).

These are clearly useful guidelines for anyone planning or implementing change. Other general areas of wide relevance in education would benefit from similar statements which avoid the simplistic prescription of the approach of the American synthesis of research on teaching and learning published as *What Works* (US Department of Education, 1986).

But what kind of reviews are required? The words synthesis and review are often used interchangeably, but they are not the same. At times a synthesis, aiming for a statement pointing to specific action, is appropriate, but perhaps more often a useful review is one that describes different findings and the circumstances in which they were found. This brings me to the subject of how reviews are conducted and the form they might take.

Review methodology

For research reviews to have the roles suggested above and to be more dependable than either the results from a single study or from a group of studies that happen to be accessible to the reviewer, they need to be conducted in a more systematic way than the traditional approach. It is in the methodology employed for this that it is possible for education to learn from other professions, especially medicine.

Early in 1996 a lively debate was stirred by David Hargreaves' comparison of the practice of medicine and the practice of education in relation to research. Hargreaves (1996) claimed that medicine is a research-based profession whilst this is not the case for education. Whilst not wishing to enter this particular debate (which for those wanting to know more, was well represented in the BERA Newsletter of July 1996), there is a more fruitful area where education could learn from medicine: the way in which research is reviewed, synthesised and made available to medical practitioners.

In medicine there is an increasing interest in setting down guidelines to advise

clinicians on best practice by drawing on the immense and growing volume of research rather than relying on the consensus developed through practice. This marks an attempt to move from consensus-based guidelines, based on expert opinions but not drawing systematically on research, to evidence-based guidelines. The practice of producing guidelines on clinical treatment through systematic review of research is not universal, although attempts have been made to introduce it over many years. It remains the case that many guidelines are developed by expert groups without the benefit of thorough reviews of research. Thus no claim is being made that this is a widespread practice, only that it is one worth considering by educational researchers, policy makers and practitioners.

In Scotland the process of developing evidence-based guidelines for medical practice is undertaken by a multidisciplinary project, the Scottish Intercollegiate Guidelines Network (SIGN). The procedures being implemented by SIGN to develop guidelines for treatment through systematic reviews of published literature involve making explicit the strategies used in the search of possible relevant studies and the criteria by which studies are included or excluded. To this extent, the approach has much in common with that advocated and practised by Robert Slavin called 'best-evidence synthesis' for reviewing educational research (Slavin, 1986). In turn Slavin borrowed his ideas for 'best evidence' from the profession of law:

> In law, there is a principle that the same evidence that would be essential in one case might be disregarded in another because in the second case there is better evidence available. For example, in a case of disputed authorship, a typed manuscript might be critical evidence if no hand-written copy is available, but if a hand-written copy exists, the typed copy would be inadmissible because it is no longer the best evidence (since the hand-written copy would be conclusive evidence of authorship).

> I would propose extending the principle of best evidence to the practice of research review. For example, if a literature contains several studies high in internal and external validity, then lower quality studies might be largely excluded from the review (Slavin, 1986, p. 6).

There is no promise here of a process which can avoid value judgements about what is 'best evidence' and of course there never can be. There is also the problem of what to do if there are no studies matching the 'best evidence' criteria. Slavin suggests that in such cases 'we might cautiously examine the less well designed studies to see if there is adequate unbiased information to

come to any conclusion' (ibid.). It is clearly a contentious matter as to whether or not to include studies which do not match up to what is expected in terms of rigour and lack of bias. Indeed it is one of the main criticism of meta-analysis that all studies are included with no exclusions based on conceptions of quality. Eysenck (1994), for example, has strongly and cogently criticised the practice of meta-analysis for being too mechanical and for including 'all relevant material, good, bad and indifferent'. Although this criticism has been rebuffed, it none the less draws attention to the need to be selective in reviewing research. It is here that the medical project provides a solution worth considering. The evidence from various studies is graded as being at one of several levels according to how it is obtained. The following descriptions of levels are currently used by SIGN and are based on the US Agency for Health Care Policy and Research (1992). They clearly reflect the criteria of quality that are used in medical studies:

Level	Type of Evidence
Ia	Evidence obtained from meta-analysis of randomised controlled trials
Ib	Evidence obtained from at least one randomised controlled trial
IIa	Evidence obtained from at least one well-designed controlled study without randomisation
IIb	Evidence obtained from at least one other type of well-designed quasi-experimental study
III	Evidence obtained from well-designed non-experimental descriptive studies, such as comparative studies, correlation studies and case control studies
IV	Evidence obtained from expert committee reports or opinions and/or clinical experience of respected authorities

(US. AHCPR, 1992)

148

Recommendations for medical practice derived from the evidence are then graded according to the level of the evidence on which they are based. This gives the guideline users the opportunity to consider the strength of recommendations of possible courses of action in deciding which to take in a particular case. Thus level IV evidence would be seen as the least dependable but might still be used in the absence of research studies meeting the criteria at the higher levels. In the medical context, random controlled studies and meta-analysis of their results are given most credibility. However, this would not be the case in all professions and is challenged even within those related to medicine. The recognition that not all medical treatments are amenable to being tested in randomised trials means that the criteria for grading evidence have to be developed further.

The purpose in describing this approach here is not to advocate one or another set of criteria for judging the value of studies for a particular use, but rather to suggest the consideration of applying the approach in education. For instance, criteria might be established in particular cases and made explicit to users when research is reviewed. The criteria could reflect different approaches to reviewing beyond the meta-analysis embraced by several reviewers, to include the best-evidence synthesis and the hypothetico-deductive approach preferred by Eysenck (1994). The latter has a particular attractiveness in education since it enables qualitative and quantitative studies to be combined. It involves looking not at the size of effects in a particular study, as in meta-analysis, but at the possible causes of effects and considering alternative hypotheses to the ones underlying the study. Evidence supporting or refuting the various hypotheses is then sought in other studies.

To consider how this might be applied in education, suppose that, for a particular focus of review, whether it is a content area, a generic theme or methodology, the reviewers establish criteria of quality. We would expect some criteria to be common across all areas of reviewing — such as reference to: consistency in the choice of methods, the nature of data and the manner in which data are analysed, interpreted and reported, with aims of the research and philosophical stance; systematic rigour and avoidance of bias in the conduct of the investigation or scholarly study; addressing and advancing existing knowledge and theory; and accordance with ethical principles. But there would be variations from one context to another, for example in the size of samples, the duration of the research and the number and selection of cases studied.

When the criteria of quality have been made explicit, the studies relevant to the focus of the review would be identified and graded in terms of the extent

to which they meet the criteria. Then the evidence from those studies fully meeting the criteria would be accorded most confidence; that from studies not matching all the criteria of quality, less confidence; and where research is particularly patchy or where there is only informed opinion to go on, least confidence of all.

The confidence given to any action or explanation based on a hypothesis might be defined in this kind of way:

A All studies meeting the criteria of relevance and quality provide clear support for the hypothesis.

B Studies meeting some but not all the criteria of relevance and quality provide some support on balance for the hypothesis.

C Studies meeting few of the criteria of quality and/or the opinions of relevant committee reports or individuals support the hypothesis.

Reviewers could draw conclusions at one of these levels, indicating that, for example, in the absence of level A evidence, then the best we have is at level B or C. But only in the absence of level A evidence would evidence at the other levels be used, and even then, as Slavin said, 'cautiously'. It might, of course, be the case that level A evidence is equivocal in relation to the hypothesis, whilst biased evidence at level B lends support to it. In my view this would be the value of the approach, that is, to prevent users accepting every piece of research evidence as being as good as any other. This is particularly important where decisions about change or about taking one course of action or another are to be based on the interpretation of research evidence.

In conclusion

The volume of published research in education in the UK is increasing rapidly mainly through the influence of the Research Assessment Exercise (HEFCE, HEFCW, SHEFC, DENI, 1996). This paper has been concerned with the use of this growing volume of information in relation to its 'instrumental' function in the context of educational change, although reference has been made to its value for adding to understanding and challenging existing ideas, or the 'enlightenment' function, of research, as Nisbet (1988) put it. For either function there has to be some way of accessing, sorting and bringing together research on a particular topic, if, as has been argued, single studies are inherently influenced by their particular contexts. In the case of the instrumental function, there is a danger that, if this is not done, then policy

makers and practitioners will either give up the effort to use research or will use it selectively. In both cases the result will be that educational decisions are less well informed by research than should be the case and are more influenced by opinion and political persuasion.

I have described a possible practice, drawn from the medical profession, which could help to ensure that proposed changes are informed by the highest quality and most relevant research there is available. Of course this practice is not entirely new to education; many of the most thorough and influential reviews (e.g., Tizard, 1975; Crooks, 1988; Slavin, 1987 and 1990; Black, 1993) implicitly reflect attention to the matters which are made explicit in the procedures suggested.

Procedures alone will not effect the change that is needed to make better use of better research. There has to be the will to use research, to see it as important, and there has to be a change in the culture in which research takes place, which at present gives more support and prestige to empirical studies than to systematic reviews. There is some sign of this happening in the contracts recently given for reviews by OFSTED and the SOEID (e.g. Howe, 1997). The further development of this trend is to be welcomed for it would not only make life easier for users of research but it would also develop healthy debate about quality and relevance of studies that hopefully would lead to improvement in both. This in turn would enable research to serve educational change more effectively.

References

BLACK, P. J. (1993). 'Formative and summative assessment by teachers', *Studies in Science Education*, **21**, 49-97.

BOALER, J. (in press). 'Setting, streaming and mixed ability teaching.' In: MAGUIRE, M. and DILLON, T. (Eds) *Becoming a Teacher*. Buckingham: Open University Press.

BRITISH EDUCATIONAL RESEARCH ASSOCIATION (BERA) (1992). *Ethical Guidelines for Educational Research*. Available from the BERA Office, SCRE, Edinburgh.

BROWN, S. (1994). 'Research in education: what influence on policy and practice?' *Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization*, **7**, 4, 94-107.

CROOKS, T. J. (1988). 'The impact of classroom evaluation practices on students', *Review of Educational Research*, **58**, 4, 438-81.

EYSENCK, H. J. (1994). 'Meta-analysis and its problems', *British Medical Journal*, **309** September.

FULLAN, M. (1995). *Change Forces*. Brighton: Falmer Press.

HARGREAVES, D. (1996). *Teaching as a Research-Based Profession: Possibilities and Prospects*. London: Teacher Training Agency.

HARLEN, W. and MALCOLM. H. (1997). *Setting and Streaming: A Research Review*. Edinburgh: Scottish Council for Research in Education.

HIGHER EDUCATION FUNDING COUNCILS (HEFCE, HEFCW, SHEFC, DENI) (1996). *1996 Research Assessment Exercise: The Outcome*. Bristol: HEFCE.

HOWE, C. (1997). *Gender and Classroom Interaction: A Research Review*. Edinburgh: Scottish Council for Research in Education.

HUBERMAN, M. (1994). 'Research utilization: the state of the art', *Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization*, **7**, 4, 13-33.

McGLYNN, A. S. (1991). 'Research and policy-making in Scottish education: the role of the Research and Intelligence Unit (RIU)', *Scottish Educational Review*, **23**, 2, 112-17.

NISBET, J. (1988). 'The contribution of research to education.' In: BROWN, S. and WAKE, R. (Eds) *Education in Transition*. Edinburgh: SCRE.

SCOTTISH OFFICE EDUCATION AND INDUSTRY DEPARTMENT (SOEID) (1996). *Achievement for All: a report on Selection within Schools by HM Inspectors of Schools*. Edinburgh: HMSO.

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

SLAVIN, R. E. (1987). 'Ability grouping and student achievement in elementary schools: a best-evidence synthesis', *Review of Educational Research,* **57**, 3, 293-336.

SLAVIN, R. E. (1990). 'Student achievement effects of ability grouping in secondary schools: a best-evidence synthesis', *Review of Educational Research,* **60**, 3, 471-99.

TIZARD, B. (1975). *Early Childhood Education: A Review and Discussion of Current Research in Britain.* Windsor: NFER Publishing Company.

UNITED STATES AGENCY FOR HEALTH CARE POLICY AND RESEARCH (1992). *Acute Pain Management: Operative or Medical Procedures and Trauma.* Rockville, MD: Agency for Health Care Policy and Research Publications.

UNITED STATES DEPARTMENT OF EDUCATION (1986). *What Works: Research about Teaching and Learning.* Washington, DC: US Department of Education.

Respondents' comments

David Raffe, University of Edinburgh

I would like to congratulate the NFER on its achievements during its first 50 years, on its enterprise and flair in organising this conference, and on its boldness in entrusting the first afternoon session to two speakers from Scotland.

It is interesting that the chair and three speakers at this session, on research and educational reform, should all be from small countries (Austria, New Zealand and Scotland). In a small country it is possible to develop a different quality of relationships among researchers, policy makers and practitioners, and these can enhance the contribution of research to reform. The Scottish education policy community can meet within a single room, and many of them often do so. This closeness brings problems as well as benefits (it means that special efforts must be made to preserve the independence of research and its critical distance from policy) but it enriches the potential contributions of research to educational reform. It also reminds us that generalisations about the relationship of research to the communities of policy or practice must take account of the different ways in which these communities are organised. I return to this point at the end of my presentation.

Wynne Harlen's paper offers a valuable and challenging analysis of the relationship of research to reform in education. I will make three sets of comments. These respectively register a point of disagreement, express strong agreement with her central argument, and discuss the implications for the organisation and funding of research.

First, although we certainly need more 'careful and critical reviews of research', I would be cautious about using a formal hierarchy of quality of research methods to arbitrate between research studies. A good review is more than the sum of the research projects which it covers. It adds to these projects by identifying, and learning from, the different findings or perspectives generated by different theoretical or methodological approaches. If two research approaches produce different findings, the reviewer's first question should be, not Which finding is valid?, but Why do these findings differ? In trying to answer this rather broader question the reviewer is likely to arrive at a better understanding of the subject under review, and perhaps a better theoretical framework, than by considering only the conclusions of the 'better quality' approach and ignoring those of the other approach. Of course, the reviewer may eventually answer the question in terms of quality; that is, s/he may conclude that one approach produces a different result from another because it uses a better and more valid method. But to determine in advance

on 'quality' as the only arbiter of conflicting research findings would be to diminish the creative and generative potential of reviews

Second, I agree strongly with Wynne's principal argument, that to understand the influence of research on policy it is necessary to look more broadly than a single project. There are many reasons why this is the case, including the differing timescales of research and policy or practice, the specific and changing nature of policy agendas, the uncertainty of conclusions from single studies, the cumulative nature of much research, the influence of research on conceptual frameworks and ways of thinking, the osmotic influence of research and so on. These reasons have been discussed by Wynne and earlier speakers and I do not need to elaborate them here. I will simply reinforce them by appealing to the personal experience of researchers. If we have had any influence on policy and practice this is likely to have been exerted less through the findings of individual projects than through the body of knowledge and understanding to which these projects have contributed; less by communicating individual research 'findings' than by applying the expertise developed across a range of research activities, and by contributing to broader conceptual and explanatory frameworks; and less through answering policy makers' and practitioners' questions than through helping them to ask the right questions. For research to have maximum influence, it follows that the relationships of researchers to policy makers must be diffuse, broad-based and interactive.

In such a context, research may have an impact that is nevertheless difficult to quantify. Six years ago I was a member of a group of six people, including five academics, who wrote a policy paper called *A British Baccalauréat*. This was not a research report, but its analysis drew strongly both on research findings and on our knowledge and judgement gained in the practice of research. It proposed a unified system of education at 16-plus. A system on similar lines is about to be introduced by the government in Scotland; a paler version is now the policy (at least as a long-term objective) of the opposition parties in England and Wales. There were, of course, numerous other influences and pressures on the policy process. The timing of the *British Baccalauréat* and the fact that it found a receptive audience were probably as important as the specific arguments that it offered. The conclusion I draw from this example is not that research had an influence (although in this case I am sure that it did), but rather that the processes of policy formation are so diffuse and disparate that it is impossible to quantify its impact with any rigour. Indeed any definition of impact that lends itself to easy measurement is almost certain to be so narrow as to trivialise the concept.

155

My third set of comments considers the implications of Wynne Harlen's argument that the influence of research on policy and practice transcends individual projects. I will list five implications for the organisation of research.

The nature of research As Wynne has commented, there is a need for more reviews and syntheses of existing research. In addition, new projects need to be designed with broader, more strategic and longer-term objectives. An important criterion for judging any new research activity should be its capacity to contribute to the long-term development of theory, knowledge and methods in the field. This criterion should apply to all forms of research, including applied research. In practice, much so-called 'applied' research turns out to be inapplicable, because it is based on the fallacy that a single project can answer the policy question, as well as on the assumption that the question has been adequately defined and will still be on the agenda by the time the research is completed.

The organisation and funding of research Most educational research in the UK is organised according to a 'fragmented–competitive' model, which breaks research into (increasingly) small 'projects' and distributes these on a competitive quasi-market basis. This model is an inefficient and ineffective way to organise a national research programme. It inhibits long-term horizons and planning of research; it is wasteful of resources; it makes it difficult to provide research careers and the development of knowledge and expertise that goes with them; and it narrows and distorts relationships with users. In other words, it inhibits the very process of developing knowledge across different projects which is essential for research to be of practical value. There are several possible alternatives to the fragmented–competitive model, including greater concentration of research, as proposed by Seamus Hegarty (above), and more selectivity within higher education. Whatever the disadvantages of these alternatives, they are better than the present arrangement.

Disequilibrium in the research market No market can function effectively if it remains in long-term disequilibrium. The problems described above are aggravated by the chronic over-supply in the research market — the imbalance between the number of researchers seeking funding and the availability of funds. This imbalance results from a variety of factors including the research assessment exercise, the academic reward structure and HE institutions' jockeying for status. The result is a grotesque misuse of resources. For example, on my calculation, more resources were devoted to bidding for and allocating projects in the ESRC's *Learning Society* programme, than were actually made available for research through the programme. The balance

sheet for Socio-Economic Research within the EU's Fourth Framework Programme is probably similar.

Relationships with policy makers and practitioners Relationships between researchers and users must be organised to reflect the diffuse, broad-based and interactive nature of the process described above. It should not be organised mainly on the basis of projects; even less should it be limited to market relationships of customer and supplier which are favoured by the fragmented–competitive model. This underlines the need for a change in the organisation and funding of research. At present my main responsibility as director of a research centre is to maintain its viability in the market for funding; most of the activities which develop and sustain these broader relationships with users are unfunded, and it is hard to give them priority under the current funding regime.

The context of policy and practice Finally, I return to the point with which I began. The relationship of research to reform depends fundamentally on the social, political and economic organisation of the reform process itself. One of the most important conditions for an effective relationship is an attitude of openness on the part of government or other users of research. We can contribute nothing of value to users whose criterion of useful research is that it confirms policy decisions which have already been taken. If only education departments, authorities and institutions would follow the logic of their own rhetoric and seek genuinely to become learning organisations, the potential contribution of research would be enormous.

Karl-Heinz Gruber, University of Vienna

Sandwiched between Barry McGaw's analysis of the nature and function of educational research and Ramsay Selden's survey of educational reform (Impact and dissemination), it looked as if all that was left of the grand topic Educational Research and Educational Reform for Wynne Harlen was the explication of the word 'and'. Quite wrong. Harlen elegantly solved her assignment by first identifying the constraints and influences on the producers and the users of educational research, and then by pleading for more sophistication in the production and utilisation of research reviews. I agree with her analysis to such a high degree that all I can offer are a few loosely connected, outlandish remarks which reveal two facts:

1   The context from which I am commenting is the 'culture of education' of the German-speaking countries where educational research and educational reform, while adhering to the universal quest for rationality, enlightenment and improvement, have different self-concepts and different modes of interaction, different in the sense that they do not share the 'self-evidence' and common assumptions implicit in the English language discourse on education which allows the British, Scottish, American and Australian members of this conference to communicate with ease and confidence.

2   Seamus Hegarty took a certain risk when he invited somebody from Vienna to this conference. The intellectual climate of Vienna is notorious for its moodiness, oscillating between the sobriety of Sigmund Freud (whose memorial next to the University of Vienna bears the much ignored inscription 'The voice of reason is quiet') and the frivolity of the 'Fledermaus' of Johann Strauss in which one character proclaims, 'Happy the man who forgets what he cannot change'. Being neither much of a Freudian nor of a waltzing disposition, I take the liberty of approaching Harlen's earnest Scottish erudition with Viennese semi-seriousness. This unfortunately means that instead of observing her plea for 'best evidence' I shall resort to anecdotes, daring comparisons, single cases and other sorts of 'worst evidence'.

Trust and alienation

Harlen uses a bipolar model to illustrate the relationships between researchers on the one side and users (policy makers and practitioners) on the other. Perhaps a triangular model would be more suitable to bring out the fundamental differences between these three cultures or communities with

their specific agendas, legitimation, time frames, gratifications, client pressures, claims of responsibility, etc. (always keeping in mind that each one of these domains is highly differentiated in itself and that individuals or groups may define their role in a distinctive way). The 'distance' and 'closeness', trust and confidence, between these three cultures varies from country to country. From a Continental European point of view the relationship between the three communities in the UK traditionally seemed to be close and the trust between them high, the outcome of many factors such as

- English pragmatism;
- more cross-fertilising and confidence-building career mobility between the three domains than in most other European countries;
- a healthy epistemological balance between nomothetic and idiosyncratic approaches (surveys and single case studies);
- educational administration/policy studies as a mediating discipline;
- the existence of respected mediating, developing and implementing agencies such as the former Schools Council.

Is the emergence of the term 'educational establishment' an indication of an erosion of this relationship of trust? Over the past 15 years some politicians use this multi-purpose expression of contempt also to show their disrespect for research or to devalue it as politically biased if it does not support their own political agenda.

Misunderstandings, conflicting interests and cultural clashes are certainly not new, as Husen and Kogan (1984) have shown, but during the last two decades the alienation between research and policy formation seems to have grown throughout the OECD and led to what a recent study calls 'a crisis of confidence', a troublesome constellation of underfunding, unrealistic policy expectations, unrealistic research promises and a rich array of mutual accusations (OECD, 1995).

This loss of confidence is particularly noticeable in the German-speaking world, where the trust had not been particularly strong in the first place. Teichler comments on the Federal Republic of Germany: '...Administrators tend...to believe in the strengths of administrators to solve almost all problems without relying on research and mistrust the political inclinations of researchers' (Teichler quoted by Trow in Husen and Kogan 1984, p. 262).

Large sectors of the German and Austrian educational research community

withdrew in frustration from systemic and macro-political issues when it became apparent that in the wake of a neo-conservative revival the short-lived educational renaissance of the 1970s with large-scale school experiments, curriculum development and a political readiness to commission and utilise research was coming to the end. In Austria a coalition government decided against the continuation of the experimental comprehensive school programme and in favour of retaining early selection in lower secondary education, despite the fact that the findings of more than ten years of unbiased evaluation were overwhelmingly pro-comprehensive.

The relationship between educational researchers and practitioners is not trouble-free either. One personal experience helped me to overcome my Continental European romantic notion of the closeness of research production and research utilisation in the UK. When Jerome Bruner in the late 1970s conducted his action research Oxford Pre-school Project (in which I was a fly-on-the-wall observer for one year), some of the pre-school teachers felt uneasy about their 'partner' role and wondered to what extent they were more than the usual deliverers of research data. One of them told Bruner this joke. A hen and a pig walk down a high street. The hen sees a sign at a cafe which announces 'Ham and Eggs — only 99 pence'. When she says to the pig 'Look, ham and eggs for only 99 pence, what a bargain!' the pig gets angry and responds, 'What do you mean by "bargain"? For you this may be just a day's work, but for me it is a matter of life and death.' The professor was startled by the implication that the teacher saw her involvement in the project as 'existential' while the researchers might have just a fleeting data-gathering interest in it. Quite concerned, he asked her: 'I sincerely hope you are not accusing us of "hit-and-run" research'. To which she responded: 'Oh no. Actually I do not know much educational research, and the little I know is mostly of the "miss-and-run" kind'. For the rest of the afternoon the great man looked rather thoughtful, possibly speculating about the essay Susan Fuhrman would write 15 years later on 'Uniting Producers and Consumers' (Fuhrman, 1994).

'Useful' and 'annoying' research

As an 'organic developmental process' (in the somewhat poetic words of a recent American publication) educational research is a highly differentiated, multi-purpose enterprise (National Research Council, 1992). The way in which institutions and individual researchers give priority to the many (and partly conflicting) functions and roles available to them depends, as Harlen shows, not only on their beliefs and self-concepts but also on the priorities

and aversions of those who demand, initiate and through funding ultimately enable certain kinds of research.

In the process of re-mandating the US Office of Educational Research and Improvement the roles of educational research were conveniently listed:

Research ...provides warnings of problems in education...,

> ...informs policy debates by testing the assumptions that underlie arguments on all sides of an issue...,

> ...evaluates the consequences of programs and policies...,

> ...provides insights into basic processes of individual and organisational functioning... and

> ...provides enlightenment — new perspectives, new ideas, new conceptualisations of problems and new priorities' (National Research Council, 1992, p. 14).

In the German-speaking countries educational research sets much greater store by its 'critical' function. The prevailing disenchantment between research, policy making and practice has affected the self-concepts of the three major types of research production. The few, big independent research institutes continue their mostly evaluative functionalist work with reduced public funding and with unspectacular impact on the macro-politics of education. The other two producers of research, the predominantly curiosity-driven university research and the institutes under the control of Länder administrations, have drifted noticeably apart. While the former is often independent, small-scale, critical and in danger of becoming irrelevant to practice, the latter may be too close to their political–administrative paymasters to raise 'inopportune' issues. Very little energy is presently invested in the prospective, creative aspects of research. Policy research — a fledgling sub-discipline trying to establish itself *vis-à-vis* the self-confident, legalistic routines of the bureaucracy of education — is still an endangered species.

When a scientific working group of the generally respected Deutscher Bildungsrat (German Education Council) recommended decentralisation within the highly centralised Länder and an empowerment of the individual schools, this suggestion of a 'self-amputation' by the political–administrative system infuriated conservative politicians so much that they decided to dispense with scientific policy guidance altogether and abolished the Bildungsrat.

It may be one of the ironies of history or just a demonstration of the non-linear and retarded impact of research that over the past few years schools in Germany and Austria (as in other OECD countries) are receiving the autonomy for site-based innovation which caused so much turbulence 20 years ago. Perhaps one could mis-quote Carol Weiss and Sigmund Freud at the same time by saying 'The noise of reason is quiet and works slowly'.

What is 'international' research?

In the second part of her paper Wynne Harlen is concerned with the improvement of research reviews and their utilisation. Her proposal to take an international view and 'to pool ideas across the globe' by including research conducted in other countries does not only give due recognition to the universality of science but is also a sad reminder that large parts of 'international' research remain hidden and unacknowledged behind scarcely penetrated language barriers. English may be the international lingua franca of (educational) research but it is also a powerful filter for 'non-Anglo' research. *Quod erat demonstrandum*, as they used to say in the former lingua franca of scholarship.

Passing through London on my way to this conference, I visited the bookshop of the Institute of Education. Looking for a recent research review, I thought I had found the ideal example of Harlen's raised standards:

- It is written with the methodological aspiration of 'best evidence analysis'.

- Its approach is interdisciplinary.

- It contains a chapter titled International Reviews.

- It begins with the most promising sentence: 'There is a vast international literature on the effects of grouping pupils by ability...'.

- As a very personal bonus, it made me feel 25 years younger: in the early 1970s I had spent a year in England asking the very question which happens to be the title of this review — *What Do We Know About the Grouping of Pupils by Ability?* (Hallam and Toutounji, 1996)

This is a fine piece of research, but its 280 references do not really reflect the 'vastness of the international literature' mentioned in its first sentence. Only three of the titles are not in English (intriguingly two are in Norwegian and one in Dutch) and the chapter International Reviews draws almost exclusively on American studies. The truly vast body of highly relevant grouping and

differentiation research conducted in the German-speaking countries in the course of the evaluation of the comprehensive school experiments of the 1970s and 1980s remains outside the international aspirations of this study. This is, to be fair, the usual state of affairs as anyone will find out who bothers to look for French or German references in English- language literature.

Finally, I cannot help mentioning a publication mysteriously missing among the references: Alfred Yates' *Grouping in Education*, a book which at its time received so much Europe-wide attention that it became one of the few English education books (together with the Plowden Report and Michael Rutter's *Fifteen Thousand Hours*) to be translated into German. But this is of course none of my business but something for the NFER to brood over: Alfred Yates was after all one of its former directors.

References

HALLAM, S. and TOUTOUNJI, I. (1996). *What Do We Know About the Grouping of Pupils by Ability?* London: Institute of Education.

HUSEN, T. and KOGAN, M. (Eds) (1984). *Educational Research and Policy. How Do They Relate?* Oxford: Pergamon.

FUHRMAN, S. (1994). 'Uniting producers and consumers: challenges in creating and utilizing educational research and development. In: OECD/OERI *Education Research and Reform: an International Perspective*. Washington, DC: US Government Printing Office.

NATIONAL RESEARCH COUNCIL (1992). *Research and Educational Reform: Roles for the Office of Educational Research and Improvement*. Washington, DC: National Academy Press.

OECD (1995). *Educational Research and Development: Trends, Issues and Challenges*. Paris: OECD.

YATES, A. (1966). *Grouping in Education*. Stockholm: Almqvist and Wicksell.

6.

International educational research

Education is a national concern but many common issues arise across countries. One way of illuminating these issues is to examine the priorities and programmes of the international organisations concerned with educational research. In this section, leading figures from the main bodies describe the work of their organisations and highlight issues of common concern.

__Jarl Bengtsson__ presents the history of educational policy priorities at OECD and the opportunities for the educational research community to interact with evolving policy agendas. __Erhard Schulte__ outlines educational research at European Union level, specifically the Leonardo and Socrates programmes and the Targeted Socio-Economic Research programme. EURYDICE is an information network on education systems and policies in Europe, and __Luce Pépin__ demonstrates its relevance to the world of educational research.

The International Association for the Evaluation of Educational Achievement (IEA) is a network of institutions rather than a single organisation; __Tjeerd Plomp__ outlines how IEA works, what sorts of studies it conducts and likely future directions. __John Smyth__ presents an overview of UNESCO's involvement in educational research, drawing particular attention to the need for international education statistics. Finally, __Maris O'Rourke__ gives an account of the World Bank's three roles with regard to educational research and evaluation — consumer, facilitator and producer — and sets out future directions for the Bank's involvement.

OECD experiences and perspective on the interaction between education policy and research
Jarl Bengtsson, OECD

I am very pleased to be here on this important occasion and it has been a pleasure to cooperate with NFER over the years, particularly with Seamus in our recent CERI/OECD work on education R&D systems. NFER is 50 years old and soon it will be the same for the OECD. It's about 50 years since General Marshall gave his speech at Harvard on the importance of US aid in reconstructing Europe after the war. The so-called Marshall plan was the start of the OECD.

I thought I should briefly present the history of education policy priorities at OECD and the way the educational research community has interacted with these changing policy agendas, how it works today and some of the challenges ahead. There is no need, in this group, to elaborate on the mission and purpose of OECD. Most of you know perfectly well that the 'E' in OECD stands for economic and not educational. This fact has meant that work on education with the OECD has always been seen in the broader socio-economic context of the OECD mandate. Of course, recent debates and developments concerning knowledge societies and learning economies are making it increasingly difficult to separate education and learning from the economy.

Work on education started at the OECD in the late 1950s following the Sputnik event. Questions were raised whether the Western world was lagging behind the Soviets in science and technology. Was the curriculum in these subjects advanced enough, and was the training of scientific personnel adequate?

These kinds of issues and problems led to the first educational studies carried out by the OECD. During the 1960s, further studies were launched in the field of economics of education, manpower planning, rate of returns to education. The educational research community, which at this time had limited empirical and quantitative research experience, was little involved in this period of OECD work on education. Most of the research work was carried out by economists. A few exceptions were educational researchers involved in research on the 'reserve of talents'. However, towards the end of the 1960s, the situation changed and strong concerns were expressed by education authorities that there was a need for more qualitative information from the educational system and process. The Centre for Educational Research and Innovation (CERI) was created at the OECD in 1968, followed by the Education Committee in 1970.

The following decade became the decade of education expansion and the educational research community became heavily involved in these developments at the OECD. Education policy and educational research experienced a period of strong interaction and cooperation. Some of the key substantive areas dealt with were: equal educational opportunities; curriculum development; and recurrent education.

But by the end of the 1970s, the policy agenda had shifted. Three new priorities emerged. First, there was a new concern for raising standards and quality in education. OECD got involved in this as did the educational research community. Later on, this work led to the development of education indicators at the OECD, where a research community that was by now well versed in educational measurement played a very active role.

A second new policy concern was unemployment and the rapid development of labour market programmes, not least training programmes. Most researchers involved here were labour economists and, on the whole, the educational research community was very little involved although there were and still are many important questions about adult and youth learning and training of both a motivational and cognitive nature.

A third major development in the 1980s, in which OECD education work got involved, was that of competencies and skill development at the enterprise level. Enterprises were increasingly competing on the basis of the skills and competencies of their work force, and human resource development became important. The consulting industry benefited together with business economists, but few educational researchers were involved in this work. Today in the middle of the '90s, these three developments are still with us. But in the educational policy community, they are brought together under the umbrella concept of lifelong learning. When the Ministers of education met at OECD last year, they all agreed that the top priority for their education policy should be the implementation of lifelong learning for all.

The educational research community is involved in this important endeavour, but perhaps in many OECD countries it has been taken by surprise at how fast the policy community has shifted its priorities in favour of lifelong learning. It is also a matter of concern to note that, in many cases, the educational research community seems less equipped to meet and respond to this new policy priority.

In addition to this short overview of the links and interactions between policy developments in OECD education work and the role of the educational

research community, one can also look at this interaction from the perspective of different modes or categories of work that OECD carries out. At least four principal categories of work are worth mentioning.

A first category is OECD education policy reviews of individual member countries. Over the years, an excellent cooperation between 'examiners' with research and policy experience has developed. A matter of concern is perhaps that the pool of such examiners who need solid experience in policy, administration and international developments is somewhat small.

A second category of work relates to the analysis and clarification of specific policy areas. For instance, such areas today include: school to work transition; school failure; tertiary education; and early childhood education and care. The educational research community is heavily involved in this aspect of OECD work, but other social science communities are also involved. What is becoming increasingly important here for the researcher is to have a clear understanding of both the research and the policy issues that arise.

A third category is the recent booming area of international comparative education statistics and indicators. The OECD publication coming out of this work, *Education at a Glance* (OECD, 1996a), is today an OECD best-seller — selling better than OECD economic publications. In particular, the measurement-oriented educational research community is strongly involved in this work and it has strengthened considerably the international cooperation among educational researchers and statisticians, both at the national and international levels.

A fourth category is studies on significant innovations in the education system, referred to as What Works in Education Innovation. Studies here include Innovations in School Choice; Parents as Partners in Schooling; Teacher Training; and Innovations in Science, Mathematics and Technology Education. Educational researchers take an active part, but those people who also bring in some journalistic skills in terms of writing 'short and sharp' reports on these innovations are particularly sought out. A major problem with these studies on innovation is the unresolved question of how they can be transformed into mainstream education. This is a big challenge for educational researchers and perhaps there is a need for a Schumpeteran school of educational research.

Let me finish with another form of OECD cooperation with the educational research community, namely that on the education R&D system itself. Within CERI, we started a specific project a couple of years ago on education R&D.

The background was one of increasing concern that education R&D systems were in some sort of crisis with under-funding and some criticism as to output from the demand side of research.

Two important CERI/OECD publications have come out of this work. The first one, called *Educational Research and Development: Trends, Issues and Challenges* (OECD, 1995), makes it clear that the linear model between the educational researcher and the policy maker is on its way out and is being replaced by a much more interactive model in which three sets of stakeholders work together: researchers, practitioners and policy makers. This report also identifies and describes an emerging international market for educational research. The second report, called *Knowledge Bases for Education Policies* (OECD, 1996b), has two principal messages. First, and building on the report above, the processes of knowledge production, mediation and use need to be further understood and clarified for the three stakeholders referred to above. Second, the report sets out to describe briefly the emerging knowledge society and learning economy and how important it will be for the educational research community to take an active part in the understanding and clarification of the role of knowledge in a learning economy and society.

Today there is much debate about these concepts and developments, but very little is actually well understood partly because relatively little research is carried out by the social science community. For instance, if knowledge will be, as many policy makers claim, the principal raw material in the modern OECD economy of the next century, then there exists an urgent and challenging need to strengthen the research to understand such an economy, which is not very well captured and understood by present theories, concepts and paradigms in social sciences. For instance, if knowledge is like love — it grows when it is being shared — that would pose serious challenges for many economic theories, not to mention educational theories.

There is no doubt in my view that the educational research community, if it would be willing to be more risk taking, could make a significant contribution to the understanding of the knowledge society and the learning economy.

At the OECD, we see the understanding of this new economy and society as a key priority and hope that the educational research community will become partners in this work. I am absolutely convinced that if educational research can make a significant contribution to this work nationally and internationally, there will be no further talk about crisis or under-funding for educational research.

References

OECD (1995). *Educational Research and Development: Trends, Issues and Challenges.* Paris: OECD.

OECD (1996a). *Education at a Glance: OECD Indicators* (4th edn). Paris: OECD/CERI.

OECD (1996b). *Knowledge Bases for Education Policies.* Paris: OECD.

Educational research within the European Union
Erhard Schulte, European Commission

It is a great pleasure for me to attend this Jubilee Conference, which celebrates an outstanding institution in the field of educational research. I also feel honoured for being included in a list of outstanding representatives of the scientific community. A number of important issues are on the agenda of this conference and have already been addressed. From my point of view these issues reflect the importance which the subject deserves. It is often stated and still continues to be relevant: education and training are fundamental in nearly every respect, from the economic perspective as well as from that of the individual who needs to build perceptions, aspirations and guidelines in order to live life in a satisfactory way.

The recent document produced by the European Commission relating to education and training takes a broad range of relevant aspects into account. The White Paper *Teaching and Learning — Towards the Learning Society* puts emphasis on the relations between education and training on one hand and the economic and technological challenges on the other, underlining the problem of unemployment and educational needs in the emerging information society. But this document clearly states that 'to examine education and training in the context of employment does not mean reducing them simply to a means of obtaining qualifications. The essential aim of education and training has always been personal development and the successful integration of Europeans into society through the sharing of common values, the passing on of cultural heritage and the teaching of self-reliance.'

You may now realise that I am not only addressing the theme I have been invited to speak about but wish also to touch on its philosophical aspects. The latter is appealing, but captivating as well. It is my task at this conference to report on educational research within the European Commission. This task seems to be more apt for short, dry listing than for intellectual escapades. However, I will try to do both, to survey what is being launched in this field of research and to shed some light on the aims and context of educational research within the EU.

First and foremost, I have to define the subject I am speaking about. Educational research at EU level encompasses a great variety of activities. There are, for instance, those educational studies which relate originally to non-educational fields of European policies, such as employment or telecommunications. In Directorate General V for instance, which deals with

employment, industrial relations and social affairs, a variety of projects and activities are funded on an *ad hoc* basis, relating to special qualification aspects. I am going to leave out these activities, since in our context they are rather peripheral to educational research. These projects would not be classified as educational research activities in this specific sense.

In order to outline the main efforts made at Community level, I will concentrate on two areas where research on education and training plays a prominent role. These are the programmes *Leonardo da Vinci* and *Socrates*, both of which are carried out in Directorate General XXII of the Commission, and the *Targeted Socio-Economic Research* programme, usually called by its acronyme *TSER*, in DG XII. If the TSER programme is given more attention in this paper, the reason may not only be the fact that I am working for this programme in Brussels, but also the different characteristics of educational research in these two areas.

The programmes *Leonardo da Vinci* and *Socrates* are the most important Community activities in the field of education and training. While looking for a common denominator which also defines the differentiation from what is being done within the TSER programme, you could say that both programmes are action-oriented. Both are striving towards improving the education and training systems and their capacity for innovation by promoting related actions such as pilot projects, exchange measures and other activities. Both programmes, based on Articles 126 and 127 of the Treaty on European Union and Council (and Parliament) Decisions at the end of 1994 and the beginning of 1995 respectively, cover the period from 1995 to 1999.

In order to promote vocational training at the Community level, the *Leonardo da Vinci* programme has its centre in two activities: transnational pilot projects concerning for example the development of common training modules, the adaptation of content and methods of vocational training, and the training of trainers; and transnational placement and exchange programmes, addressing different groups of people involved in vocational training and giving them the opportunity to improve their work through international exchange.

But there is also a third goal which the *Leonardo da Vinci* programme is aiming at: the development of knowledge in the area of vocational training. How this activity is placed in the *Leonardo* programme and what it comprises in detail is shown in Table 1.

Table 1: Leonardo da Vinci programme

Strand III

Support for the development of language skills, knowledge and the dissemination of innovation in the field of vocational training

Strand III.2

Development of knowledge in the field of vocational training(a) Surveys and analyses

14 fields of activities, e.g.

● skill and qualification needs

● new types of vocational training apprenticeships

(b) Exchange of comparable data in the sphere of vocational training

4 fields of activities, e.g.

● the development of comparable concepts

Moving to the *Socrates* programme, it is not possible to give detailed education areas — school education and higher education as well as adult education. Included are so-called horizontal measures, such as the promotion of language learning, open and distance learning, and the exchange of information and experience. Tables 2 and 3 give an overview of what I think is interesting with regard to studies in the area of education.

Table 2: Socrates programme

Chapter III, Action 3

Exchange of information and experience on education systems and policy

1. Analysis of questions of common educational policy interest

2. European information network in the field of education(Eurydice)

3. Study visits for educational decision makers (Arion)

4. Network of National Academic Recognition Information Centres (Naric)

Table 3: Socrates programme

Chapter III, Action 3.1
Analysis of questions of common educational policy interest
Theme A: Young people without adequate qualifications/ underachievement in schools
Theme B: Evaluation of the quality of school education
First field of activities: studies and analysis
Second field of activities: operational activities
exchange of experts, study visits, colloquia, workshops, pilot projects

I am now going to present the *TSER programme*, which entails research on education and training as Area II, that is, one of three areas of the whole TSER programme. Before going into detail I would like to give a short description of how this programme started and its key objectives.

Socio-economic research including research on education and training is a new element within the research and technological development activities of the European Community. This programme was created by the Council Decision of December 15th 1994 as part of the fourth research and technological development (RTD) framework programme. It is worth quoting from this Council Decision so as to highlight the political context of the TSER programme. It is stated that 'this programme may make a significant contribution to the stimulation of growth, to the strengthening of competitiveness and to the development of employment in the Community, as indicated in the White Paper on *Growth, Competitiveness and Employment'*. The document, to which it makes reference, is the so-called Delors White Paper from 1993, the seventh chapter of which deals with the adaptation of education and vocational training systems. This chapter itself opens on to the third main part of this White Paper which is headed by Employment.

With a view to the Council's approval of the TSER programme and the related paragraphs of the TSER work programme, it becomes apparent that the TSER programme has been embedded in traditional policies of the Community. In other words, the TSER programme and, therefore, research on education and training, have to be seen in the context of the overall economic, employment

and social efforts made at Community level. Thus, educational research funded within the TSER programme is not an end in itself but has to follow a broader scope of economic and social relevance. I will return to this aspect later in relation to the present stage of implementation of the TSER programme and its Area II (research on education and training) in particular.

Now I would like to draw your attention to the structure and content of the TSER programme with special reference to Area II. Table 4 shows how the whole programme is structured and how the funding of about 100 MECU for this four-year programme is distributed. Area II has been allotted about 25 per cent of the entire budget.

Table 4: TSER is divided into 3 areas —> 105 MECU

I   Evaluation of Science and Technology Policy Options in Europe   48% of budget
II Research on Education and Training   23.5% of budget: 25 MECU
III Research into Social Integration and Exclusion in Europe   28.5% of budget

I am now going to describe the structure and content of Area II. Area II is divided into three sub-areas (Tables 5 and 6). In total, these three sub-areas include 30 specific research tasks.

Table 5: TSER - Research on education and training

Sub-areas	
Area II.1	Effectiveness of policies and actions, European dimension and diversity
Area II.2	Methods, tools and technologies: quality and innovation in education and training
Area II.3	Education, training and economic development

Table 6: TSER - Research on education and training

Fields of research
Area II.1
1. European policies in the field of education and training
2. Adaptation to change
3. European unity and diversity
Area II.2
1. Educational effectiveness
2. Innovation in education and training
3. Education, training and new technologies
Area II.3
1. Evaluation of economic needs
2. Organisation and training

For each sub-area one finds main objectives identified. In sub-area II.1, Effectiveness of policies and actions, European dimension and diversity, the two objectives are:

● to improve the understanding of the way education and training systems in Europe respond to the new and emerging needs of European society and citizens;

● to enhance understanding of the more specifically European aspects of education and training in Europe, and the various components of the European dimension in this field.

Here you will find an explanation of what 'European dimension' should mean: it is about the 'variety of approaches', the 'cultural differences', and 'common values and needs'.

Thus some of the research tasks in this sub-area concern issues such as the analysis and reformulation of educational goals in the light of anticipated

developments in society, comparative research on educational and training policies, common values in Member States' educational philosophies and the positive impact of European cultural diversity.

It is not possible to number all the research tasks listed in Area II of the TSER programme in a limited time. I would like to highlight some of the main issues, significant for and relevant to the conception of research activities within this programme. I shall concentrate on the research tasks of Sub-area 3. This area appears to be the predominant one thus far, judging by the response of the scientific community (Table 7).

Table 7: TSER - Research on education and training

Sub-area II.3: Education, training and economic development Research tasks
● Building scenarios concerning labour market demands and policy measures
● Comparative research on methods, procedures and techniques to determine labour market requirements concerning vocational training
● Comparative research on the effectiveness of labour-oriented training programmes for the long-term unemployed
● The transition from school to work: comparative research on mechanisms
● Scientific and technological literacy: research on the social and cultural aspects of the teaching of technological knowledge
● In-company training strategies and the learning organisation
● Research on the concept of the learning organisation and the way it is applied
● Conceptualisation and empirical investigation of core skills for certain branches of industry
● Comparative research on cooperation between universities and corporations aimed at the training of top specialists

I would like to conclude with some final remarks on two aspects: first, the stage of implementation of the TSER programme and, secondly, an indicative assessment of the educational research within this programme.

The programme came into force at the beginning of 1995. The first call for tenders was launched in the following March. There was an unexpected response from the scientific community. In Area II, we received about 200 proposals, out of which 12 were finally selected for funding, given the limited funds available. On 15th October this year we started the second call. The third and last call is foreseen for autumn 1997.

I mentioned above that the TSER programme in general and therefore the specific kind of basic research on education and training within the TSER programme are new elements in the European RTD framework programmes. It is not surprising that it needs some time for further consolidation, which in my opinion has been significantly developed with the ongoing second call. This recent call aims at focusing on a limited number of research tasks and stressing the originally intended policy context.

The success of a programme like TSER depends largely on how researchers in Europe respond and how they make use of the opportunities presented by this programme. Most of the possibilities are to be derived from its European character, including the precondition of any funding that research projects must be carried out in cooperation with partners of different countries. Developing a common knowledge base by comparative research and identifying good practice are to be considered as corner stones of these efforts. This constitutes the European added value. This also facilitates a great variety of approaches and should encourage interdisciplinarity as well.

I would like to finish by referring again to the Council Decision. For Area II of the TSER programme it is stated there that the objective of the Community's research activities on education and training must be to support the efforts made by the Member States to strengthen the links between research, education and training and to improve their education systems through research and disseminating good practice and innovations.

These goals are highly ambitious but they are not beyond our reach.

EURYDICE
Luce Pépin, EURYDICE European Unit

EURYDICE is not a network centred on educational research. However, in its role to support policy making and policy makers, it does share preoccupations and objectives which have been presented in this forum.

EURYDICE is an information network on education systems and policies in Europe. Its primary role is to present and observe national education policies and their common trends. Through its work, EURYDICE aims at showing as clearly as possible the diversity of education systems without hiding their many convergences. You will agree that this is an ambitious mission but an essential one for the development, strengthening and deepening of cooperation in the field of education at the European Union level. With Action III.3.1, EURYDICE is one of the key horizontal actions of the Community action programme in education. These two actions are actually two sides of the same coin because they both aim at supporting policy making at EU and national level:

● through reliable descriptive and comparative information and indicators on education systems, policies and reforms — this is the role of EURYDICE;

● through the promotion of research — this is the role of III.3.1 and of course of the Targeted Socio-Economic Research (TSER) of DG XII.

Support of policy makers' work through relevant information and research is not an easy task, whether at national or international level. Our role is to make those concerned aware of the potential for their policy making (short- and longer-term) and of the available resources in the form of both factual and comparative information and research results. The latter are doubly valuable in that they can make both a direct and an indirect contribution to the formulation of policy, by providing examples worth following and in some cases warning against some directions which have been revealed elsewhere as inconclusive. While education systems in Europe are marked by their diversity, we have to recognise that many of them are facing the same problems in practice and there could therefore be some advantage in examining solutions proposed or tried elsewhere.

What is the concrete work of EURYDICE? Based on the new orientations set in the SOCRATES programme, the activities of EURYDICE are centred on:

- **The production of descriptive and comparative analyses** on issues of common interest to policy makers at EU level. The list of publications appended clearly shows the variety of topics which we have worked on, through a comparative approach (e.g. the role of school heads; consultative councils and other forms of social participation in education; a decade of reforms in compulsory education; pre-school and primary education, etc.). Two other studies are in the pipeline: the situation of secondary education, and the role of parents in education.

- **The development of indicators** in the field of education through the production of the annual report *Key Data on Education in the European Union* for the European Commission. This has been a major component of EURYDICE's work since 1994, in this case in close cooperation with Eurostat. The originality of this report *vis-à-vis* other existing products on education indicators needs no further demonstration. While the basic statistics used are taken from the same joint UNESCO/OECD/ Eurostat questionnaire, the indicators developed in *Key Data* and their presentation acquire their particular value from the addition of and combination with the very rich and relevant qualitative information coming from EURYDICE. You will also find very interesting time series and regional graphs which add to the value of the report when thinking of policy makers' needs. This work will continue and be strengthened in the future.

- **Our Community database on education systems — EURYBASE.** This database, updated on a yearly basis by our national units, is a very rich source of reliable, basic information for all those interested in education. We will open access to this database for the wider public in the first half of 1997, through the Internet and on CD-ROM. We are pleased that you will have access to it. We feel that it is a tool of great value to the work of the research world.

All this activity is the result, not of one centralised office in Brussels, but of interactive, collective work with the Member States. The great value of EURYDICE is its network character. EURYDICE is a good example of a collective construction between the European Commission, which set up the European Unit, and the Member States, which are responsible for their national unit(s). The latter are mainly based in ministries of education. There are exceptions like England and Wales, where the ministry of education decided to base the unit, a very dynamic and efficient team, in the NFER, and Spain, where the unit is in the CIDE. So, in several cases, national units are closely linked with research centres. This is a good approach to promote operational links between information and research.

EURYDICE is at the moment covering 24 countries (15 European Union, three EFTA/EEA and six associated countries of Central and Eastern Europe) and will soon be opening its activities to the three Baltic States and to Slovenia, Cyprus and Malta. This rapid expansion (we will have gone from 12 countries in 1994 to 30 in 1997) is the real challenge of the present and the future not only for EURYDICE but for education cooperation in general at EU level and for Europe as such. We will have to make sure that we use to the full this increasing wealth of information resulting from the expansion and that we open up our comparative work to include new, and sometimes not yet stable, educational policies and practices. We have, above all, to maintain the cohesion of our network. This will be a priority.

We hope our work will become increasingly known in the future and recognised by the world of educational research as a sound basis for its investigations. We will do our utmost to secure this objective. I wish again to thank very much the Director of the NFER for having enabled us to take a first step in this direction at this Conference.

List of EURYDICE network studies. 1993-1997

1. **School heads in the European Union (*).**
   1996, 96pp. Available in DE-EN-FR. ISBN 2-87116-252-2.

2. **Consultative councils and other forms of social participation in education in the European Union (*).**
   1996, 92pp. Available in DE-EN-FR. ISBN 2-87116-250-6.

3. **A decade of reforms at compulsory education level in the European Union (*) – 1984-1994.**
   1996, 320pp. Will be available in DE-EN-ES-FR. ISBN 2-87116-255-7. To be published in the course of the first quarter of 1997.

4. **Key data on education in the European Union (*) – 1995.**
   (Produced for the European Commission, DG XXII: Education, Training and Youth). 1996, 200 pp. Available in DA-DE-EN-ES-FI-FR-GR-IT-NL-PT-SE. ISBN 92-827-5591-6. Catalogue number: C2-90-95-348-EN-C. This document is on sale at the Office for Official Publications of the European Communities and its national sales points, at the price of 25 ecus.

5. **Supplement to the study on 'Pre-school and primary education in the European Union'. The situation in Austria, Finland and Sweden and in the EFTA/EEA countries (Iceland and Norway).**
   1996, 64pp. Available in DE-EN-FR. ISBN 2-87116-234-4.

6. **Thematic bibliography: 'The European dimension in education'.**
   1996, No 1/96, 31pp. Available in EN-FR. ISBN 2-87116-233-6.

7. **Structures of the education and initial training systems in the European Union (*).** Eurydice and Cedefop. 1995, 464pp. Available in DE-EN-ES-FR-GR. ISBN 92-826-9319-8. Catalogue number CY-86-94-828-EN-C. This document is on sale at the Office for Official Publications of the European Communities and its national sales points, at the price of 20 ecus.

8. **Key data on education in the European Union 1994 (Dossier: Teaching of Languages).**
   1995, 110pp. Available in DA-DE-EN-ES-FR-GR-IT-NL-PO. ISBN 92-826-9142-X.

9. **In-service training of teachers in the European Union (*).**
   1995, 200pp. Available in DE-EN-FR-GR. ISBN 2-87116-224-7.

10. **Thematic bibliography: Teacher training.**
    1995, No 1, 22pp. Available in EN-FR. ISBN 2-87116-228-X.

11. **Organisation of school time in the European Union (*).**
    1995, 44pp. Second Edition. Available in DE-EN-FR. ISBN 2-87116-229-8.

12. **Pre-school education in the European Union. Current thinking and provision.**
    1994, 160pp. Studies No 6. Available in DE-EN-FR. ISBN 92-826-8427-X.

13. **Pre-school and primary education in the European Union.**
    1994, 120pp. Available in DE-EN-ES-FR-IT-NL-PT. ISBN 2-87116-217-4.

14. **Private / non-state education: Forms and status in the Member States of the European Community.**
    1993, 95pp. Available in DE-EN-ES-FR-IT-PT. ISBN 2-87116-192-5.

15. **Administrative and financial responsibilities for education and training in the European Community.**
    1993, 30pp. Available in DE-EN-FR-IT. ISBN 2-87116-207-7.

16. **Requirement for entry to higher education in the European Community.**
1993, 70pp. Available in DE-EN-FR-PT. ISBN 2-87116-203-4.

17. **The main systems of financial assistance for students in higher education in the European Community.**
1993, 32pp. Available in DE-EN-FR. ISBN 2-87116-198-4.

18. **The teaching of modern foreign languages in primary and secondary education in the European Community.**
1993, 92pp. Available in DE-EN-ES-FR-IT. ISBN 2-87116-194-1.

19. **Measures to combat school failure - A challenge for the construction of Europe (Summary analysis and study).**
1993, 160pp. Available in DE-EN-ES-FR-PT. ISBN 92-826-6833-9.

Distribution: As far as stocks permit, distribution will be free on a priority basis to organisations and information relays in the education sector so as to ensure a multiplier effect on the information to its widest possible target public. Most of these documents will be available on the Internet via the European Commission's Europa server (http:// europa.eu.int) in 1997.

(*) The EFTA countries are playing a full part in the activities of the Eurydice network under the Agreement on the European Economic Area and they are covered in the network's publications.

The case of IEA
Tjeerd Plomp, IEA

Introduction
The NFER Jubilee Conference provides an opportunity to discuss the mission and activities of the International Association for the Evaluation of Educational Achievement (IEA) in the context of the first results of IEA's most ambitious study, the Third International Mathematics and Science Study (TIMSS), which have just been released. Some of these results will be summarised in the next section. They will be used to illustrate the relevance of IEA's mission and the type of comparisons IEA studies allow for. Possible functions of IEA studies will then be discussed, many of them relevant for educational policy makers as well. Finally, IEA's cycle of studies will be summarised, followed by some concluding remarks.

Third International Mathematics and Science Study (TIMSS)
TIMSS is the largest and most ambitious study of comparative educational achievement ever undertaken. The TIMSS achievement testing in mathematics and science included:

- 45 countries;
- five grade levels (third, fourth, seventh, eighth and final year of secondary school);
- more than half a million students;
- testing in more than 30 different languages;
- more than 15,000 participating 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 of the way. Rigorous procedures were applied to translate the tests, and numerous regional training sessions were held in data collection and scoring procedures. Quality

control observers monitored testing sessions. The samples of students selected for testing were scrutinised according to rigorous standards designed to prevent bias and ensure comparability. This monitoring of the quality of the study resulted in the marking of countries who did not meet all the quality criteria in the tables with results.

The first results of TIMSS, namely those for the seventh and eighth grade students, have been published by the International Study Center at Boston College (USA) in Beaton *et al.* (1996a,1996b). Some of these results are summarised here: Table 1 contains achievement test results for science in the seventh and eighth grade, while Figure 1 presents the multiple comparisons for eighth grade mathematics achievement.

Table 1 and Figure 1 illustrate one of the purposes of international comparative achievement studies, namely providing 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 the 'mirror' function: countries can determine whether or not they like the picture or profile of their country as compared with other countries.

Table 1 just gives 'horse race' data, with England for eighth grade science on the tenth place and for the seventh grade on the eleventh place. Figure 1 provides more information, namely for a certain country which other countries have mean achievement scores which are significantly lower or higher than those of the comparison country, or which show no statistically significant difference. Applying the approach of Figure 1 to England results in an overview of countries which are performing significantly better/worse or statistically not different from England (see Table 2).

This type of information informs policy makers in England how well their country is doing in comparison with other countries. It shows also that league tables like Table 1 contain limited information and may result in misleading interpretations, as it does not reflect any statistical information. Figure 1, which does reflect this type of information, shows not only that England is not really performing better than the USA or worse than Germany, but also that EU partners like Ireland, Belgium, The Netherlands and France are performing significantly better.

However, information in tables and figures like these does not help policy makers, curriculum developers and educational practitioners to understand *why* their educational system is performing as it does — for example, why England is performing poorer than many of its EU partners. This leads to the

Table 1: TIMSS — average achievement in science

Eighth Grade		Seventh Grade	
Singapore	607	Singapore	545
Czech Republic	574	Korea	535
Japan	571	Czech Republic	533
Korea	565	Japan	531
Bulgaria	565	*Bulgaria*	531
Netherlands	560	*Slovenia*	530
Slovenia	560	Belgium (Fl)	529
Austria	558	*Austria*	519
Hungary	554	Hungary	518
England	552	*Netherlands*	517
Belgium (Fl)	550	England	512
Australia	545	Slovak Republic	510
Slovak Republic	544	United States	508
Russian Federation	538	*Australia*	504
Ireland	538	*Germany*	499
Sweden	535	Canada	499
United States	534	Hong Kong	495
Germany	531	Ireland	495
Canada	531	*Thailand*	493
Norway	527	Sweden	488
New Zealand	525	Russian Federation	484
Thailand	525	Switzerland	484
Israel	524	Norway	483
Hong Kong	522	New Zealand	481
Switzerland	522	Spain	477
Scotland	517	Scotland	468
Spain	517	Iceland	462
France	498	*Romania*	452
Greece	497	France	451
Iceland	494	*Greece*	449
Romania	486	Belgium (Fr)	442
Latvia (LSS)	485	*Denmark*	439
Portugal	480	Iran, Islamic Rep.	436
Denmark	478	Latvia (LSS)	435
Lithuania	476	Portugal	428
Belgium (Fr)	471	Cyprus	420
Iran, Islamic Rep.	470	Lithuania	403
Cyprus	463	*Colombia*	387
Kuwait	430	*South Africa*	317
Colombia	411		
South Africa	326		

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

Figure 1. Multiple comparisons of mathematics achievement — upper grade (eighth grade)

Read *across* the row for a country listed horizontally to compare performance with the countries listed vertically. Countries are ordered by mean achievement.

	Mean achievement significantly higher		No statistically significant difference		Mean achievement significantly lower
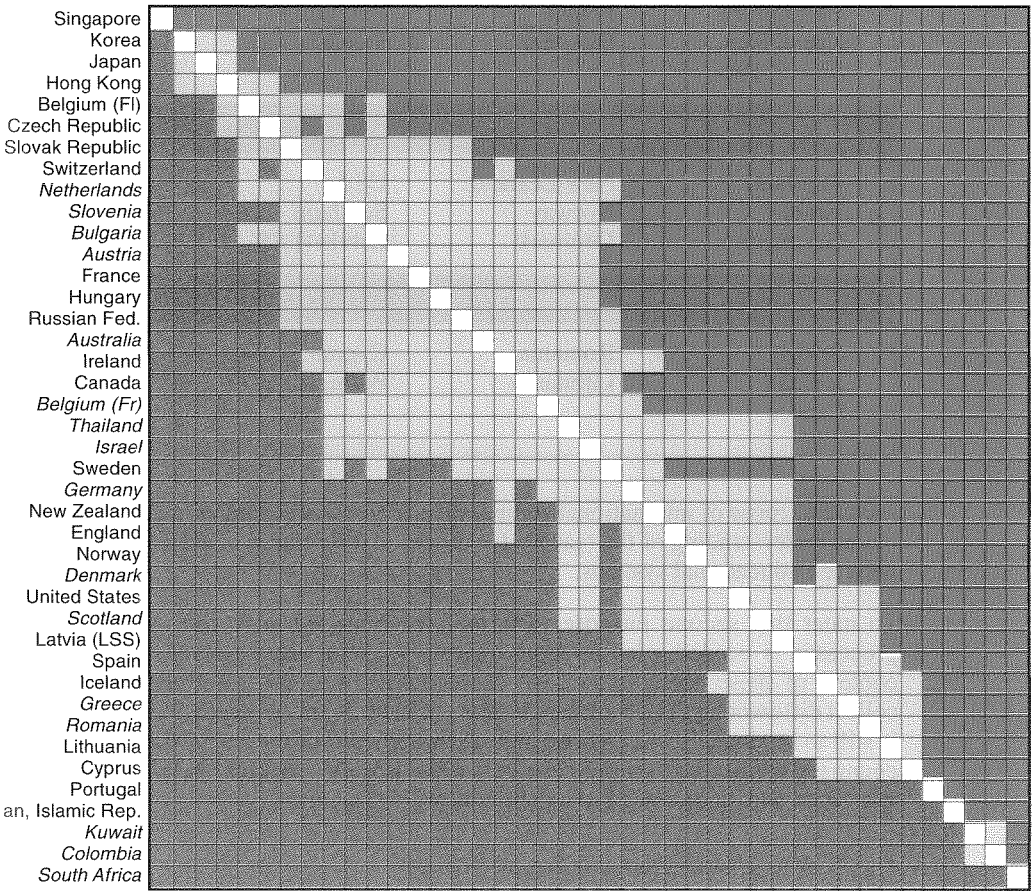
Because coverage falls below 65%, Latvia is annotated LSS for Latvian Speaking Schools only.

Countries shown in italics did not satisfy one or more guidelines for sample participation rates.

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

Table 2: TIMSS — mathematics: England vs. other countries

Significantly higher achievement:		
Singapore	Switzerland	Russian Fed.
Korea	Netherlands	Australia
Japan	Slovenia	Ireland
Hong Kong	Austria	Canada
Belgium Fl.	France	Belgium Fr.
Czech Rep.	Hungary	Sweden
Slovak Rep.		
No significant difference:		
Thailand	New Zealand	USA
Israel	Norway	Scotland
Germany	Denmark	Latvia (LSS)
Significantly lower achievement :		
Spain	Romania	Cyprus
Iceland	Lithuania	Portugal
Greece		

second purpose of international comparative achievement studies, namely to get an understanding of the reasons for observed differences in achievement between educational systems.

To meet this second purpose, information about learning and teaching processes is needed, plus in-depth analysis of achievement results in the context of these background data. IEA international comparative studies collect different kinds of background data. In TIMSS, this in-depth of analysis is still to be done. IEA considers this an important task for the participating countries themselves, as they can best articulate the research and analysis questions which are relevant to their education systems.

IEA: what it is and does: mission, history

IEA, the International Association for the Evaluation of Educational Achievement, is the organisation that conducts international comparative studies in which educational achievement is assessed in the context of process

and input variables. It has developed over 37 years as a cooperative of research institutes representing at present 53 educational systems (see Husen and Postlethwaite, 1996, for a concise description of the history of IEA). Many countries are represented nowadays in the IEA General Assembly by policy makers. Member institutes are often the most prominent ones in their country; some of them are part of their Ministry of Education, others are linked to universities or are independent research centres.

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

Next to this, IEA also conducts other studies which are not curriculum-based. Examples are the Pre-Primary Project and the Computers in Education study, of which a successor (the Second Information Technology in Education Study) is under preparation. See the IEA Guidebook (1993) for more detailed information about IEA and the studies.

As mentioned above, IEA recognises two purposes of international comparative achievement studies: (i) to provide policy makers and educational practice with information about the quality of their education in relation to relevant reference groups; and (ii) to assist in understanding the reasons for observed differences between educational systems (which serves policy makers' needs, but is clearly a researchers' interest).

In its studies, IEA strives for two kinds of comparisons. The first one consists of straight international comparisons of effects of education in terms of scores (or subscores) on international tests, as is illustrated for TIMSS in Table 1 and Figure 1. The second kind of comparison 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, but such analyses can be done in an international comparative approach.

A typical IEA study deals with grade levels in three populations: elementary education, junior secondary education and senior secondary education.

IEA was founded as a research cooperative. Initially, it was primarily interested in international comparative studies from a research perspective. But in the second half of the 1980s IEA started to recognise the increased interest of policy makers in educational indicators. Since then, IEA has taken it as a challenge also to serve the interests of policy makers through its studies. The inclusion of IEA achievement indicators in publications of OECD is an illustration that IEA has started to become successful in this way. The last edition of OECD's *Education at a Glance* (1996) presents a number of indicators based on the TIMSS results. Examples of IEA publications which address relevant policy questions are Postlethwaite and Ross (1994) and Keeves (1996); another relevant source is Kellaghan (1996).

Not every study should have a size and a design as comprehensive as TIMSS. But IEA has a strong belief that the conceptualisation of its studies (which is summarised at the end of the next section) allows for designing studies which meet the needs of policy makers.

Functions of IEA studies

As already stated above, the relevance of IEA studies reaches much further than just making straight comparisons in the form of league tables. Some more reasons for the importance of international comparative achievement studies (and of educational indicators) can be mentioned.

(i) *Description*

To provide policy makers and the education community with information about the status of 'their' educational system: this is considered by many already interesting in itself. In the past, IEA was not interested in 'league tables' or 'horse race data'; for example, in the Second International Mathematics Study (SIMS), IEA never reported such tables. But IEA has now recognised that this kind of information is not only of great interest for policy makers, but also a good starting point for generating questions for in-depth analysis ('mirror' function). The broad interest world-wide in the TIMSS results illustrates the relevance of this function.

(ii) Benchmarking

This function can best be illustrated with an example. Within TIMSS, some Asian countries and, within Europe, Belgium Flemish and the Czech Republic have the highest test scores for mathematics. If another country is interested in improving its education in mathematics, it can analyse its own case against the case of the Asian countries and/

or the European countries on many variables, related to curricular aspects of mathematics education (including curricular materials), pedagogical approaches and instructional processes, school variables, teacher background, teacher training (and in-service training). Such analyses may result in proposals for change, although no easy answers should be expected. For such countries, an important question in a next IEA study would be whether it is then performing closer to the reference countries chosen.

(iii) Monitoring quality of education

One step further than benchmarking is monitoring: the regular assessing of educational processes on different levels in the educational 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, that is, 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).

(iv) Understanding reasons for observed differences

Policy makers may want to understand differences between or within educational systems from the prespective of national policy making (this function should be distinguished from the next one: cross-national research). This function is again one step further than just collecting data for monitoring purposes: it serves ultimately policy makers' needs, but is clearly also a researchers' interest. A nice example is the analysis in the USA of the data of the IEA Second International Mathematics Study (SIMS) resulting in a monograph, *The Underachieving Curriculum* (McKnight *et al.*, 1989). Again, easy answers should not be expected here on what measures should be taken to improve education in a country. But this kind of research may lead to policy decisions about changes in education ('informed decision making'), or to initiatives like in the USA, where the National Council for the Teaching of Mathematics developed the well known standards for the teaching of mathematics.

(v) Cross-national research

This function refers to exploratory and/or in-depth research of the IEA databases. Many examples can be found in the IEA volumes. Here only two other examples are mentioned. Postlethwaite and Ross (1994) did an exploratory study of the IEA Reading Literacy database in an

effort to find indicators discriminating between more effective and less effective schools (in reading). The second example is Keeves' (1996) monograph, *The World of School Learning: Selected Findings from 35 Years of IEA Research,* in which he discusses on the basis of all IEA studies conducted up to 1994 ten key findings with suggested implications for educational planning.

Almost all of the functions mentioned above need measurement of educational achievement and other outcomes of education on three levels of the education system:

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

IEA studies are adressing all three levels by distinguishing three aspects of the curriculum:

● intended curriculum: what should be taught and learned — measured by analysing documents such as official syllabuses, course outlines, text books;

● implemented curriculum: what actually is being taught or taking place in schools and classroom: content, time allocations, instructional strategies, etc. — to be measured through questionnaires (or observations);

● attained curriculum: what students attain or learn in terms of cognitive skills, attitudes, etc. — to be measured through tests.

For more information about the conceptual approach of IEA, see for example Robitaille and Garden (1989), Plomp (1992) or the IEA Guidebook (1993).

In a typical IEA study, many activities have to be completed to provide data and indicators of good quality. These include: 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.

Cycle of studies

It is IEA's policy to have a core cycle of studies, in which certain basic subject areas are studied regularly. The present cycle contains reading literacy, mathematics and science. Knowing that governmental organisations like OECD have an interest in the same areas, IEA has expressed its willingness to attune its studies to the wishes and plans of OECD.

IEA's present studies are:

- Third International Mathematics and Science Study (TIMSS), of which the results of the seventh and eighth grades have been released in November 1996; in June 1997 the results of the third and fourth grade levels will be published, while the results of the final grade of secondary education and performance testing (at the elementary and junior secondary levels) will follow later in 1997.

- Language of Education Study of which Phase 1 resulted in a monograph with country profiles of the teaching of languages (Dickson and Cumming, 1996); Phase 2, the achievement testing, is on hold due to a lack of interest among policy makers resulting in insufficient budgets for conducting this phase.

- Second Civics Education Study is now (December 1996) in its first phase, the development of country profiles, with Phase 2, achievement testing, in preparation.

Another study, different in scope, is the Pre-Primary Project, a study of policies and practices in child care, while a new Reading Literacy Study and a second Information Technology in Education Study are under preparation.

An important feature of IEA studies is the training of National Research Coordinators (NRCs). In most cases this is an essential component of the study, as many NRCs appear to be unfamiliar with the methodology and especially the specifics of international comparative studies. A special project was the creation of the IEA Network in Central and Eastern Europe (IEANCEE), which provided a training programme for the National Research Coordinators of IEA studies in this group of countries, most of which did not have any tradition in IEA-type research.

Concluding remarks

A few concluding remarks will be made, to emphasise IEA's present orientation in planning and conducting its studies.

- IEA started as a cooperative of researchers and developed into a cooperative of research institutes. IEA was initially driven by the interests of researchers, of course inspired by the presumed relevance of their studies for policy making. Since the mid-1980s IEA clearly recognises and values the interest of policy makers in its research, and now takes that as one of the starting points of its studies.

- IEA's only agenda is that it wants to contribute through its studies to enhancing the quality of education. It provides a network of institutes and of individuals, which together represent much experience and intellectual capacity. In that way it is a meeting place for policy makers, educators, scientists and researchers.

- The functions of international comparative studies as described above make clear that IEA needs to have a cycle of 'core' studies which provide trend data. Such an approach is not only necessary for monitoring the quality of education, but allows also the study of many interesting research questions.

- IEA is world-wide, and can therefore serve the needs of the international community represented in organisations like UNESCO and the World Bank. But next to that, IEA is also ready to serve the needs of groups of countries organised in organisations like OECD, European Union or the Asia–Pacific Economic Co-operation Group (APEC). Being broader than any such organisations, IEA can link through its studies the interests of member countries of such organisations with its world-wide membership.

References

BEATON, A. E., MARTIN, M. O., MULLIS, I. V. S., GONZALEZ, E. J., SMITH, T. A. and KELLY, D. L. (1996a). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study.* Chestnut Hill (MA, USA): Center for the Study of Testing, Evaluation and Educational Policy, Boston College.

BEATON, A. E., MULLIS, I. V. S., MARTIN, M. O., GONZALEZ, E. J., KELLY, D. L. and SMITH, T. A. (1996b). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study.* Chestnut Hill (MA, USA): Center for the Study of Testing, Evaluation and Educational Policy, Boston College.

DICKSON, P. and CUMMING, A. (Eds) (1996). *Profiles of Language Education in 25 Countries: Overview of Phase 1 of the IEA Language Education Study.* Slough: NFER.

HUSEN, T. and POSTLETHWAITE, T. N. (1996). 'A brief history of the International Association for the Evaluation of Educational Achievement (IEA)', *Assessment in Education,* **3**, 2, 129-41.

IEA GUIDEBOOK (1993). Amsterdam: IEA.

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,* **3**, 2, 143-60.

McKNIGHT, C. C. *et al.* (1989). *The Underachieving Curriculum: Assessing US School Mathematics from an International Perspective.* Champaign (IL, USA): Stipes Publishing Company.

OECD (1996). *Education at a Glance: OECD Indicators* (4th edn). Paris: OECD/CERI.

PLOMP, Tj. (1992). 'Conceptualizing a comparative educational research framework', *Prospects,* **XXII**, 3, 278-88.

POSTLETHWAITE, T. N. and ROSS, K. (1994). *Effective Schools in Reading: Implications for Educational Planners.* Amsterdam: IEA.

ROBITAILLE, D. F. (Ed) 1993. *Curriculum Frameworks for Mathematics and Science.* TIMSS Monograph No.1. Vancouver: Pacific Educational Press.

ROBITAILLE, D. F. and GARDEN, R. (Eds) (1989). *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics.* Oxford: Pergamon Press.

For more information about IEA:

The IEA Secretariat
Herengracht 487
1017 BT Amsterdam
tel: +31 20 625 3625; fax: +31 20 420 7136
e-mail:Department@IEA.nl or 101776.2527@compuserve.com

UNESCO's role in international educational research
John Smyth, UNESCO

This year (1996) is also the fiftieth anniversary year of the United Nations Educational, Scientific and Cultural Organisation, and it may be recalled that in the NFER's early days some of the distinguished educators associated with the NFER were also associated with UNESCO. Thus it is especially a pleasure for someone from UNESCO to participate in the NFER's jubilee conference.

You may want to know first why UNESCO should have any role at all in international educational research. Therefore, let me begin with a few words on UNESCO's overall purposes and mandate, quoting from its Constitution.

UNESCO was founded on the idea that '... a peace based exclusively upon the political and economic arrangements of governments would not be a peace which could secure the unanimous, lasting and sincere support of the peoples of the world, and that [it] must therefore be founded, if it is not to fail, upon the intellectual and moral solidarity of mankind'. UNESCO's founders, therefore, 'believing in full and equal opportunities for education for all, in the unrestricted pursuit of objective truth and in the free exchange of ideas and knowledge ... agreed ... to develop and to increase the means of communication between their peoples and to employ these means for the purposes of mutual understanding and a truer and more perfect knowledge of each other's lives'. In consequence they decided to create UNESCO, specifically for 'the purpose of advancing, through the educational and scientific and cultural relations of the peoples of the world, the objectives of international peace and of the common welfare of mankind for which the United Nations Organization was established and its charter proclaims'.

UNESCO, therefore, is not an educational research organisation. Nor does it exist for the purposes of encouraging international educational research as such. It is mandated to encourage international cooperation in educational research only in so far as such cooperation helps to advance 'the objectives of international peace and of the common welfare of mankind'. However, it can do this only in so far as its 185 or so Member States can reach a consensus on the contents and modalities of a programme.

For most of its 50 years' existence, UNESCO was handicapped by two sets of disagreements among Member States: those resulting directly from the cold war, and those consequent on the great disparities in development between the different regions of the world. As regards the cold war, some of the

audience present today may remember the difficulties it caused for educational cooperation, including cooperation in educational research. The history of the IEA is illustrative: originally founded with UNESCO's encouragement in the 1960s — its first home was at the UNESCO Institute of Education in Hamburg — it was regarded from the beginning with suspicion by the countries of the Soviet bloc, and with the exception of one or two mavericks like Hungary, these countries never joined the IEA studies of the 1960s, 1970s and 1980s. Since the end of the cold war of course, especially in the case of TIMSS (Third International Mathematics and Science Study), the countries of East and Central Europe, not to speak of Russia itself, have shown a very active interest in cross-national educational research of all kinds, not only the studies carried out under the IEA's auspices, but also a great number of bilateral comparative studies on topics ranging from educational management and financing to civic education and the teaching of human rights and democracy.

Still, the cold war is now history and there is not much point today in belabouring the opportunities for international cooperation in educational research that were lost. More to the point is the second set of constraints which I mentioned earlier: those relating to the global disparities in development, in particular disparities in the provision of what the Preamble to UNESCO's Constitution referred to as 'education for all'. It is to this second set of constraints, and their implications for international cooperation in educational research, that my remaining remarks are directed.

First, I shall briefly mention why, and in what way, the global imbalances in the provision of 'education for all' have held back the development of international cooperation in educational research, and then I shall briefly describe how this situation is beginning to change. Finally, I shall consider how this change is beginning to impact on UNESCO, and what could be the possible implications for UNESCO in the future.

In an ideal world, one where there existed that 'intellectual and moral solidarity of mankind' referred to in UNESCO's Constitution, there would be a lot of international cooperation in educational research. In reality, though, we are far from attaining such solidarity. The disparities in education between the different peoples of the world are too great. The countries represented here today account for barely 10–15 per cent of the world's population. Yet around a quarter of the world's adults — over 800 million persons, two-thirds of whom are women — are illiterate. Moreover, although the percentage of the world's adults who are illiterate is slowly declining, the absolute number of illiterate adults in the world, because of population growth, is virtually

197

stationary. Many of the world's formal school systems are in such poor condition, lacking teachers and appropriate teaching materials of all kinds, that they can barely stem the flow of young people into the ranks of illiterate adults. In any case, over 130 million children of school-going age (6–11) are estimated today to be out of school. The situation is especially grave in those countries which the United Nations has classified as 'Least Developed' — 47 or so countries mostly located in Sub-Saharan Africa and the poorer parts of Asia, Latin America and the Caribbean. In twice that number of countries — 90 or so, roughly half the membership of the United Nations — it is still the case that probably a majority of children barely receive more than four years of basic formal education, the minimum which could be said to be necessary for acquiring permanent literacy and elementary numeracy.

For a very large number of UNESCO's Member States, therefore, in fact probably the majority, the main challenges for educational policy are relatively straightforward. They are challenges, simply, of sheer provision: of providing at least some kind of education for a majority, if not yet all, of their youngsters. For most of these countries, educational research is a luxury; at least, up to now it has been largely considered to be so by national educational policy makers. During the cold war especially, policy makers in a majority of developing countries were suspicious of the urgings made in favour of increased educational research by their colleagues in the industrial countries. Such urgings were often regarded simply as insensitive, or worse, as an effort to divert countries from their real needs and keep them in a subordinate position in the global economic and social order. It is hardly surprising, therefore, that during the 1960s, 1970s and 1980s, UNESCO's programme made very little provision for international cooperation in educational research.

Such cooperation was not wholly neglected though: in the 1980s it came to be regarded mainly as a challenge for countries in the South, with the industrial countries of the North considered in part as having educational research interests and priorities that could not be legitimately applied in the countries of the South. UNESCO therefore set up several regional networks for cooperation in educational research, in Africa, the Arab States, Asia and the Pacific, Latin America and the Caribbean. In the Europe region, which for UNESCO's purposes included the United States, Canada and Israel, a formal network as such was not established, but instead UNESCO joined with the Council of Europe in sponsoring periodic conferences on educational research, each time focusing on a particular theme or topic: for example, in 1990 (International Literacy Year) the theme was 'literacy and functional literacy in the industrial countries'. Besides the regional networks, UNESCO also initiated programmes to provide training in educational planning, especially

for officials in developing countries. In that connection, for example, the International Institute for Educational Planning (IIEP) was established under UNESCO's auspices in 1963. In recent years, the IIEP as well as UNESCO's other institutes (the International Bureau of Education, Geneva, and the UNESCO Institute of Education, Hamburg) have been quite active — though on a modest scale — in trying to stimulate educational research in some of UNESCO's poorer Member States.

The record of success for UNESCO's South–South networks has been mixed. In general, however, the setting up of the networks did establish an important precedent, namely, that inter-governmental mechanisms for cooperation in formulating common goals and areas of interest in educational research could function at all. The most successful network has probably been APEID (the Asia–Pacific Network of Cooperation for Educational Innovation and Development). CARNEID (the Caribbean Network of Cooperation for Educational Innovation and Development) and a similar programme in Latin America also have been quite successful. The corresponding programmes in the Arab States and Africa regions (there is some overlap) have never really taken off. Adequate and steady financing has been a problem for the networks almost from the beginning, partly because they came into existence at a time when UNESCO entered a period of grave difficulties over its budget with the withdrawal from the Organisation of the United States, United Kingdom and Singapore. Generous extra-budgetary financing by Japan has been an important factor in the success of the Asia–Pacific network. In contrast, the African network (NEIDA — Network for Educational Innovation and Development in Africa), after an initial impulse of funding from UNDP (United Nations Development Programme), has had no major extra-budgetary sponsor and in consequence today really only exists in name.

Probably, the networks were premature. Except for a few countries in Asia and a handful of countries in Latin America and the Caribbean, the majority of network members simply did not give educational research much priority in their own national budgets to enable their few researchers to participate effectively in the networks. That being said, there are some signs now that the situation could be changing. National educational policy makers in the South, especially since the holding of the World Conference on Education for All (Thailand, 1990) — and one could mention too the report of the South Commission (1990) chaired by Julius Nyerere — are revising their interests and priorities, and are beginning to take a more positive view of the value of educational research and of possible international cooperation in this area.

Education is currently high on national policy agendas in most countries, whether they are poor countries still struggling to provide a basic education for all, or advanced industrial countries increasingly anxious about their ability to compete in the emerging global economy. National policy makers are looking more intensively at their education systems. There is widespread concern over the quality of education. This concern is manifested in many ways: e.g. debates about 'standards', reconsideration of school curricula, new arrangements or procedures for student testing and examinations, innovations in teacher education, and new approaches to the management and financing of education. In some countries, e.g. Eastern Europe and the Russian Federation, radical reforms are under way.

National policy makers are also looking outward. Even in the largest countries which have a wide range of different educational approaches within their borders there is increasing interest in the approaches followed by other countries. Policy makers are asking, What are the best practices — in regard to effective teaching and pedagogy, assessment strategies, approaches to equity and fairness in educating ethnic minorities and so on? One consequence, among others, has been pressure on the existing international educational information system in which UNESCO plays a critical role. How effective is this system? Can it respond adequately to the growing demand for information from national policy makers? What are the prospects of it being able to respond better in the future? What needs to be done in order to enable it to respond better? Let me close my presentation by trying to answer these questions.

The demand is for both quantitative and qualitative information, including of course quantitative information on the quality of education. The whole question of international education statistics has come to the fore, prompted in part by the experience gained among industrial countries participating in OECD's INES (Indicators of National Education Systems) Project, in particular the difficulties they have encountered in making valid international comparisons of educational performance, and in part by the experience gained among developing countries in following up on the World Conference on Education for All, in particular the difficulties which many countries have experienced in monitoring effectively their progress towards the goals agreed to at that conference. UNESCO has been engaged in the collection, collation, presentation and publication of international education statistics virtually since the date that the Organization came into being, but in recent years this programme has languished; the case for strengthening it has been highlighted in a review carried out last year (1995) at UNESCO's request by the Board

on International Comparative Studies in Education of the US National Academy of Sciences.

International education statistics are not of course the same thing as international educational research. Nevertheless, it is not too far-fetched to say that it is policy makers' needs for improved international education statistics, in particular comparative statistical indicators of their education systems' characteristics and performances (whether the focus is on inputs, process or outcomes), that is beginning to drive an increasing interest in the possibilities of international cooperation in educational research. The outstanding example of this interest is the use which is now being made, by both UNESCO and OECD, of the findings of the international research on student achievement carried out by the IEA (International Association for the Evaluation of Educational Achievement).

Future international cooperation in educational research is likely to take its lead from policy makers' needs for improved international education statistics and indicators. What is the current situation in that respect? There are three major challenges or clusters of challenges to be overcome.

First, there is unbalanced development of educational information systems between countries. Broadly speaking, those countries which still have a long way to go towards the goal of 'education for all' are also those which have the weakest information base. This weakness is first of all felt by the countries themselves in terms of their own needs for management and planning of their educational development. (Many countries have approached UNESCO for assistance in improving their education statistics. The Organization, in collaboration with extra-budgetary sources, currently has a modest programme for providing such assistance in selected African countries.)

Secondly, existing international classification systems for education statistics are out of date. While hardly any two educational systems in the world are the same, and most are significantly different from each other, such differences need not be a bar to the development of international education indicators or, more generally, to the communication and exchange of educational information and experience, if there exist appropriate internationally agreed definitions and systems of classification. The International Standard Classification of Education (ISCED), which was developed by UNESCO in the early 1970s, is a crucial instrument for ensuring international comparability in education statistics, but it needs updating and revision especially in the areas of post-compulsory and higher education. (This is now under way in

UNESCO, with the cooperation of OECD and EUROSTAT as well as experts from Member States.)

Thirdly, existing education statistics (both national and international) provide unbalanced coverage of education itself. Existing international education statistics are mainly demographic. The World Conference on Education for All (Thailand, 1990) stressed the importance of effective education and learning achievement, yet in a majority of countries, both developed and developing, only limited statistical information is available on the educational process and even less on the outcomes of education. The financing of education is another area where international statistics are weak.

Overcoming these challenges or constraints will require much international cooperation. Still, the international climate today is probably more favourable to such cooperation than at any other time in the last 50 years. In UNESCO's global perspective, much careful thought needs to be given to devising cooperative mechanisms that could overcome the first set of challenges just mentioned: those relating to the present imbalances between countries in their educational information systems and research capacities. In other areas outside of education, for example, agriculture and health, successful mechanisms have been devised which could serve as useful models for cooperation in educational research: for example, the Consultative Group on International Agriculture Research (CGIAR), the Special Program for Research and Training in Tropical Diseases (TDR), the Onchocerciasis (River-blindness) Control Program (OCP), and the Global Programme on AIDS (GPA). Each of these mechanisms is basically an association of countries, international organisations and foundations designed to match funding to developing countries' research needs. The United Nations Specialized Agencies involved are FAO and WHO, along with the World Bank, which typically acts as the main channel of financing. It surely would be possible to establish a similar association for educational research involving UNESCO.

Let me close by mentioning two closely linked factors which are likely to be critical for the success of a possible future CGIAR-type association in the area of educational research. First, it will be vital that the research priorities, subject of course to their feasibility, be seen to be established largely by the countries involved, and not just by the funding agencies based in the industrial countries. Secondly, there will need to be, at least initially, substantial provision for cooperation between mature educational research institutions in the industrial countries on the one hand, and newly established institutions in the developing countries on the other, in training the latter's young researchers.

This brings me, finally, to mention our hosts, the NFER, one of the world's largest and most experienced educational research bodies. Surely we may hope that in the NFER's second half-century, its experience will be contributed in some form or other to assisting young researchers in institutions in poor countries that now are starting out on their own roads of educational research and inquiry.

Reference

SOUTH COMMISSION (1990). *The Challenge to the South.* Oxford: Oxford University Press.

Educational research and evaluation: the World Bank's role[1]

Maris O'Rourke, World Bank

Why the World Bank is interested in educational research and evaluation

Since 1963, the Bank has lent US$23.3 billion for 560 education projects in 112 countries, and Bank lending for education stands at ten per cent of all its lending in FY96, up from four per cent a decade ago. As the single largest source of external financing for education in the world, the World Bank is able to better realise its development objectives by being a *consumer, facilitator and producer* of educational research and evaluation.

The four main objectives of the Bank's research programme are:

- to support all aspects of Bank operation, including the assessment of development progress in member countries;
- to broaden understanding of the development process among Bank staff, member countries, and the academic and development community;
- to improve the Bank's capacity to give policy advice to its members by incorporating internal and external research and evaluation; and
- to assist in developing indigenous research capacity in member countries through its lending operations.

The World Bank and educational research and evaluation: three roles

The World Bank's role with respect to research and evaluation takes three forms:

- *consumer*: Bank staff learn from all sources of knowledge, including research facilitated and produced through Bank activities;
- *facilitator*: Bank staff operate in a 'hands off' manner to promote production of research through Bank lending, the Special Grants Programme, and consultants hired through Trust Funds to assist borrowers prepare projects;
- *producer*: Bank staff carry out and/or supervise production of Original Research, Economic and Sector Work (ESW) or Policy Work.

Increasingly, the Bank is shifting its emphasis from being a producer of research and evaluation to being a facilitator and consumer of educational research and evaluation. Significant amounts of resources and staff are devoted to building the Knowledge Management Network, facilitating research studies and sharing information with clients.

Educational research and evaluation help create more effective lending operations

The Knowledge Management Network of the Bank depends upon Original Research, Economic and Sector Work (ESW) and Policy Studies produced by the Bank; experience gathered from successful project design and implementation; client-conducted research facilitated by the Bank; and externally conducted research.

The Knowledge Management Network harvests and synthesises research from these sources for the realisation of the Bank's development objectives, design and implementation of successful projects, and design and implementation of effective policy.

Lessons from Bank experience in countries and sectors are fed back into the Knowledge Management Network to enhance the quality of the Bank's operations. The Knowledge Management Network also leads to the production and facilitation of new research through identification of knowledge gaps.

The World Bank as a consumer of educational research and evaluation

There are four main areas where the Bank is a consumer:

- the Knowledge Network;
- the Bank's international training wing — the Economic Development Institute (EDI);
- the development of policy advice and guidance; and
- direct Bank lending operations, where staff help borrowing countries design and implement education projects, policies and programmes.

The Education Sector Knowledge Management is part of the Bank's Knowledge Management System. The aim of the Human Development Network is to maintain and enhance the internal capacity of the Bank in terms of people, quality of operations, sector strategy and knowledge.

The strategy of the Knowledge Management System is to move towards *a*

culture of information sharing, publishing, knowledge reuse, and knowledge development and generation. High priority will be given to the identification of 'knowledge gaps'.

The first set of Strategic Thematic Areas in Education are:

- Technology in Education
- Economics of Education
- Effective Schools
- Access and Equity in Education
- Post-Basic Education
- Early Childhood Development
- Strategies for Educational Change
- Education Project Design/Delivery.

Tools include:

- Web pages/best practices/ideas/examples
- Help desk
- Indicators/databases.

The Bank has made a significant commitment to increase its ability to be an informed consumer of educational research and evaluation in FY97, in both Education Sector Knowledge Management (US$3.5 million — 22 per cent of the proposed Human Development Network budget) and EDI's education sector work (from US$1.5 million in FY96 to US$3.0 million in FY97). All EDI's activities are largely research-based: knowledge is disseminated to clients, including academic and member-country counterparts.

The World Bank as a facilitator of educational research and evaluation

The Bank facilitates educational research and evaluation with three major instruments:[2]

- *The Special Grants Programme (SGP)* of the World Bank supports partnerships with other donor agencies in key areas of development research. The World Bank contributes up to 15 per cent of the total research requirement and the partners contribute the remainder. In the past, the SGP has supported a limited number of education projects, but this will change for FY97.

- *Trust Funds* included in this analysis are the project preparation trust funds (PHRD TFs) and the Consultant Trust Funds (CTFs). Both facilities complement the World Bank's own resources for technical assistance activities, pre-investment studies and project preparation assistance to support the lending programmes.

- *Bank-Funded Education Projects* represent ten per cent of all Bank lending. These projects fund educational development in the following sub-sectors: primary; general secondary; vocational secondary; vocational post-secondary; teacher training; and higher education.

The largest share of support for facilitating educational research and evaluation comes from World Bank support to clients through its lending operations. In fact, the share of Bank-funded education projects with research components has nearly doubled in the 1990s, from an average of 24 per cent for FY86-90 to an average of 48 per cent for FY91-95.

The World Bank as a producer of educational research and evaluation

The Bank has also had a modest role as a producer of educational research and evaluation, in three areas:

- *Original Research* This comprises all analytical work at the Bank designed to produce results with wide applicability across countries and sectors. The studies are directed toward recognised and emerging policy issues and are focused on yielding better policy advice. Most are carried out in the Development Economics Vice Presidency (DEC), Policy and Human Resources Division of the Policy Research Department (PRDPH).

- *Economic and Sector Work (ESW)* This is a combination of original research and synthesis, adapting the product of research to specific country and sector settings.

- *Policy Studies* These synthesise the product of research studies and adapt it to specific projects and country settings in order to design successful projects and effective policy.

Bank-generated research has made an enormous contribution to educational literature. Some recent contributions include work on the importance of textbooks in determining educational outcomes, on high rates of return (especially private) to education, measured either as wage differentials or as farmer productivity, on private/public provision of education, and on cost-effectiveness of various school inputs.

Examples of Original Research

The Impact of Women's Education on Fertility and Contraceptive Use: a Study of 14 Sub-Saharan African Countries (Martha Ainsworth, Kathleen Beegle and Andrew Nyamete, 1996).

This study examines the role of female schooling in determining levels of fertility and contraceptive use in 14 countries. High fertility and the high demand for children are reflected in Sub-Saharan Africa's rapid rate of population growth, at 3.2 per cent a year. Recent surveys reveal that women's 'ideal family size' ranges from six to nine children, despite important advances in the provision of family planning and the implementation of population control policies in most countries. This paper examines the role of female schooling in determining levels of fertility and contraceptive use. The research results show that increasing female schooling will have a significant impact on the lowering of fertility and the spread of contraceptive use.

Impact Evaluation of Education Projects: Research in Decentralisation and Privatisation Issues (Elizabeth King and researchers from PRDPH, SA1PH, HDD, LAT *et al.,* 1997).

This project, jointly conducted with operational staff, uses experimental and quasi-experimental impact evaluation methods to examine four Bank-funded projects (in Colombia, Pakistan, Peru and Tanzania) that change the way central governments allocate financial planning and educational decision-making power to local governments, communities and schools. The methodology allows the Poverty and Human Resources Department of the Policy Research Department (PRDPH) to address policy areas that have been relatively unexplored empirically, including implementing voucher schemes, fostering decentralised decision making, inducting private school supply and increasing managerial efficiency. Baseline school and household data from control and treatment populations have been collected and are being analysed, and follow-up surveys to ensure accurate evaluations of the effectiveness of these projects are in preparation.

Example of Economic and Sector Work

Primary Education in India: Achievements and Challenges (Marlaine Lockheed) is the product of the World Bank's first sector work on basic education in India. It results from a lengthy, unique, highly sensitive and fully collaborative process involving the Bank, the Government of India, key Indian research institutions and internationally known scholars in India.

With agreement from the Government of India, the process was undertaken along two tracks, with scholarly papers written by Indian authors presented at national and international seminars and published by Indian institutes, and an open and transparent synthesis of this research prepared by the Bank. Thus, the process resulted in a significant degree of client ownership.

Although the study initially set out to research key issues facing basic education (as support for the preparation of the first District Primary Education Project (DPEP)), the research grew dramatically as new questions emerged over two years. Issues of research included: local constraints of school quality; special barriers faced by girls and disadvantaged groups; and baseline indicators for access, attendance and achievement in order to measure progress, learn from experience and have a basis for shifting resources from areas of low performance to areas of high potential. In addition, ways to strengthen key elements of the education system — teachers, textbooks, management and financing — needed to be identified so that steps could be taken to ensure that reform efforts would be effective.

The Bank's sector work synthesised research undertaken to prepare for the DPEP, and commissioned special studies covering areas for which new information was needed. Ultimately, 46 new studies, in addition to the district and state-level studies commissioned as part of the DPEP preparation, were completed involving dozens of Indian research scholars.

Example of Policy Work

Priorities and Strategies for Education: a World Bank Review by Nick Burnett (1995) reviews the literature and data on the contribution of education to development and on the financing of education in developing countries. It outlines policy options and the following key reforms for increasing access to education and for improving equity and efficiency in educational services:

- *A higher priority for education.* Global economic and political liberalisation combined with technological change makes education increasingly important for economic development and social cohesion.

- *Greater concern with results.* Instead of focusing so heavily on the supply of education, more attention should be paid to the quality of learning and its impact on preparing people for employment and society.

- *More emphasis on basic education.* Within the education sector, free primary and early secondary should be a priority for public spending.

- *Greater attention to equity.* More needs to be done to increase the enrolment of girls, the rural poor, linguistic minorities, street children.

- *Greater parental involvement.* Since it is parents who take decisions about enrolling their children much more parental and community involvement is needed.

- *More local autonomy.* Schools, universities and other educational institutions should be able to make their own decisions about how to allocate resources.

The relative importance of Original Research, Economic and Sector Work, and Policy Work in the education sector

Of the three instruments for producing research and evaluation, operationally relevant Economic and Sector Work is the most significant. In FY96, ESW comprised the largest share of research and evaluation produced in the Bank in the education sector, with approximately 43 per cent of all research funding allocated to ESW.

The number of ESW reports completed annually parallels that of the number of education projects completed annually. However, the 1970s showed an increase in the number of ESW reports while the 1990s showed a decline relative to the number of projects. Thus, although the total number of education projects is increasing, the ESW supporting these projects has declined.

Future directions for the Bank's role in educational research and evaluation

In all sectors, the Bank will assume a smaller role as **producer** of research. Expenditures on the production of *all types of research and evaluation* have decreased by approximately five per cent (almost US$10 million) between FY93 and FY95. Support for *Original Research* has declined from US$33.8 million in FY93 to US$26.1 million in FY95, a decrease of almost 23 per cent. Support for *Economic and Sector Work* (ESW) has also decreased from US$126.8 million in FY93 to US$109.9 million in FY95, a decrease of almost 13 per cent. Expenditure on *Policy Work*, however, increased by nearly 50 per cent from FY93 to FY94 and stayed stable in FY95.

For education, support for Original Research is expected to increase slightly, while ESW and Policy Work are expected to decline.

The Bank is expanding its role as **facilitator** of research and evaluation through the establishment of the International Programme for the Improvement of Educational Outcomes, which will be supported by the Bank's Special Grants Programme and other donors; and through the Education Knowledge Management System, which will identify areas where research is needed and facilitate knowledge development and generation in those areas.

Most importantly, as a research **consumer**, the Bank is increasing its role dramatically — through a new Human Development Network which includes a new Knowledge Management System and through an expanded Economic Development Institute (EDI).

At the 1996 Annual Meeting, Mr James D. Wolfensohn, President of the World Bank, said: 'We will work with others to build a global knowledge partnership, in which the Bank Group can play the role of "connector" — to capture and disseminate development expertise and experience for our clients and partners'.

So in addition to being a development institution and a financial institution, the World Bank is also a *knowledge-based institution*. The Bank has lending and non-lending services and accompanies its financial assistance with advice. This advice is based on its experience of working in more than 140 countries and in virtually all sectors of national economies.

According to the *Report on the World Bank Research Program: FY94 & FY95*, 'in many ways, the real value-added of the Bank is based on its ideas, knowledge and experience — increasingly true in a world of expanding global markets'.

---

[1] Based on presentation prepared by Marlaine Lockheed, Maris O'Rourke and Ayesha Vawda.

[2] Additional sources employed for educational research facilitation are not listed because the data are not readily available.

7.

Learning from other areas of study

Maurice Kogan, Centre for the Evaluation of Public Policy and Practice, Brunel University

I was asked to discuss how far educational research can learn from adjacent areas of inquiry. This will be a partial and inexpert contribution; the difficulties of the task arise partly from the fact that it is difficult to grasp what is going on in any one area of study, let alone try to make comparisons with subjects outside. In mainstream education, studies range from curriculum analysis and learning theory to the political science of governing bodies. Research on schools and on higher education pursue quite different agendas and the research communities are different. As we come to major divisions of subject area — and I will dwell largely on health and education — we find different power and funding structures which are intimately related to the dominance of particular forms of expertise. To take my subject seriously we would need to create a matrix with the substantive areas on one axis and common characteristics of organisation, or knowledge concerned with what is actually delivered to client groups, on the other. How is a lesson given in a school to be compared with a medical prescription or nursing care or a social work intervention? The differences as much as similarities would stand out. As I will argue, it is hazardous to believe that education can ever simulate the power that medical science has with medical practitioners.

This is, however, a theme that deserves serious attention. Studies that converge on one ultimate theme, the human condition, become framed by specialised perspectives. Specialisation is certainly necessary for the handling of mightily complex materials, but each area could yield wisdom to others if some of us looked to the broader scholarship of more than one area. I cannot in any way lay claim to that ambition, but some funder, and some really bright scholar, ought to work on it.

It is a subject entailing fascinating and deep issues about what conditions the nature of different forms of knowledge. In part it must be the nature of the phenomena being studied. Human psychology deals with quite different objects from those researched in metallurgy. Part of the difference lies in the extent of control of variables that is possible, and the regularity and predictability of the characteristics of that being observed. In part, too, the formation of knowledge arises from social dimensions such as its objectives and users.

I will first recite my conclusion and then try to find some facts to support it. It is as follows: research in medical and health sciences differs from that in other personal social services, such as education or social work, because it starts from a hard and determinant knowledge base, a biological scientific base, which can be translated into practice. (Economics has a similar appeal.) But work in the health field is not only about the human biological system but about the systems for allocation and delivery of health care and in these aspects is not far different from studies in other areas of human concern. The interesting point is, however, that whilst such studies are not dissimilar in substantive focus from many undertaken in education, the power of the biomedical model extends to them.

In pursuing that theme, and relating it to educational concerns, I will consider the production and use of knowledge in different areas by noting four dimensions of commonality and differentiation. They are:

- the knowledge base and the knowledge rules, or epistemology, of the different areas;

- the degrees of applicability to practice;

- the extent to which policy makers concern themselves with the sponsoring, objectives-setting and use of research, and the differences in power exercised by different science communities;

- some particular examples from evaluation studies.

My main areas for comparison will be education, higher education and health, although I will take a sideways glance at research on social services.

Nature of the knowledge

First, the nature of the knowledge. Here the concerns are broadly epistemic — the rules governing the acceptability of knowledge. By what kind of criteria are the statements derived from research thought to be true? The nature of knowledge in the different areas can be placed on a spectrum. It ranges from that which is capable of being subjected to experimental testing and/or randomised controls, through that which may disclose an internal logic, or which may seem to be right, but which also clearly depends upon its context of social construction. On that spectrum, near the hard end, one has the collection and ordering of data within a positivistic mode, so that, for example, the Chief Inspector of Schools can firmly inform us that England is behind in reading ability compared with Singapore. By contrast, social anthropological

studies are essentially directed to explaining and giving an account rather than proving replicable truths. Or research or other more conjectural forms of inquiry, linked with values analysis, might discern the normative models of nonconformity among our young people that could sustain a live democracy.

The conversion of subjectivity into objective and verifiable statements is what we are all about, but some subjects have a harder time than others. The potency of a statement in our field need not depend on the rigour with which it is formulated. Take, for example, the fact that for most of the time the press and politicians will say that standards in schools are dropping. That fact is testable empirically by processes that first set up criteria of success or failure, determine the extent to which the criteria are to be subject to normative adjustment according to contexts, and then go in and measure on some time series. But even if not proven in that way, the statement takes on a truth quality by virtue of the fact that it is widely believed and thus conditions political actions. It becomes what Cohen and Lindblom (1979) call Ordinary Knowledge which is then converted into Usable Knowledge. The social scientist attempting to evaluate standards may then have to take it as datum that there is popular belief that standards are dropping because the belief in itself affects teacher behaviour and their expectations of pupils. Standards may thus be affected by it. It is like the 'unreal' market in economics where sentiment in the City will affect the price of stock and the exchange rates despite what might be happening to production or exports.

A not dissimilar set of issues arises in dental studies; Aubrey Sheiham and others (see the literature reviewed in Rosenoer and Sheiham, 1995) have established there is poor association between reported satisfaction with teeth and the number of standing teeth. Whilst orthodox dentistry tends to recommend the replacement of most missing teeth, whether with teeth or porcelain, 'subjectively perceived needs are important determinants in assessing requirements for the prosthetic replacement of missing teeth'. Even among those at the lowest levels of oral heath status (15 missing teeth or deep periodontal pockets), 40 per cent of individuals still felt that their oral health was excellent (Reisene and Baillitt, 1980). The normative figure with which patients may be satisfied is a shortened dental arch of about 20 teeth (Kayser, 1981, 1984). What will constitute, then, a true statement of dental effectiveness?

Obviously, this problem of truth emerges in different forms in different subject areas. Some can emancipate themselves from it by deliberately divorcing the phenomenon from its social context and keeping a tight grip over the range of variables involved.

I now will exemplify the knowledge range by looking at the widest of subject areas, that is, the knowledge base of health studies. These range from biomedical science to studies of the delivery of health services. They can be divided into several sectors which I will boil down to three:

A. Scientifically based studies of the human biological system and diseases that might attack it and procedures for tackling or preventing diseases or clinical studies.

B. Studies of social and economic contexts, of lifestyle or cultural influences on health.

C. The organisation and economics of health.

There are several other areas such as the study of the health professions, their recruitment, training and careers which could be added to this list. I will allude to health education later on. Let us look, however, at these main blocks:

A. *Knowledge about human biological mechanisms, diseases and cures* is firmly tied into science producing and testing systems. The mode of knowledge generation is that of the scientist starting with quite fundamental scientific problems, such as those associated with molecular biology, DNA and 'the book of life', or the identification of viruses and their pathological impacts, and hence their potential cures.

These studies are controlled in laboratories or in scientifically controlled clinics and the results are capable of being replicated if the conditions of the experiments remain the same. Therapies are tested through double-blind, randomised, controlled tests.

We have no analogous certainties in educational research. Presumably the nearest that one gets to deriving practice propositions from research is in learning theory, but that would not claim to be rigorously predictive. We do not see anybody, at least anybody with scientific credibility, with a prescription pad, although the air is constantly full of prophecy.

Nor are the kind of studies of behaviour that underpin theories of social work capable of being more than strongly suggestive. A sophisticated legal attack (Mnookin, 1975), from California, of course, noted that whilst the law gives social workers enormous determinate powers in generating adoptions or fostering or taking into care, the scientific base of their judgements is indeterminate. Here we can note how the epistemic issues interact with pragmatic issues and issues of empowerment. Hard science is convincing.

But soft science may be treated as determinate if socially insisted policies, such as effective social control, are being reinforced.

This hard mode of scientific inquiry generates power in a way not to be found further along the spectrum. It is esoteric in procedures and its language is not accessible to the laity. Yet it leads to a capacity to make desirable therapeutic interventions in the human body. It thus generates power and some protection from political interference not to be found in other areas of study.

B. Alongside this 'hard' scientific base, there is work drawing upon *studies of social and economic contexts, of lifestyle or cultural influences on health* that go from the individual to public health states. The most influential of these are related to epidemiological studies and rely on the organisation of samples and control methodologies. The classic studies of Doll on the link between smoking and lung cancer, and the many epidemiological studies in cardiology, are obvious examples. The power of cross-national sampling is exemplified by studies which are able to associate genetic links with cancer or other diseases. These are different from clinically or laboratory-based studies which seek to establish the mechanisms which govern health states. Instead they establish powerful inferential links between social or lifestyle factors or genetics and disease. In principle this range of studies follows fairly closely the models of physical and biological science, in its demands for sampling and replicability.

Here education draws closer to health studies, at least in terms of substantive content and methodology. We can know the distribution of measured ability and educational performance in any population and we can, and have, related these to socio-economic backgrounds. This kind of social epidemiology has been used positivistically, as in the social engineering variety of policy making prevalent from the 1950s to the 1970s. The health studies, however, also have additional power, deriving perhaps from individual concern for the avoidance of disease, and from the scientific form of the operative statements. Thus knowledge about the factors making for lung cancer or heart disease has become didactically important and has caused individuals to change their lifestyles.

In these areas, however, of both health and education, there is always the possibility of challenge in that the modes of enumeration may incorporate social constructions, or that in the use of such knowledge indicators may become vindicators. The challenge has been sharper in education than in health because the health studies seem to have a greater power to predict than do the equivalent in education.

C. There are studies of *health organisation, funding and the like.* In the UK at least, government has been prepared to spend quite a deal of money on them. Since its inception in 1945, the British National Health Service has undergone a whole procession of changes in organisational structure, and this has often been accompanied by ministry-financed research projects of the likely effects and consequences of organisational change. Examples are some of the kinds of studies in which my own and neighbouring groups have been involved. These included an evaluation of an attempt to improve organisation and delivery through management advisory services; the critical analysis of medical audit schemes; an evaluation of a Total Quality Management pilot scheme in the National Health Service; a developmental project on clinical audit in four professions.

Here there is comfortable common ground. The study of educational systems has generated whole tax-free salary earning industries in Paris and Brussels. The studies of school organisation and their impact on educational procedures, from Elizabeth Richardson onwards, are also commonplace. They are not far different in objectives and style from the considerable work on internal structures of health organisation or the formidable research on the delivery of nursing or other health profession care. The difference is not one of knowledge style but of funding and sponsorship; I will return to that point later.

The first of these modes of knowledge generation is strongly related to the world of scientifically based practice. The medical school is the admired meeting point of much of the research and of practice. The production of knowledge occurs through scientific procedures in scientifically based medical schools or units and laboratories or the pharmaceutical industry. The process of mediating knowledge among medical and health scientists of all types is well remarked and strong. Medical and health scientists have strong networks. Knowledge is mediated largely through journals and academic and practitioner conferences, exchanges of protocols and collaboration, some of it taking place through IT networks. All of these can be found in not only the major specialties but also in quite detailed sub-specialties.

Another important feature of medical and health research is its multidisciplinarity and search for cross-cutting themes. In a recent unpublished study, I noted nearly universal demonstrations of how in specific areas of health (e.g. cardiology, cancer and neurology), whilst clinical services are driven by largely organ-related symptoms, medical science is multidisciplinary in terms of methodologies and techniques and relies on considerable association with neighbouring areas of research. Thus work on

single organs involves association with work on other organs and recruits help from a wide range of discipline-based areas such as epidemiology, physics, molecular biology, genetics and basic biochemistry.

British medical science culture enhances cross-cutting attitudes because medical scientists also practice general medicine, unlike many of their continental counterparts. They are thus likely to avoid narrow monodisciplinary or single-organ approaches. It is claimed that multidisciplinarity is fully accepted as an essential frame of reference. The nineteenth century view of disciplines such as physiology or anatomy is not a good starting point for solving problems. It is necessary to define a problem and put a team together from different areas to solve it. There is still merit in the older structures, but problems require a wide group of expertise. There is a need to pull in a group of people who must then give up their allegiance to wherever they came from originally. I do not know how far educational studies actively seek out equivalent multidisciplinary approaches to particular problems.

Applicability to practice

In health studies, the relationship between knowledge generation and use is in theory unproblematic. Much new knowledge is, of course, strongly mediated through the commercial selling systems of the drug houses and the rest. David Hargreaves (1996) has noted the differences between doctors and teachers in their reliance on research findings, although the commitment of doctors to keeping up with the science should not be exaggerated.

Knowledge use is a salient characteristic of the health area at all levels, at the practitioner level as well as at the systems direction levels; there are pressures and rewards to ensure that practitioners take notice of the new knowledge. Knowledge of new drugs or pathologies or modes of treatment is deemed essential to good medical functioning. In education, the burden of proof, as to usefulness, is on the researcher. In the UK there is at present a campaign within the medical profession for the strengthening of evidence-based practice; in comparison with education or social work or housing policy, it can be said to be already there, and is, indeed, sometimes criticised for its certainty as well as for its inadequacies. There are strong and publicly funded Centres for Evidence-Based Medicine and for Review and Dissemination. The reviews undertaken by the latter range from health prevention and education to specific treatment areas. A similar aspiration is presumably behind the Teacher Training Agency's initiative for teacher-based research, but keeping up with

research findings is hardly written into the regular practice of teachers or social workers. I do not think this can be ascribed to practitioner idleness. There is simply no analogy in education or even social work to the intervention in one's personal physical system attempted by prescribing a drug or any clinical intervention. Social workers make or initiate decisions that profoundly affect people's lives, but see how contested they are, because they do not derive from a determinate scientific base. Where knowledge is uncertain, value preferences become more powerful.

Teacher knowledge is far more diffuse than that of the health practitioner, although context and environment are increasingly important in those areas too. Teaching and learning are affected by a far wider range of variables, and many of them are incommensurate with each other. There are large problems of control and predictability. The gap between what teachers know and what they can do is wide. There has to be strong reliance on the development of practical wisdom, and its knowledge characteristics are not far different from those of, say, the professional historian whose techniques are not so much technical but involve an ability to interpret events and link ascertainable facts to broad contexts. The problem of transferability of this kind of wisdom–knowledge is huge. Yet the need for disciplined inquiry is equally huge.

Since writing this paper I have had the benefit of David Hargreaves' critique which is based on his own recent research on the training of medical practitioners. He rightly points out that the training given to junior doctors is strongly concerned with instilling craft skills, and that medical practice largely depends on the successful application of skills and techniques generated by experience and practice. 'Clinical practice has an ultimate foundation in the natural sciences, which feed into medical and clinical sciences. These do not, however, feed directly into clinical practice but are mediated by medical craft knowledge ... the favoured practices of one's peers; and one's own experience, much of which is acquired by trial and error' (Hargreaves, below).

This is an important and secure perspective which has the advantage of pointing the way to fruitful comparisons between the development of medical and teacher practice. It does not, however, impair the point being made here which is that the power to inspire confidence depends not only on manifest competences created by experience — which are important in the formation of all scientific workers, including the most abstract of theoreticians — but also the possession of esoteric and testable knowledge. Using the battery of drugs, or applying diagnostic procedures, including those taken for granted such as taking blood pressure or tests for cholesterol, are not an exploitation of individual experience but of scientific development. David Hargreaves

would be correct in saying that experience would guide the doctor in determining when to administer the test or how seriously to take the results. Perhaps that knowledge becomes implicit and taken for granted in much the same way as the powerful always deny the extent of their power. It is, however, clearly present as a differentiator between professional and client in medicine in a way not obvious in the relationship between teacher and pupil or parent, and also is an important component of status differentiations within medical hierarchies.

The differences in power exercised by different science communities

The more 'scientifically' based forms of health knowledge are firmly in the hands of the medical scientists who largely set their own agendas on research councils and other funding bodies. Although scientists complain about the general shortage of funds, in the main policy makers let the medical scientific community get on with their own research agenda even though there have been recurrent attempts to create policy frames, and to avoid the duplication of research effort and gaps in funding. In the UK, in the 1970s the government attempted to create a customer (government)–contractor (scientist) relationship (the Rothschild experiment) for the production of clinical and health organisation knowledge (Kogan and Henkel, 1983). This failed, although attempts are being made to create variants of it now.

More recently, the Culyer Report (1994) proposed the creation of a single stream budget for R&D and a central R&D Committee to set a strategic framework and advise on priorities. The financing of medical research is, to some extent, affected by political priorities. AIDS and mad cow disease research can be rapidly promoted by government demand, under pressure often from the media. The scientists complain that the EU puts its resources behind what medical scientists regard as fashionable areas, rather than those dictated by scientific development.

As far as health organisational studies are concerned, funding is considerable. For example, the impressive growth in nursing studies has been stimulated largely by government sponsoring through the health departments. The NHS has been more able and willing to fund evaluations of its organisational changes than have the education and higher education systems. Government has done little to assess the effects of, say, the binary system or its demise, or the growth of managerialism in universities or the impact of the so-called education reform acts. Whilst the NHS has funded evaluations of audit and

quality assurance, I know of no attempt to evaluate the impact of OFSTED on children's learning or teachers' teaching.

The reason for the differences in government funding of research on organisational and service-delivery systems given to us by one knowledgeable insider is that doctors in the NHS and the academic systems would not willingly contemplate a system in which key issues are not subject to research, and they are powerful within the Department of Health. It would be fair, however, if sad, to say that whilst these works are referred to by people in quality assurance work and presumably are used by some managers in health trusts and other health organisations, the impact on policy has been small. Here is the judgment of a former Permanent Secretary: 'I know of no strategic issue with which Ministers were concerned which was illuminated by the Health Services Research Programme' (Stowe, 1989). This judgement, made perhaps by someone at far too high a level to know the full impact of research programmes, can now be modified by the results of recent studies on the payback of research in health and social services, to which I will return later.

The reasons for the lack of impact take me a bit outside my remit, but, briefly, it has been noted that in most cases the officials commissioning the research have moved on, and there is no continuity of reception of the findings of research that they had commissioned; evaluative schemes created by the Department of Health and Social Security have been poorly conceptualised; and there are conflicting policies at work within single government departments. This 'tribalism' had been noted in our earlier study (Kogan and Henkel, 1983). Yet in the attempt in the 1980s to install the Rothschild customer–contractor system, an attempt which ultimately failed because of Medical Research Council opposition, it was seen as possible to get scientists and policy makers together (in research liaison groups) to make a study of the current pattern of research, the gaps in research that are required for practice and policy, and to commission work directed to filling the gaps.

Much work on health evaluation could be read over into the other personal services. Studies of organisation are likely to use similar constructs of the distribution of power, of working relationships, of the texturing of structures according to the tasks being tackled. As one moves away from strictly clinical intervention studies, of, say, the effects of particular drugs or surgical procedures, so the scientific base becomes more indeterminate and context-related, but there are similar problems that might yield to similar conceptualisations and research approaches.

But there is a division between medical scientists and the others in that, for

example, some medically based workers consider an evaluation to take place only when controls have been used. Some studies of social work (Goldberg, 1984) asserted, against the long-standing social anthropological model (e.g. Abrams, 1984), which sought to explain rather than assess the outcomes of interventions, the importance of establishing tight criteria in advance of evaluation and arriving at judgements on the micro-effects of interventions. This was pursuing the classic model of evaluative research. Another division in the broad field of medical audit is that doctors emphasise output measures whilst the other professions, more involved in generic aspects of treatment and treatment systems, still concentrate on the processual dimension of outcomes.

To some extent, there is movement towards the medical model in education. There is an acceptance of, or at least an acquiescence in, the quantitative dimension of evaluation, and the emphasis on outputs; OFSTED has created methods by which the performance of both pupils and teachers can be rated and put into time series. Our own Higher Education Funding Council too has commissioned work on the impact of research assessments that will not stop at asking academics how it was for them, but will track the difference in resource use in departments in relation to the outcomes of the research assessment exercises; the data for these are available.

Some possible examples to follow

So far I have given arguments demonstrating the difficulty of following examples from other fields. Let me turn to two sets of studies in health and social services which may provide more encouraging examples from other areas. They centre on some rigorous UK studies of impacts of research under the heading of payback. There has been a cloudburst of applied research on this issue financed by the National Health Executive and encouraged by the Economic and Social Research Council.

From Buxton and Hanney's study (1994), which is based on eight cases, we can draw examples of both an evaluative frame for the outcomes of research and at least one example of useful methodology about outputs that seems to me to go well beyond the somewhat limited output methodologies established in education.

Their studies of payback of research were primarily based on economics modelling, but a multidisciplinary team undertook the case studies. It is a valuable exemplar of how to go about studying who put what into and who

got what, and how, from research. The range of subjects is wide, from heart transplants to social service care management.

They classify payback into five categories:

● knowledge benefits

● research benefits

● political administrative benefits

● health and social service sector benefits

● broader economic benefits.

Let me briefly give some examples of some of these categories, and note how far they apply to education studies. Knowledge benefits can be measured through the familiar techniques of patents analysis and bibliometrics. The research benefits include the better targeting of future research, the development of research capacity and a critical capacity to utilise existing research. The political and administrative benefits include creating a better information base for policy formation, and legitimation of policies if they take account of research findings. The benefits to the health and social service sector include cost reduction, qualitative improvements in service, increased effectiveness and improved allocation of resources. The broader economic benefits may include commercial exploitation of innovations and the benefits of a healthy work force.

Some, but not all, of these paybacks could be transferred to analysis of research in education. Some are measurable, whilst others must depend upon qualitative studies with key informants or client groups. The authors of this work do not underestimate the difficulty of making it usable. I simply report that they have done so in this different field. Their eight case studies showed that a number of the researches had a directly significant impact on policy and executive decisions. Others became part of a body of evidence leading to clinical guidance. But much of the research was largely ignored in the policy debate. Several factors were associated with high payback:

● continuing support from customers

● liaison with stakeholders

● appropriateness and quality of research

● brokerage

- appropriate dissemination

- ongoing programmes in their own right and as a context for specific projects.

This study thus gives some relief to educational researchers whose scientific base is not determinate. It shows, as we inferred in our OECD Report (1995), that the social and organisational arrangements for research have an important effect on ultimate outcomes. The payback area of analysis could be directly transferred to education. So could perhaps some of the research models which were analysed in the study.

One case study, undertaken by my colleague Mary Henkel (1994), of the Kent Community Care Project in which the Personal Social Services Research Unit at the University of Kent cooperated with the local authority, showed how an experimental evaluative project on community care for the frail elderly had high payback over the long term and made a major contribution to community care policy and, potentially, practice, at local, national and international levels and has earned a significant academic reputation. The case study examined how far an input–output model can be used to identify factors that affect payback at various stages. The design was conceived as an experiment in micro-budgeting: an application of arguments about the provision of authority, responsibility and accountability and the building in of incentives to match resources better to needs. The design was tried out in two further studies.

My general point here is that rigorous work in other areas may well be generating methods that could at least be tried out in cognate educational areas. My other point is that these studies are demonstrating a combination of quantitative and qualitative evaluation which takes them outside a narrow inspectorial base for the forming of judgements.

A common agenda?

There are, then, certainly common themes and methodological issues running through these diverse policy and practice fields. Health research has the advantages of the determinate scientific base, although that may also restrict its reflexivity in areas where it is desirable, and of directly and usefully addressing conditions from which all of us seek relief. As researchers move away from the substantive technical base towards systems and institutions, the possibility of borrowing techniques and concepts becomes more feasible. This would not constitute a common agenda, but it does suggest that some

systematic raiding parties on other areas of study to establish common issues and research approaches might be rewarding.

These initial ruminations also invite thought about more systematic study of connections between fields. It could start with a taxonomy of human needs — some are already available in the different fields — and wants which would identify both unities and conflicts. From there it could compare a selection of different areas of research and scholarship directed to the analysis and meeting of those needs. The analysis of knowledge and its use would identify the different values, technologies, organisational structures, client groups and power bases of the different knowledge sectors. It could then attempt to show how far these differences conflow into holistic views of human needs and the ways of meeting them. Such a project would depend less on new research than on the creative application of scholarship to what we already know.

References

ABRAMS, P. (1984). 'Evaluating soft findings: some problems in measuring informal care', *Research Policy and Planning*, **2**, 2.

BUXTON, M. and HANNEY, S. (1994). *Assessing Payback from Department of Health Research and Development*. Uxbridge: Health Economics Research Group, Brunel University.

COHEN, D. and LINDBLOM, C. E. (1979). *Usable Knowledge*. New Haven, Conn.: Yale University Press.

CULYER REPORT (1994). *Supporting Research and Development in the NHS*. Report prepared by the NHS Task Force, chaired by Professor Anthony Culyer. London: HMSO.

GOLDBERG, E. M. (1984). 'Evaluation studies: past experience and possible future directions', *Research Policy and Planning*, **2**, 1.

HARGREAVES, D. (1996). *Teaching as a Research-Based Profession: Possibilities and Prospects*. London: Teacher Training Agency.

HENKEL, M. (1994). 'Case Study Two: PSSRU research in care management: Kent Community Care Project.' In: BUXTON, M. and HANNEY, S. (op. cit.).

KAYSER, A. F. (1981). 'Shortened dental arches and oral function', *Journal of Oral Rehabilitation*, **22**, 469-80.

KAYSER, A. F. (1984). 'The minimum number of teeth to satisfy functional and social demands.' In: FRANDSEN, A. (Ed) *Public Health Aspects of Periodontal Disease*, p. 135. Chicago: Quintessence Books.

KOGAN M. and HENKEL, M. (1983). *Government and Research: the Rothschild Experiment in a Government Department*. London: Heinemann.

MNOOKIN, R. H. (1975). 'Child-custody adjudication: judicial functions in the face of indeterminacy', *Law and Contemporary Problems*, **7**, 3, 226-93.

OECD (1995). *Educational Research and Development: Trends, Issues and Challenges* (edited by TUIJNMAN, A. and KOGAN, M.). Paris: OECD.

REISENE, S. T. and BAILLITT, H. L. (1980). 'Clinical oral health status and adult perceptions of oral health', *Social and Medicine*, **14**, 597.

ROSENOER, L. M. and SHEIHAM, A. (1995). 'Dental impacts on daily satisfaction with teeth in relation to dental status in adults', *Journal of Oral Rehabilitation*, **22**, 469-80.

STOWE, K. (1989). *On Caring for the National Health*. Nuffield Provincial Hospital Trust.

Respondents' comments

David H. Hargreaves, *University of Cambridge*

Professor Kogan's paper, of which I saw a draft version, deals with the topic across a large canvas that reflects his substantial intellectual range and the breadth of many years' experience. This response is narrower in scope, limited to a small portion of that canvas. These remarks reflect my current work in the field of medical education, where I am seeking to improve the quality of postgraduate on-the-job training of doctors in hospital. In this country, doctors who, after initial qualification, wish to become specialists have hitherto undergone a lengthy training of 12 years or more under the supervision of a consultant before they achieve that status as an autonomous specialist. Naturally, to that work I take what I have learnt about teacher education and educational research; and from my experience in medical education, I am taking back ideas to teacher education and educational research. Thus far the reception among doctors to the ideas I have carried from education has been warmer than that of the educationists to whom I have brought ideas from medicine.[1]

Four questions arising from Kogan's paper (and linking to Seamus Hegarty's opening address) are of considerable interest in any comparative study of the medical and teaching professions. All concern the knowledge base that underpins professional practice.

1. *What kinds of knowledge base underpin professional practice?*

    We have no reason to assume that the knowledge bases of different professions are structurally very similar.

2. *How does research contribute to the knowledge base of a profession?*

    Whilst relevant research might reasonably be treated as one component in a professional knowledge base, its relations to other components might well vary between professions.

3. *How does the knowledge base contribute to professional practice?*

    Since the pioneering work of Herbert Simon and Donald Schön, we know this to be an enormously complex issue.

4. *Through a comparative analysis, can one profession learn from another to improve the quality and effectiveness of professional practice?*

    This I take to be an important question within the broad scope of Kogan's analysis. Early in his paper, he recites his conclusion, which is that medical research differs from educational research because 'it starts from a hard and determinant biological scientific base which can

be translated into practice'. I am not convinced that this is strictly accurate. It may well *start* from such a basis in the foundation natural sciences, but as knowledge moves through clinical science to clinical practice, the 'hard and determinant' elements are compounded by many softer ones. This is not, of course, to deny that the role of science in the knowledge base in education compared with medicine differs from and bears a different relation to professional practice. I sketch, in highly simplified form, selected aspects of the knowledge base of doctors in hospitals and contrast it with that of teachers in schools. (A continuous line indicates a strong relationship and a dotted one a weak relationship.)

The knowledge-base in the teaching and medical professions:
Some simplified schematic relationships

Figure 1 — Medicine
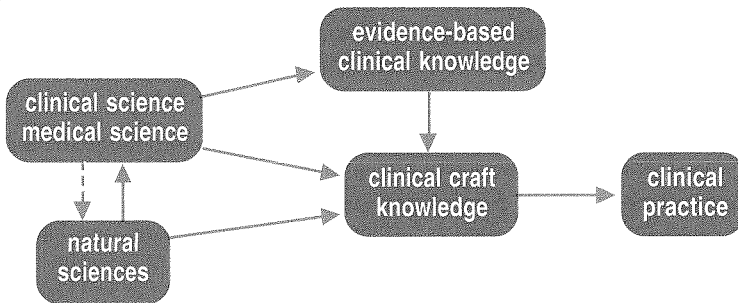
Clinical practice (Figure 1) has an ultimate foundation in the natural sciences, which feed into medical and clinical sciences. These do not, however, feed directly into clinical practice but are mediated by medical craft knowledge, which consists of the traditions of a doctor's specialty; the ways in which one was trained; the favoured practices of one's peers; and one's own experience, much of which is acquired by trial and error. This last component, trial and error, is probably an essential element in the acquisition of any professional skill; but in the case of medicine, its importance and salience are carefully protected from public scrutiny in the interests of patient confidence. The long training of medical specialists involves the acquisition of an extensive amount of clinical science; but it also includes the acquisition of clinical craft knowledge which is far from being wholly rooted in 'hard and determinant biological science'. The scientific knowledge base is thus not

directly translated into clinical practice, but is heavily modified as it is filtered through clinical craft knowledge.

There is often a weak scientific or research base to help the medical practitioner decide precisely what therapeutic treatment or management technique to use with a particular type of patient in particular circumstances and for what scientific reason. During the two years I have spent working with doctors in clinical settings in hospitals, I have been deeply surprised at how often clinical decision making is grounded in craft knowledge, not research-based science. Like other professionals, including teachers, doctors are interested in what works in what circumstances — thus the growth in evidence-based medicine which seeks a scientific, not craft, answer to the 'what works?' question.

In recent times, evidence-based medicine is becoming a component of doctors' knowledge base. This requires (i) that practitioners – researchers investigate the impact and effects of different treatments on different kinds of patient in different circumstances, (ii) that such evidence is collated and made readily available to other practitioners, who (iii) take account of it when relevant to a clinical decision concerning a unique patient. (Wynne Harlen's paper provides an interesting account of some aspects of evidence-based medicine in Scotland.)

Of course generalisations that emerge from evidence-based medicine are not close cousins of the universal generalisations so important in the basic natural sciences: they are limited and highly qualified generalisations, which get established, refined or qualified as further research accumulates on what works in what circumstances. It will be a long, slow process to develop a more substantial evidence base. As Henry Miller, Dean of Medicine and later Vice-Chancellor of the University of Newcastle expressed it: 'The fact is that medicine consists of a few well-lit islands of scientific certitude surrounded by a boundless ocean of uncertainty and ignorance.'[2]

Figure 2 — Education
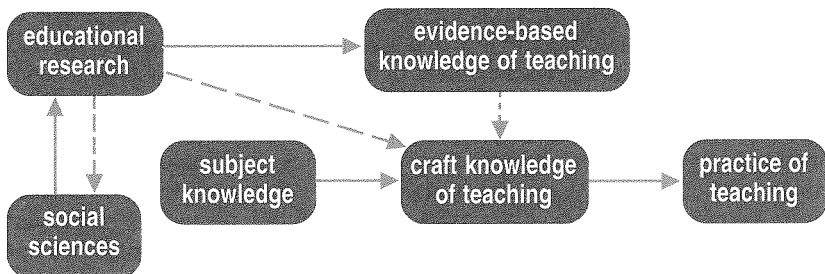
The knowledge base of teachers (Figure 2) has different components and different relations between those components. The main component is the subject the teacher teaches — mathematics, geography, literature, etc. This is mediated through craft knowledge into the practice of teaching. The craft knowledge is acquired through many processes, but there is a heavy reliance on personal experience and trial-and-error learning. The social sciences make very little impact — at least in any direct or explicit way — on craft knowledge; indeed, the social sciences play a less salient role than ever in the initial training of teachers. There is thus no line at all in the diagram. Educational research, itself profoundly affected by the social sciences, makes some impact on teachers' craft knowledge, but little of this is direct or explicit, and so must be regarded as no more than a weak influence. Of course educational research has already generated some evidence-based knowledge (though it is not usually given such a title) which bears on craft knowledge. Indeed, it is my view that it is this particular component which is most in need of strengthening if educational research is to improve the knowledge base of teachers: this is the route to the enhanced professional utilisation of educational research.

There are thus significant differences between the knowledge bases of the two professions. That one is fed by the natural sciences and one by the social sciences is not, I contend, the most notable or consequential difference from the point of view of research. Rather, I suggest that the differences that matter are that in medicine (i) clinical researchers are usually also practitioners; (ii) researchers and practitioners are interested in the question: What works in what circumstances of practice? and (iii) they are creating a body of evidence which contributes to the knowledge base that bears powerfully on clinical practice. By contrast, in education most researchers are not practitioners and there is a much weaker commitment among researchers to producing evidence about what works in what circumstances and so a relatively small effort is being devoted to this task. Indeed, there seems to be a widespread scepticism, at least in the UK, about the capacity of educational researchers to produce a sound evidence-based knowledge of teaching.

I think they are mistaken. I am an admirer of Kurt Lewin, who is seen as the father of action science/action research and as the father of experimental social psychology. Lewin believed that research could *both* inform professional practice for human betterment *and* generate high-quality social science. The school of action research associated with the name of the late Lawrence Stenhouse seems largely to have abandoned the second of Lewin's twin goals, and thus gravely damaged the power of action research. Sound research on

evidence-based teaching should contribute 'forwards' (so to speak) to an improved professional practice of teaching and 'backwards' into social science in the form of better theories to explain such practice. Happily an evidence-based approach to improved teaching flourishes in other European countries, as in The Netherlands to judge by Bert Creemers' *The Effective Classroom* (1994) and Jaap Scheerens' *Effective Schooling* (1992).

Teaching and educational research could learn from medicine and develop an evidence-based approach to professional practice and its improvement. British educational researchers constantly emphasise and exaggerate, as I think Kogan does, the differences between the professions and so underestimate what one might learn from the other. The qualities needed by the good researcher have been memorably described by Nobel laureate zoologist and immunologist, Sir Peter Medawar:

> A sanguine temperament that expects to be able to solve a problem;
> power of application and that kind of fortitude that keeps scientists
> erect in the face of much that might otherwise cast them down; and
> above all, persistence, a refusal bordering on obstinacy to give up and
> admit defeat.[3]

On the basis of the evidence about the directions that British educational research is taking, I tentatively conclude that such a movement towards research-based and evidence-based teaching is more likely to flourish outside the UK than within it. If so, it will be a healthy step forward for British educational researchers to learn more than they have been accustomed to from their sanguine, brave and persistent Continental colleagues

---

[1] HARGREAVES, D. H. (1996). *Teaching as a Research-Based Profession.* London: Teacher Training Agency.

[2] Quoted in LOCK, S. and WINDLE, H. (eds) (1977). *Remembering Henry.* London: British Medical Association.

[3] MEDAWAR, P. (1984). *The Limits of Science.* Oxford: Oxford University Press, pp. 9-10.

232

Margaret C. Wang, Temple University Center for Research in Human Development and Education, Philadelphia

Maurice Kogan provides a visionary perspective on educational research and development by examining the research processes, products and theoretical advances of other disciplines and areas of inquiry. In his paper, he advises that we go beyond the familiar terrain of our academic and professional specialisations and seek wisdom from adjacent fields. Kogan's call for multiple ways of inquiry as a key strategy in expanding the social problem-solving role of the educational research enterprise, and his attention to what Lindblom and Cohen referred to as 'usable knowledge', particularly his insightful analysis on the state of practice of educational research, posed an impressive set of challenges for deliberation at this conference and next-steps strategies for advancing educational R&D.

I would like to focus my remarks on elaborating three aspects of Kogan's contribution. They are: production of usable knowledge; development of a procedural knowledge base; and a transdisciplinary approach to educational research and development, highlighting some of the emerging developments in those areas in educational R&D in the United States and internationally.

Production of usable knowledge

Kogan begins his analysis of how educational research can benefit from an examination of other areas of inquiry by considering the nature of knowledge produced in given areas of inquiry. He notes that education, like other social science disciplines, is based on knowledge that comes from diverse sources. New understandings may be the product of studies with scientific rigour, or socially constructed accounts of efforts to solve complex problems. He further notes that the knowledge base in education, like that of other areas in the social sciences, also includes widely held beliefs, referred to by Lindblom and Cohen (1979) as ordinary knowledge. Such knowledge does not typically emerge from rigorous scientific studies.

One promising strategy pointed out by Kogan is the power of the 'ordinary knowledge' that can be drawn from large-scale studies utilising multiple investigative methods from varied disciplines. He uses as an example the multidisciplinary approach employed in studies of epidemiology and the health sciences and how their contributions to 'ordinary knowledge' provide influential inputs into social problem solving and policy formulation. For example, epidemiologists study the prevalence and control of diseases, and they also conduct large-scale studies of social, cultural and lifestyle variables

233

that are associated with particular diseases.

Similar examples of the value of such large-scale studies for furthering our understanding of conditions of education and policy formulation can be found in educational research. Large-scale studies of social and economic contexts that influence educational performances, such as the National Assessment of Educational Progress, High School and Beyond, and the National Educational Longitudinal Study (NELS: 88) are illustrative of the power of using multiple methods of gathering information from different disciplinary perspectives. These studies and subsequent secondary analyses of their databases have contributed to policy debates on topics of significant national importance. Policy formulation on desegregation, school choice, educational reform and the sustaining effects of educational interventions are examples of the 'payback' values of such large-scale studies.

A recent example of using multiple methods of learning and knowledge that has generated widespread discussion is the Third International Mathematics and Science Study (TIMSS). TIMSS involved testing more than half a million students across 41 countries. TIMSS is one of the world's largest, most comprehensive scientific studies about teaching, learning and achievement, comprising a database that provides a wealth of information on the usage of textbooks and curricula; surveys of teachers, students and school administrators; and observations of the practices that shape education and achievement around the world. The publication of the first of a series of TIMSS reports on findings by the National Center of Education Statistics (NCES) in the United States on November 20, 1996, for example, has generated unprecedented interest in how TIMSS data can provide invaluable international benchmarks defining world-class performance in science and mathematics education in the United States.

Development of a procedural knowledge base

Studies of delivery of effective education and related social services are a high stakes area of educational research in terms of research payback. In such studies, the organisational structures and governance of institutions are examined to develop an understanding of how the delivery of services can be enhanced in the service of student development and learning. Much of educational research results in what can be termed 'declarative knowledge', that is, a knowledge base of ideas and facts about effective policy and educational practices. Although important and potentially useful, declarative knowledge usually requires further steps for transformation into procedural knowledge — knowledge about the processes or routines that are used to

implement a set of activities that lead to intended outcomes (Glaser, 1989).

Educational research has built an impressive knowledge base in declarative form about pedagogy, school climate, child development, family and community influences on development and learning of children and youth, and other topics. But much of the declarative knowledge base must be transformed for use by schools and other service delivery agencies. Developmental work is needed to provide procedural guidelines and products that are usable in situation-specific circumstances. This phase of the research utilisation process is glaringly lacking, and needs to be recognised and built into the professional roles of researchers, the school staff, university faculty and other related service providers (Wang and Gordon, 1996).

Development of procedural knowledge is receiving increased attention by educational researchers in efforts to improve the delivery of education and related social services. Chubb and Moe (1993) conducted a secondary analysis of high school students' academic performances, organisational features of schools and the leadership qualities of school administrators. These researchers asked how bureaucracy and school autonomy affected the delivery of educational services to high school students. Their research results have heavily influenced the national dialogue on school choice. The past decade has also seen a flurry of research studies on the coordination of educational, health and social services. Research on the prevalence, nature, implementation and effects of coordinated, school-linked services has contributed evidence to ways of reducing the inefficiency, ineffectiveness and fragmentation of service delivery to children and their families (Haertel and Wang, 1997). These studies have examined how schools can serve as the center of a nexus of educational, health, psychological, legal and other social services.

The research community has not contributed sufficiently to the resolution of policy debates or the enhancement of educational practice and decision making. Tomlinson (1994) attributes the decline in federal support for educational research in the US and elsewhere in part to the educational research community's record of 'poor pay-off'. In response to the large gap between the knowledge that educational research has yielded and its use in guiding practice and formulating policy, Tomlinson recommends that the research community reduce its isolation from school practitioners and engage in rigorous, collaboratively designed research projects that would shed light on the learning process.

Taking this one step further, one could argue that reducing isolation among the educational research and practitioner communities through increased

communication would greatly enhance the all-important concept of reciprocal payback and collaboration for researchers and practitioners. Teachers and school administrators would have better access to new knowledge that is scientifically established. Researchers would have better access to the practical knowledge that school personnel use on a daily basis. Improved communication among researchers and practitioners is sure to enhance the usefulness of the knowledge base as a guide to practice and policy formulation. The reduction of isolation is essential to building a procedural knowledge base, and perhaps in a more pragmatic way it can be viewed as a key strategy for increasing the potential for payback from educational research.

The case of a transdisciplinary approach to social problem solving

The third area of Kogan's paper that I would like to address is the merit of 'learning from other areas of study'. I would like to press beyond his call for a multidisciplinary view of social problem solving towards forging a transdisciplinary approach to advancing the 'usable knowledge' agenda of educational R&D. Broadening the base of knowledge that informs educational practice and policy to include the wisdom of practitioners is only the first step in increasing the usefulness of the knowledge base for improving educational practice and policy. In his paper, Kogan stresses the importance of multidisciplinary inquiry as a source of new knowledge and methods that can lead to improved educational research. The need to connect with and learn from other areas of study beyond the traditional sphere of education is at least twofold: it widens the substantive knowledge base of educators; and it diversifies the methodological approaches used to provide verifiable evidence.

Gathering knowledge and insights from other professions, disciplines and areas of inquiry can significantly increase the payback value of educational R&D. There is an emergent emphasis on such broad-based, multidisciplinary, long-term programmes of research. This is particularly evident in the most recent round of federally funded National Research Centers in the US, and some of the highly visible programmes of research funded by private foundations. I would like to describe the work of two such programmes of research as illustrative of the potential payback from a transdisciplinary approach to addressing complex social problem-solving issues. The first is a programme of research on teachers' use of pedagogical and content knowledge which employed a cognitive model of teacher problem solving that evolved from observations of physicians diagnosing illnesses (Shulman, 1986, 1987; Elstein and Shulman, 1971).

The second is a transdisciplinary programme of research being conducted at the Center for Research in Human Development and Education (CRHDE), an interdisciplinary R&D centre at Temple University. At CRHDE, scholars from a wide range of social sciences and related fields, including economics, psychology, sociology, urban development, and educational and social policy analysis, collaborate with educational and social service providers in research projects designed to identify factors that contribute to life opportunities and educational resilience of children and youth. This programme of research brings together the substantive knowledge, methods and perspectives of several areas of inquiry in order to better understand the ecology and changes in rural and urban America; it also examines the implications for improving the capacity for development and education in the service of children and youth in some of the most disadvantaged rural and inner-city communities.

Going beyond education to understand teaching better

The work on understanding teaching as a tool for school reform led by Professor Lee Shulman of Stanford University is one example of 'learning from other areas of studies' that have led to the kinds of advances envisioned by Kogan. Shulman and his colleagues believe that there is a growing knowledge base on teaching and learning which could frame reform efforts and guide the evolution of education, especially classroom practice (Holmes Group, 1986; Carnegie Task Force, 1986). Such a knowledge base can be culled from rich and extensive fields of expertise, including: scholarship in subject area disciplines; the materials and settings that are common to educational processes (e.g. curricula, texts, school organisation and governance, and professional structures); research on learning and teaching, human development, social organisations and schools; and the wisdom of practice itself.

In conceptualising what the knowledge base of teaching would entail, Shulman has provided an example of the value of multidisciplinary inquiry. His early research on physicians' diagnostic and problem-solving skills (Elstein & Shulman, 1971) has contributed much to the current call for emphasising problem solving and pedagogical reasoning in the teacher education curriculum. Shulman argued that the skills needed by a physician are similar to those needed by a teacher or other professional who must interpret an individual's symptoms or behaviours and link them to a knowledge base on acquiring subject matter-specific competencies. During the 1980s, Shulman collected evidence about how novice teachers' knowledge of pedagogy and content develops, based on observations tracking the journeys of college

students from teacher education students to novice teachers. Shulman identified six key components of pedagogical reasoning and action–comprehension, transformation, instruction, evaluation, reflection and new comprehension (Shulman, 1987).

By learning from other areas of study beyond education, Shulman's examination of the techniques employed by physicians in solving diagnostic problems yielded a new perspective for problem solving by teachers, and informed the development of an expanded model of pedagogical reasoning and action. By incorporating — or at least considering — alternative methods of problem solving, Shulman enlivened the more traditional approaches taken by educational researchers. His work provided a clear demonstration of how drawing expertise and methods of investigation from disciplines outside of the traditional boundaries which had heretofore defined the field (be they teaching, school administration, research or policy making) could provide a 'value added' payoff to educational R&D.

A transdisciplinary programme of research on development and learning of children in the inner cities

An example of a broad-based, long-term programme of transdisciplinary research that focuses on building a usable knowledge base on field-based implementation of improved practice and policy is our programme of research on fostering educational resilience among children and youth in inner-city communities at the Center for Research in Human Development and Education at Temple University. The overall goal of this programme is to identify ways to significantly improve the capacity for education in the inner cities by focusing on solutions through a transdisciplinary programme of research and development. A basic premise is that cities are complex ecosystems, whose economic, political and social climates bear directly on education, and that coordination among all elements of the ecosystem, including families, schools, and public and private sector agencies in the near and far communities, is essential to improving the life circumstances and education of children and youth in inner-city communities. Furthermore, approaches to addressing the ubiquitous research question of 'How to significantly improve our capacity for education in the inner cities?' must be similarly integrated, collaborative and transdisciplinary.

Our research shows that the picture of US cities that emerges from the research base on the development and education of children and families in the inner cities is a startling juxtaposition of despair and hope, disorganisation and potential. Nowhere are the problems and needs of children and youth as

238

great as in the inner cities across this country. Families, neighbourhoods, community agencies and schools are desperately depleted of resources and spirit; the familiar litany of unemployment, crime, child abuse and neglect, and addiction to drugs and alcohol is all too prevalent. They confound and overshadow the problems of widespread academic failure in the schools, thus crippling the next generation's best hopes for transcending the problems they and our society face. Yet cities are rich in promising resources. In spite of problems that surround them, many remarkably resilient children in the inner cities mature into healthy, competent adults, capable of identifying, nurturing and maximising what is positive in their lives to rekindle hope, and the prospect of schooling success. Although schools must be a primary focus of research efforts, research must also take into account that significant learning takes place outside of schools, in homes and in the larger community. The capability of schools can be greatly enhanced through a better understanding and appreciation of community resources and influences, family educational goals and other factors that influence the learning success of children.

Researchers at our center (CRHDE) benefit from the expertise of a multidisciplinary team of researchers and collaborating practitioner colleagues who make use of diverse knowledge as they participate in a long-term programme of research and development. They merge knowledge, theory and methods from areas of inquiry as diverse as demography, ethnography, sociology, economics, political science, educational and developmental psychology, curriculum and instruction, and social work.

Through merging theory with research and practical wisdom from diverse disciplines and professions, our programme of research aims to achieve several specific outcomes, including: delineating problems more fully; contributing new understanding about the nation's inner cities; raising consciousness about opportunities in inner cities, especially among policy makers and service providers; designing and testing interventions; and improving educational practice in inner-city classrooms and schools. Among the salient payoffs from the transdisciplinary approach to R&D is the breaking down of the barriers of isolation between researchers of different disciplines and field-based professionals, and the forging of genuine partnerships among researchers from diverse scholarly and professional backgrounds and educational practitioners and social service providers.

Building on the frames of reference from varied disciplines and professions facilitates researchers in strengthening their efforts in identifying variables, contexts and relationships that influence learning. As a result, what is tried by the educators in the schools is linked with what the sociologists tell us

about changing demographics in the city, connected to what the economists observe about shifts in employment opportunities, and tied to what the psychologists tell us about childhood and family stresses (Wang, 1995). Going beyond the diverse substantive contributions made by researchers from various disciplines and professions, these scientists also bring a range of methods that facilitate examining phenomena in important ways.

The presence of a multidisciplinary team of researchers has allowed us to embark on an ambitious research agenda that examines the influence of a wide variety of variables and contexts on inner-city children's educational success. Our multidisciplinary research agenda extends from psychology- and education-based studies of resilience to economic and demographic studies of the social ecology of urban neighborhoods and their relation to the labour market; to studies focused on policy formulation and development; to effectiveness, and nature of school–family–community programmes; and to exploring the relationship of socio-economic contexts of schooling to educational attainment.

What have we learned from the transdisciplinary programme of research on building the capacity for development and education in the inner cities to date? And what are our next steps? The following are examples of some key observations from our work that we believe have significant implications for shaping next-step efforts for improving our capacity for educational research and development in general, and research to improve the capacity for inner-city situations in particular.

Key findings on the state of practice

- The research and practical literature on inner-city life and schooling are scattered, limiting the capacity of researchers and practitioners to keep abreast of advances in theory, research and practice.

- Research-intensive universities have exhibited a limited capacity to mount transdisciplinary research efforts and place their findings in the service of broad-based community efforts.

- The current educational system has failed in its efforts to implement what is known about effective schools, classroom instruction and policy. This failure is especially devastating in inner-city schools where there are limited resources and families often cannot compensate for watered-down content, poor instruction, outdated curricula, obsolete and limited technology, and poorly implemented policies.

● Federal and state governments' influence on education and related service delivery systems has been narrowly categorical (e.g. Title I, bilingual, special education, etc.). This approach to meeting the diversity of learning needs among students has caused disjointedness and inefficiency at local levels. The policies and their implementation have been procedural rather than substantial; too often they identify, measure and reward problems, not solutions. Urban schools are disproportionately affected by this narrow, categorical and fragmented approach to improving the school's capacity for meeting the needs of the increasingly diverse students schools in urban communities are challenged to serve.

● The transition of youth from high school into the workplace or college is a crucial life passage. Most schools fail to equip students for this transition. Inner-city youth have the same desire to work as non-inner-city youth, but intervening environmental factors frequently impede the fulfillment of this desire. It is critical to provide students with strategies to find work and mechanisms to facilitate employer contacts. Schools, particularly those in the inner cities, need to facilitate a smooth transition into the world of work.

The next-steps research and development agenda

Findings from our work suggest three major next steps in R&D to advance our capacity for education in inner-city communities:

● Focus on fostering educational resilience among children and youth who live in a variety of highly adverse circumstances that place them at risk.

● Implement effective practices that focus directly on classrooms and homes, where learning takes place; eliminating educational segregation in schools and providing for student diversity within one inclusive education system.

● Forge greater school connections with families and community.

Accomplishing these three next-step action items would result in demonstrating that: (a) research on fostering educational resilience that relates broadly to families, schools and the community is useful; (b) research that connects in all possible ways across disciplines forges the emergence of broad, cohesive views and understandings; (c) researchers are energised to be connected to social problem-solving concerns that significantly enhance the capacity for development and education in inner-city communities; (d) well-

241

focused policy discussions concerning research findings have clear implications for improving educational and social service delivery.

The ubiquitous question of what conditions are required to cause massive improvements in the learning of children and youth will continue to drive and refine our programme of research on inner-city education. The guiding principle that our work must lead to practical applications in the near term, while pursuing a long-term programme of research and development, will continue to undergird the evolving process of development, refinement and application of our research.

Conclusion

In conclusion, I would like to reiterate what Kogan has eloquently raised by outlining three broad directions for furthering the payoff for educational research and development:

- Educational research needs to be inclusive, embracing the substance and method of many disciplines, professions and areas of inquiry.

- There are a variety of sources of verifiable knowledge which can contribute to the knowledge base on effective educational practices and policies. These sources include research results on teaching and learning; information on school settings and materials; and the wisdom of practitioners.

- Ordinary knowledge, which includes widely held beliefs, common sense, casual observations and thoughtful speculation, when acted upon, becomes usable knowledge. Usable knowledge influences policy, guides practice and informs decision making. We increase the usability of knowledge by making research results accessible to consumers and practitioners, by collaborating with practitioners in the design of studies, and by stepping outside the boundaries of our specialisations to take into account the knowledge base and methods of inquiry from disciplines beyond the traditional boundary of educational research.

There is a Chinese saying that a journey of a thousand miles begins with a first step. Kogan has presented a challenge of initiating a step in his presentation to advance significantly the journey of educational research and development into the next millennium by charting an agenda on social problem solving and the building of a usable knowledge base through learning from other disciplines and beyond.

References

CARNEGIE TASK FORCE ON TEACHING AS A PROFESSION (1986). *A Nation Prepared: Teachers for the 21st Century.* Washington, DC: Carnegie Forum on Education and the Economy.

CHUBB, J.E. and MOE, T.M. (1993). *Politics, Markets and America's Schools.* Washington, DC: The Brookings Institution.

ELSTEIN, A.S. and SHULMAN, L.S. (1971). 'A method for the study of medical thinking and problem solving.' Paper presented at the Annual Meeting of the American Educational Research Association, February.

GLASER, R. (1989). 'Expertise and learning: how do we think about instructional processes now that we have discovered knowledge structures?' In: KLAHR, D. and KOTOVSKY, K. (Eds) *Complex Information Processing: The Impact of Herbert A. Simon.* Hillsdale, NJ: Erlbaum, pp. 269-282.

HAERTEL, G.D. and WANG, M.C. (1997). *Coordination, Cooperation, Collaboration: What We Know About School-linked Services.* Philadelphia: Temple University Center for Research in Human Development and Education.

HOLMES GROUP (1986). *Tomorrow's Teachers: a Report of the Holmes Group.* East Lansing, MI: Author.

KOGAN, M.C. (1996). 'Learning from other areas of study to education research.' Paper presented at the National Foundation for Educational Research Conference, Windsor, December 2-4.

LINDBLOM, C.E. and COHEN, D.K. (1979). *Usable Knowledge: Social Science and Social Problem Solving.* New Haven: Yale University Press.

SHULMAN, L.S. (1986). 'Those who understand: a conception of teacher knowledge', *American Educator*, **10**,1, 9-15, 43-4.

SHULMAN, L.S. (1987). 'Knowledge and teaching: foundations of the new reform', *Harvard Educational Review*, **57**,1, 1-22.

TOMLINSON, T.M. (1994). *Educational Research and Reform: An International Perspective.* Paris: Organisation for Economic Cooperation and Development.

WANG, M.C. (1995). *Next-Steps in Inner-City Education: Focusing on Resilience Development and Learning Success: An Impact Report.* Philadelphia: The National Center on Education in the Inner Cities.

WANG, M.C. and GORDON, D.E. (1996). 'Improving urban schools: better strategies for the dissemination and utilization of knowledge', *Educational Policy* (special issue, June).

8.

The impact of educational research

Ramsay Selden, Director, Education Statistics Services Institute, Washington

Introduction

I have been asked to comment on the ways in which educational research has an impact on education. I appreciate very much the opportunity to do that and to contribute to this important conference commemorating the fiftieth anniversary of the National Foundation for Educational Research. I would like to share the sentiments of gratitude voiced by others and commend NFER for being so generous in designing a conference which looks more generally at the status of educational research at this point in time, as opposed to reviewing their own accomplishments.

I am speaking in substitution for Marshall (Mike) Smith, who is a policy researcher in education and is presently Acting Deputy Secretary of Education in the United States. I believe this is the highest federal post to which a bona fide educational researcher has risen in our country. As a member of the Administration, Mike is in the enviable position of being able to set and to study policy at the same time!

My organising metaphor for talking about the impact of research on practice in this paper is that of a telescope: the relationship between research and practice can be thought of as a telescope which can be viewed through either end — top-down and distant, or bottom-up and close. I will develop this more, but the basic idea is that the connection between research and practice can appear to be very remote and small, if viewed from the top, or it can be seen as very close and immediate, if looked at in detail.

I believe research has a fairly extensive impact on education. The problem is, this impact does not necessarily occur in the ways we once thought it did; and, to see the impact, you have to look very closely at some specific instances.

How research can have an impact

There are three ways in my mind by which research can have an impact in education.

One of these is in revealing patterns and phenomena in the system. Here, research performs a descriptive or exploratory function, providing us with information which allows us to see how things are.

Another way in which research can have an impact is by changing our understanding of underlying processes involved in education. I must confess that we do not know very clearly *how* such a developing understanding affects practice, but it has been proven through experience that understanding basic processes plays a role in improving the nature and effectiveness of the professional practices we use to develop those processes.

A third way research can have an impact is by clarifying and guiding the roles of various actors in the educational system. By this, I mean that research can explore and explain how various roles in the education system work, and we can use that information to adjust the roles — how we perform them and how they work in relation to one another. Much educational research has been of this nature: looking into the roles of teachers, school leaders, policy makers, families or members of the community.

Specific instances of impact

Under each of these suggested modes, I think we can point to some fairly concrete and significant instances where research has had an impact on the education system, at least in the United States.

An example of how descriptive information can affect the system is our recognition and response to low achievement in urban school systems in the United States. By the 1940s or 1950s, our cities had been left with a very disproportionate share of poor children. For whatever reasons, the educational achievement of these students was a special challenge. Until the 1960s or 1970s, we did not really understand how bad this problem was because we did not have data. Then, regular administration and summarisation of standardised achievement tests became widespread, and we could see that students in urban systems were at about the twenty-fifth percentile on these tests, compared with the fiftieth percentile average. We would not have known how severe the achievement problem was if the testing had not been done and if the results had not been displayed grade by grade, subject by subject.

The response was at least a great *push* for better achievement in the cities. The federal government and the states launched compensatory education initiatives, and there was a great movement for better teaching of 'basic skills'. That push did not result in much progress in student achievement — although there was some limited success in bringing about gains in the 1970s and 1980s — but the main point is that we would not have known about the problem at all without the test results. (American cities seem to be too

overwhelmed by the problems of poverty, scale and political will to be able to do anything substantial or lasting about these achievement problems.)

There are clear-cut examples of how practice has been affected by research into the processes it is concerned with. One of the best examples in my mind is research on basic reading processes. In the mid to late 1970s, basic psychological research began to investigate and explain how reading takes place. This research moved us from a behavioural to a cognitive model in our understanding of the process of reading. We came to understand how readers integrate new information with what they already know and how their ability to guide their own reading processes affects the effectiveness of their reading.

One project in particular comes to mind. About 1979 or 1980, the National Institute of Education (NIE) received a proposal from Scott Paris and his colleagues at the University of Michigan. The proposal was to do an experimental study to determine whether, if metacognitive skills in reading were taught directly, this would have an effect on reading performance. This was a classic instance of needing an experimental study to confirm causal connections when correlational studies already had confirmed an association.

In this study, it was proposed that college students be used as the sample or target population. Those of us at NIE at the time suggested that elementary school students be used, and that the study be done in the context of a public school system. These changes were made and the study was implemented. It found that by directly teaching these skills (to monitor and manage their own reading) the ability of low-achieving students to read could be improved. Probably more important, a sort of 'enfranchisement effect' seemed to take place: students who were taught these skills had higher reading abilities than control students a year later, without any further intervention.

Other research in this area explored reading comprehension processes, looked at the quality of textbooks and tests being used in the schools, and explored cultural and background effects in reading. This research completely changed our research-based understanding of reading. Gradually afterwards, practice began to change. Textbooks and instructional programmes began to address comprehension and metacognitive skills directly. More authentic texts began to be used in reading instructional materials. Reading tests were reviewed to examine how well they measured important aspects of the reading process. Teacher training began to change, so new and serving teachers were taught how to use more research-based techniques. By the late 1980s in the United States, practice had been largely transformed.

So, even fairly basic research explaining the processes we are trying to develop in students can allow improvement in the approaches used to develop these skills.

A further example of research impact is when research explains various roles in the system. In this instance, I want to give a counter example of how research can have an impact: an example where the impact has been negative, to show how consequences can be unpredictable and counter to what we want, if we are not careful about how we interpret research implications.

This example is what we in the US refer to as effective schools research. In this instance, we had correlational and outlier research which was aimed at finding characteristics of schools which were associated with school success or unusual school effectiveness. Among the findings was that, in effective schools, principals or school heads were particularly effective 'instructional leaders'.

That was all well and good; where things went bad was in how these findings were interpreted and implemented. It led to much effort within the United States to have principals acquire the attributes of effective instructional leaders. The problem was that this was done in a simplistic or superficial way, eventually degenerating to things such as stipulations that principals should get out of their offices and go up and down hallways to visit classrooms and supervise what was going on. This, of course, became unproductive or counter-productive to genuine instructional leadership since it did not embody very effective practices. But it shows how research findings on the roles of significant agents in the system can have a tangible effect on behaviour.

Perhaps more positive examples might be found in research into teaching or into family involvement in learning. We have had very good research which has elucidated the nature of high-quality, effective teaching, and this has allowed us to refine our school improvement and teacher training approaches. We also have had good research which has given us better insights into how different kinds of households relate to the schools. This has allowed some school systems to develop improvement efforts, tailoring the practices they use to work with different kinds of families.

Another instance of impact can be found in international/comparative studies. In the US and, I think, other countries, the policy interest in international comparisons is resulting in these studies having great impact. We were largely alerted to questions about the effectiveness of the US education system by mediocre standings in international studies. Fortunately, the powerful,

descriptive nature of those studies is no longer confined to just the achievement bottom line: recent results from the IEA Third International Mathematics and Science Study prompted as much interest by the press and policy makers in comparisons of instructional practices as in the achievement rankings.

Horizontal and vertical dimensions of research impact

The process of educational improvement has a horizontal and a vertical dimension in the US, and it is possible to think of the impact of research playing out on these two directions.

Horizontally, we are working under the concept of systemic reform — this is seeing broadly how elements in the system are organised and making sure that they line up with one another coherently. By elements, I mean curriculum, the content and expectations of assessments, teacher training and licensing, accountability mechanisms, and so on. We have discovered that, in many situations, because we have a very decentralised system, these elements may not be in alignment with one another. An assessment programme might be in place which does not reflect the latest curriculum developments, or state efforts might be in conflict with local or national ones.

Research can inform and guide the nature and alignment of the various parts of the system, showing us what to do so that things are aligned more effectively. For example, if research on learning processes suggests curriculum or teaching should be changed, we know from this systemic view that assessment, teacher training and other elements may also need to be changed to support the improvement we are trying to make in instruction or curriculum. That is how the horizontal dimension of this view plays out in relation to research innovations. Research also, of course, helps us to determine how the elements themselves should be thought of, as indicated above.

On the *vertical* dimension, we have to attend to the implementation of reforms. We have found in the US that movements or exhortations at top levels of the system — state policies or national movements, for example — will not have an impact if efforts are not made to make sure teachers at the classroom level understand and implement the reforms or approaches effectively. Here, when research suggests a change, we now are beginning to think in terms of strategies which will result in knowledge, understanding and effective implementation at the most immediate levels of practice. This translates into a need for deep and faithful processes of teacher development and support, in contrast to the kinds of wand-waving and magic bullets we often tried to use in the past.

249

Research and its responsibility for the process of change

Seamus Hegarty (above) observed that research is unfairly criticised for its role in the change process. The criticism is that research should bear responsibility for having more of an impact — that if it was effective, it would have more impact. The implication of this criticism is that research should be more conscientious and useful, addressing topics and being done in ways that result in more of an effect.

I concur with Seamus that this criticism is unfair. Research is one of several types of information that are available to inform or guide practice, so the choices of practitioners in the information they choose to use is somewhat beyond the control of the research enterprise, and research by itself cannot guarantee that it will be attended to and have an impact. Contemplating this relationship, however, reveals that research has an impact which is complex, difficult, unreliable and difficult to understand. Let me develop each of these qualities.

First, that research is complex in its impact: this means simply that the processes by which impact occurs include many factors and that they play out and relate to one another in very complicated ways. I think this holds true for many of the examples we looked at: the reading research and school effectiveness examples described above certainly reflect this complexity and subtlety.

Second, that the process of impact is difficult: by this I mean that the process by which change occurs as a result of research is difficult — that it is not easy to accomplish. I think most of us would concur with this on the basis of our own experience. Again, the examples we have talked about illustrate it. The reading research example was one in which change occurred over a period of ten or 15 years, not very quickly or easily. The instance of effective school leaders underscores the subtlety of the process. And the example of urban achievement is one where desired effects have not been accomplished even now, after 20 or 30 years.

Third, that the process is unreliable: by this I mean that the process by which research has an effect in actual practice is usually difficult to predict and control. In most of our experiences, the change process unfolds in ways that cannot be anticipated reliably. The effective schools example certainly confirms this: things developed in a completely different way from what was anticipated. This poses a dilemma for us in how we look at research and its impact. It may be that much, if not all, of the impact is largely unpredictable,

or at least that the impact will take on a character which is largely different from anything we could speculate about ahead of time. Could anyone have anticipated that basic psychological research into reading would have a major impact on how reading was taught? Probably not. But if it is true that impact occurs in largely unpredictable ways, is there any rationality to the system? Can we say that there is anything systematic about the way change occurs as a result of research?

To respond to that, we identify some general kinds of impact that can be predicted. For example, we could say that basic research like that which was done on reading would generally have some impact such as we observed. We might look to similar kinds of impact from other areas of basic research, albeit not as extensive or exactly like that observed in reading. Also, there probably are areas where the impact of research would be more predictable.

Finally, the ways in which research has an effect are difficult to understand. Because of its complexity and subtlety, it is not easy for us to come to understand the change process at all, let alone the role of research in it.

Now, what *can* we say about the change process and the role of research? First, I think one of the most important things to point out is that none of us really believes in a production metaphor for this process any longer. At one time, we conceived of a linear process from basic research to applied research to development to implementation to evaluation. We all know that the actual process is more complicated, more subtle and more organic than that, and we have not often seen that linear process work in practice. On the contrary, we have seen other, more complicated processes in real life.

What are those real processes like? Let us take the example of class size to explore this. Over the years, a great deal of research and synthesis of the research have been done on the effects of class size on schooling. No simple, clear, right or wrong answers have become available, certainly not early on. Some studies said one thing, others suggested something else. Reviews of the literature and meta-analyses were done. Big, definitive studies were designed and conducted to resolve the issue 'once and for all'. Finally, some complicated conclusions began to appear: class size had certain effects within particular size ranges and had different effects for different types of students. At this point, after decades of research, review, synthesis and development of implications, we began to have some guidelines for practice out of research. But, the process of arriving at these implications was much longer, more complex, more subtle and more iterative than the production process we might at one point have had in mind.

251

To continue with this example and explore how things become even muddier, with guidance finally in hand from research, we could see policy makers making decisions in response to factors other than research. Teacher groups and parents wanted smaller classes, regardless of the research literature on the effects, and this push was often powerful enough to warrant changes, even at great extra financial cost. What is important about this development is that it shows how research — subtle, imprecise and complex as it is in its own role — is but one of many forces driving change in education.

This leads to a corollary to Hegarty's observation: that professional practitioners and policy makers are not necessarily very conscientious in *how they use research*. I believe this was said by Susan Fuhrman and others. Not only is it not fair to expect research to bear too much of a responsibility for the impact it has, it is also reasonable to expect other actors in the process to do their jobs more effectively, and sometimes this means using research findings more conscientiously.

Mechanisms by which research can have an effect

All of this then leads me to summarise several ways by which research, understood in this more complex way, can have an effect.

One of these is by expanding our theories of education and the underlying processes it affects. An example was given previously on learning to read. Research truly has made it possible for us to develop powerful cognitive models of learning, and that has had an effect on education. Another example is teaching. Research has resulted in a progression of views of teaching — a series of evolving theories, the most recent of which has been moving to a view of the subject-specific clinical practitioner from a generic view of good teaching. Other areas where this has occurred include theories of schools as organisations and work on how schools should relate to their students' families.

How should this process of theory-building be structured and supported? How can we ensure that research continues to result in breakthroughs that enable us to acquire these new understandings of underlying issues or processes in education? Actually, the process of basic research is generally a process of theory-building: research is often, if not usually, done to move beyond current theories. In this way, the process constantly ensures that research will be done which allows us to develop our theoretical understanding. In this sense, I do not think that 'management' of the process of basic research is necessary for it to continue to have an impact on education. Rather, as long as basic

work is done in psychology, sociology, anthropology, economics and so on, it will generate new understandings, some of which will be applicable to education.

I do think the connection between basic research and its theory-building and educational improvement can be enhanced. Basic researchers can be sensitised better to the potential implications of their work for education, and educators can be alerted to the potential implications of even very basic research. Researchers tend unnecessarily to see their work as being very distant, and educators look too much to applied research and evaluation for implications, and too little to basic work in the various disciplines.

Another way research has an impact is by providing information on what is going on. This is a hobby-horse of mine, but I think it is important. Empirical research offers the tremendous benefit of objective, systematic information on phenomena we are concerned about. As an enterprise, education can take place without much systematic information being collected on it. We know that only too well! The void can be filled, of course. Earlier, I talked about test scores of urban students. That is a good example of this kind of information and its impact. Another is information on distribution of services. This is extremely important in order to know where the strong points and weaknesses are in current educational programmes. We are learning how surveys, including international ones, can be used for this purpose in the US. However, very little survey activity is undertaken to find out how things are out there! States and local school systems, which operate schools in the US, could use this information to great advantage for improvement of services where it is most needed.

Another example is microscopic work on the implementation of reforms, such as that of David Cohen and his colleagues. That research is showing us that reforms are not being implemented as we imagined they would be, by observing practice in some sites in great depth and detail.

Another route is by informing the development and implementation of new approaches and products. It is not very glamorous in the academic scheme of things, but good research into the effectiveness of techniques and approaches is painfully scarce in education. When studies do come out, they are so rare and striking that they attract a tremendous amount of attention. Of course, we have had much experience with methods studies that have been flawed: factors that have not been accounted for, inquiries that have not been done in enough depth, failure to base studies on strong enough theoretical foundations. But we have learned how to avoid many of these mistakes, and we should be

doing more work which informs the selection or use of approaches in education.

What is the role of certain agents of change in the process?

There are several agents which contribute to change in education. What is their role in the process by which research has an impact in education?

One of these is information. Through professional information sources and through the general and professional media, information can play an important role. The problem is that we may have outmoded models in mind for the role information plays. We may have something analogous to the production model in mind, when in reality other processes are at work. The question is, if research is to have more of an impact, how could information help?

Let me provide one thoughtful, positive example. This comes from the State of Kentucky. Many of its reforms are founded on research, but information is structured and used in very strategic ways. Teachers are sent newsletters directly reporting in a concise, accessible way the directions of the reforms and their rationales. They are not spoon-fed or patronised, but rather given the information and rationale that they need as educators. And information is provided to them in creative ways: rather than sending it to them through the bureaucracy, it is sent to their homes where they are more likely to notice it and read it.

Another hugely important agent is training, and it seems to be in a position to play a critical role in ensuring research has an impact in the future in the United States. We are trying to bring about research-based reforms on a wide scale, but the capabilities of teachers and other practitioners is a major problem. Their capabilities simply are not up to the vision of learning which research has enabled us to develop and which we want to move toward. The problem is that we need change on a massive scale, and training appears to be the only answer. We have to develop new kinds of genuine capability in teachers, so we need to be able to use pre-service and in-service teacher training to develop that skill in a large number of teachers. We have not been able to do this in the past, however, and we do not know how to structure and use the training system to accomplish it.

Some models of how it might work are beginning to emerge, but they are very different from how we have structured teacher development in the past. One model is professional development schools, where schools essentially

are in a collaborative relationship with universities, working together around the problem of improvement of practice. Other approaches include coming up with very different models for pre-service and in-service training, with the new models aimed directly at improving the ability of teachers to teach to the new visions of learning. Long-term, professional improvement networks are an instance of this.

Another agent could be professional support — particularly using information technology. Teachers and other practitioners receive little support for their professional practice, but such support could make a great deal of difference in helping ensure that the implications of research are realised. Some programmes in the US are aimed at providing information and more structured, active forms of professional development to teachers, basing the support on the implications of research. For example, professional improvement networks work with teachers to help them develop their practice in ways suggested by research, and some states and local school systems are accompanying instructional reforms by supporting teachers to adapt their practice in desired ways, sometimes using telecommunications technology. Schools of education are adopting computer networks as a way to mentor teacher trainees and enhance the transition to practice.

Given all this, how can we help ensure that research has an impact in the future?

It helps to remember that the role of research in relation to practice and improvement can be like a telescope: we can look through from the big end or the small end — which end you use drastically changes your perspective. The top-down perspective gives us a very distant view of how research is related. However, this view is probably not relevant any longer. Instead, we need to take the 'small-end' view, requiring that we look in much more focused and immediate detail, and that we be much more modest in the scope we expect the impact of research to have. Bearing this in mind, we can see that research can have a tremendous impact —

● expanding the theoretical understanding upon which we can base practice;

● providing us with crucial information on what is going on in the system; and

● guiding our roles as educators, managers, policy makers *and* researchers.

This suggests that we can do several things to ensure that the impact of research is maintained in the future:

- continued support for basic research into processes;
- continued descriptive information to guide the system;
- continued, careful work on roles and other aspects of the system;
- careful work within the system on the implications of research; and
- work with professional communities on improvement of practice.

Respondents' comments

John Gray, Homerton College, Cambridge

Introduction

There are numerous reasons why a single piece of educational research may have less impact than its producers or sponsors might have hoped for. It may, for example, have been poorly conceived and executed; it may have addressed 'yesterday's problems' or been portrayed as doing so; or it may simply not have been disseminated in more than a minimal way. There are a few guidelines that researchers can be encouraged to follow and some examples of good practice. Not surprisingly, however, only a small proportion of what is produced survives in the market-place of ideas. Many researchers have grown to accept that what happens to their research is, to a greater or lesser extent, beyond their control. Unfortunately, when the research process is conceived of as the accumulation of insights from a large number and variety of individual studies, deeper problems about collective impact and influence, which affect the research profession as a whole, may be ignored.

At the current time there is undoubtedly a feeling that educational research is less influential than it might be. As the OECD's Ministers of Education put it several years ago: 'The potential of educational research as an integral element of improvement remains largely underdeveloped. Traditional academic research has its own special part to play (but) more important still, much research and development needs to be grounded in practice, involving staff and institutions, whether individually or collectively, in a constant process of diagnosis, comparison and analysis. To achieve this, experimentation and innovation should be given strong support' (quoted in OECD, 1995:9).

Faced by such challenges educational researchers would do well to re-examine their core assumptions about how research influences practice and to seek ways of improving their collective performance, if for no other reason than that their own survival would seem to demand it.

Rethinking some naïve assumptions

When designing a piece of research, most researchers include a dissemination phase. Almost invariably, this comes last. From the researchers' point of view this may seem logical. Until the research has been completed, what is there of value to disseminate? The consequences of this sequencing for potential users of the research are, however, considerable. Underlying them are more fundamental issues of control; the timetable for events is essentially that of the researchers. Users may not, of course, feel that they have the same

kind of influence over the ways in which events unfold. Even as a piece of research is being commissioned the seeds of its own irrelevance may be being planted.

Assumptions about the most appropriate forms of communication may also need to be revisited. Few researchers these days confine themselves to dissemination through academic journals or lengthy books and reports. They recognise the importance of tailoring their approach to their audience. None the less, approaches which are essentially 'academic' in character still surface with depressing regularity. Opportunities for potential users to engage with the researchers' evidence are often highly constrained. Lecture formats with minimal opportunities for interaction still prevail.

Building for the future in the short term

In the light of these problems some rethinking of taken-for-granted dissemination practices seems called for. Indeed, researchers might do well to pool their collective knowledge on this subject with a view to constructing some kind of dissemination guidelines on *How Adults Learn*. The following are some of the issues such guidelines would need to address:

Timing and audience Some opportunities can be planned in advance but many are only likely to emerge as the research proceeds. Planning for the unexpected may be a necessary skill of the good disseminator. And whilst conventional dissemination through journals and the like is likely to continue to be appropriate, other more innovative strategies have a place alongside them.

Locating the context The audience for a specific piece of research is likely to be limited. Most users have rather broader interests. They are likely to be based on a concern to link a particular piece of research to a rather wider set of concerns. Relying on the user to make the links and to explore the implications is likely to be unhelpful. Researchers need to accept more of the onus for making (some of) the connections.

The place of methodology How a piece of research has been constructed is central to the research community. Unfortunately, methodological discussions often obfuscate, albeit unintentionally, rather than clarify. Methodology ought to be of more concern to users than it frequently is. In practice, few users have sufficient knowledge about methodology to comprehend what is of key importance. Researchers need to accept more responsibility for communicating how the methodological dimensions of research in their field

258

determine the kinds of outcomes that can be anticipated. If they can do so without constant references to technical language, so much the better.

The acceptance of multiple perspectives It is a well-established feature of much educational research that participants will have multiple perspectives based on coherent but possibly conflicting positions. The existence of such variety needs to be more fully acknowledged when it comes to dissemination; potential users will have a range of concerns and possible uses for a particular piece of research. There is also much to be learnt about why they fail to take up particular implications. Regrettably, researchers' understanding of why their ideas fail to have the impact they might have anticipated is often limited and, on occasion, even naïve.

The greater challenge of knowledge transfer

Thinking about dissemination in the so-called 'hard' sciences has moved on in recent years. Scientists have become (and been forced to become) a good deal more sensitive to the conditions under which 'knowledge transfer' may occur. It would be convenient for social scientists to argue that the dissemination of ideas does not lend itself to the same kinds of strategies as those required for 'scientific products'. To maintain this position, however, runs the risk of ignoring the more serious challenge which knowledge transfer poses.

To create knowledge transfer, researchers must engage in two separate but related activities. First, they must consider what 'the knowledge' to be transferred actually is. Thinking in this way is likely to prove a difficult but none the less productive discipline. The 'products' of educational research will only rarely resemble those of conventional scientific endeavours; researchers will doubtless need to become more skilled at recognising precisely what it is they have to offer. Second, for knowledge transfer to be effective, researchers will need to take into account something of the readiness of the proposed recipient(s) for the knowledge it is proposed to transfer. Researchers will doubtless quickly come to realise that, for their efforts to be successful, they will need to engage in some capacity-building amongst many of those they hope to influence.

Knowledge transfer implies a more sustained commitment to developing and using research than conventional notions of effective dissemination have demanded. Such an approach, when pursued with sufficient vigour, has the capacity to transform some of the ways in which researchers and users currently relate to each other. If researchers' ideas are worth implementing,

then they would do well to take some steps to build up the understanding of key innovators in the system and support them whilst they learn more about the benefits of a research-oriented approach. They will need to be encouraged to become masters and mistresses of what is already known, especially if they are to go on to develop an appetite for further knowledge. To capitalise on the new opportunities, users will also need to recognise that they have begun to reconstruct the ways in which they themselves have conventionally worked.

Many researchers are already well placed to face up to the challenges of knowledge transfer, even though its language and assumptions may currently appear somewhat alien. If every researcher were to decide to target a local education authority or a school where they intended to ensure that a measure of knowledge transfer took place over the next three years, the profession would almost certainly find itself in a more powerful and influential position than is currently the case. It might even be able to look back with pride at its collective achievements. Research has made a difference to educational practice; researchers must now square up to the challenges of continuing to do so.

Reference

OECD (1995). *Educational Research and Development: Trends, Issues and Challenges.* Paris: OECD.

Alejandro Tiana, Spanish Open University

I have been kindly invited to respond to the paper from Ramsay Selden, which deals with the suggestive topic of the impact of educational research. The interest of the topic was clearly demonstrated by the number of references to it throughout the conference. There were many voices addressing the subject in different ways and from different viewpoints, even in sessions not dealing explicitly with it. In almost every general discussion, some thought was given to how best to disseminate and use the findings of educational research. So, in trying to respond to Selden's paper, I would like also to take into account the valuable contributions made by other participants in the course of the conference.

Some conflicting evidence concerning the impact of educational research

In my opinion, the paper from Selden and other comments raised at least as many questions as they gave answers. In fact, impact and dissemination are considered crucial matters by the research community, but neither researchers nor policy makers seem to be very sure about how to address them in our mature societies.

There is some conflicting evidence which reveals the difficult role played by educational research at the present time. I would like to refer to four pieces of evidence which exemplify my claim.

First, a growing interest in building up *a knowledge base* for educational policy and practice can be seen at the international level. The title of a recent report from OECD (1996) shows this new trend very clearly. The new term goes further than simply recognising the importance of educational research. A number of different sources of knowledge are necessary in the process of decision making. Of course, research and information systems do play a significant role in that context, as exemplified by the title of another international report, in this case from UNESCO (Reimers, McGinn and Wild, 1995). Other new instruments are receiving a growing interest. Among them, education indicators are paramount, as the success of the INES project from CERI-OECD clearly shows. But policy analyses, national or international surveys and comparative studies should not be neglected in this context. All of them contribute to developing a deeper knowledge and a better understanding of educational phenomena and to setting the ground for making informed decisions.

261

One could expect that educational research would play an important role and have a significant impact in this new situation. But it seems as if it is not clearly placed in this group of activities. As a demonstration of this displacement, one can refer to the paper by Barry McGaw. It expresses some doubt as to how well the OECD definition of R&D adapts to educational research and where the latter should be placed in the R&D continuum. In his opinion, the boundaries between research, comparative studies, 'short and sharp' reports and policy analysis are not always very clear. As a consequence, educational research must find a new role if it is to make an impact on practice.

A second evidence of conflict arises out of the relationship between researchers and policy makers. It is almost a tradition to consider them as two different, separate, almost opposed communities. As one could expect, a number of interventions were made in the conference about the different interests, cultures, approaches to knowledge and agenda of the two communities. Having working for some years at the interface between them, I have to recognise the validity of the different perspectives.

But perspectives often give a biased view of reality. In fact, one has also to acknowledge the mutual interests and concerns both communities share. Researchers look to policy makers for funding and/or support for their activities. In their turn, policy makers look more and more to researchers to get from them the necessary legitimation. As Ernest House has stated, 'organised science has achieved an unparalleled cultural authority' which leads governments sometimes to 'substitute scientific authority for lapsed traditional authority' (House, 1993, p.33). So, even if both communities perceive themselves as different and separated, they definitely need each other. Conflict is evident, but should be addressed by new approaches based on mutual recognition and understanding.

The third evidence refers to the effective impact of educational research on policy making and practice. Again, conflicting discourses and views can be identified. As a general impression, educational research is meant to have little impact on reality. The image expressed by Karl-Heinz Gruber — the research community as a pack of 'wolves howling to the moon' — reflects the self-portrait many researchers have of themselves. They consider that their work does not receive the fair attention it deserves.

But if one looks more carefully at the impact had by some lines of research, one must recognise their importance. Some examples of strong conceptual, methodological and practical influence can be found in fields such as special educational needs, assessment and cognition. The work done by researchers

like Seamus Hegarty in the field of special education has provided the education community with a number of new tools — for example, the concept of *special educational needs* — to address disadvantage. The paper from Chris Whetton clearly illustrates the contribution of research on assessment to the new practices recently developed in that field in a growing number of countries. And the influence of constructivism on current teaching and learning processes clearly reveals the significant impact that pure and applied cognitive research may have on practice.

These few cases, which could of course be confronted with a number of contrary ones, demonstrate the uneven impact that research can have in different fields. Researchers should reflect upon those differences and draw their own conclusions.

The fourth conflicting evidence relates to the question of financing educational research. Susan Fuhrman was very clear when speaking about the current situation of underfunding in USA, which also applies to the European Union. Even if the financial contribution from a growing third sector — foundations and suchlike — is significant in some countries, most of the costs of research in education are now covered by universities themselves. The situation is similar all over the developed world, as OECD indicators and several international reports demonstrate (OECD, 1995). The influence of the *value for money* principle, the high political and social expectations on the direct impact of research on teaching, and the concerns expressed about its quality have contributed to this financial decrease.

But the situation described also has its converse. In fact, experience shows that funding is available according to political interest in certain fields. Research on topics such as education indicators, international surveys of academic achievement, and curriculum development and assessment have attracted considerable funding in recent years. Authorities have invested significant amounts of money in developing such fields of study. So, while the lack of appropriate funding is a general problem, it does not affect every single project in the same way. Once again, reality is complex and multifaceted.

Dealing with complexity

These conflicting pieces of evidence come to underline what I consider to be the central statement made by Selden: 'Research has an impact in ways that are complex, difficult, unreliable and difficult to understand.' The idea of complexity is crucial here and it deserves some more attention.

263

In fact, even if we usually speak about the complex connections between research, knowledge and practice, we are still used to linear, simple, direct models to explain that relation. As a recent study showed, a significant proportion — around 40 per cent — of Spanish researchers in education share the belief in a simple, additive model of knowledge growth (Velaz de Medrano, 1996). And I guess the situation in Spain is not very different from other places, even if the specific percentage may vary. Another deeply rooted image in our educational milieu is that of a linear strategy of change, in which research, development and training are successive steps in the process of transformation.

The common feature of those images is simplicity, but reality has proven to be complex. Some authors like Fullan (1993) have stressed the fact that we need to learn how to live and deal with complexity in the field of educational change. Some question the validity of this post-modernist approach, but the debate is not merely a question of fashion. A deep understanding of the process of impact and dissemination of research demands taking account of complexity.

In my opinion, we should accept that the correct understanding of that process is an unresolved issue. The questions posed are many and we still lack the correct answers. But we should recognise that the lack of answers sometimes derives from the inappropriateness of the questions. We should focus more attention on raising the right questions, as Maurice Kogan and David Hargreaves did in their paper and response. Having raised the right questions, we can be confident of getting the correct answers eventually.

Some ideas and suggestions for discussion

It is not my intention to present here a set of *right* questions. The task is a collective one, to be undertaken in the future. I would only like to offer some suggestions or ideas for discussion, in other words some pathways to be explored in order to understand the complex ways through which educational research impacts on reality.

My first suggestion is to recognise and explore the possibilities offered by the new concept of *mediation*. Researchers and policy makers have frequently thought about their relations as if they were direct interlocutors. But very often they communicate with and influence each other through several *mediators*. Different OECD studies have insisted on this fact. The important role played by mediation in the interface between producers and users of information is nowadays recognised.

Mediators belong to very different categories. Some have a formal, institutional character. That is the case of EURYDICE, the information network from the EU, or the World Bank. As Luce Pépin and Maris O'Rourke demonstrated at the conference, both organisations contribute to fostering the production of knowledge at the same time that they disseminate and apply research findings. It is not difficult to find examples of this kind of institution at a national level, such as the Spanish Centre for Educational Research and Documentation (CIDE), well-known to this respondent.

But we can also find another type of mediator, as is the case of many researchers working in policy making for some periods of time. In the conference there were a number of them, currently working for national governments or international organisations. They play a significant role in the interface between production and use of knowledge.

These mediators, formal or informal, institutional or personal, produce a double effect. On the one hand, they provide practice and policy making with conceptual and methodological tools as well as operational instruments. On the other hand, they influence the development of new agendas for educational research. Their role is a significant one, even if it has not been traditionally recognised.

My second suggestion consists in further exploring the different models which tend to explain the process of dissemination and impact of educational research. There is no longer a single model but a coexistence of several different ones. Traditionally, the influence of knowledge on practice was seen as a simple, linear one. The sequence pure research–applied research–development–training–implementation was assumed to explain that relation. But studies carried out in fields other than education — namely health and social sciences — revealed the importance of other ways and strategies. According to that, we nowadays accept different models and approaches for impact and dissemination, such as those based on social or professional interaction, mediation and problem-solving.

The response by Hargreaves to Kogan's paper was an excellent example of how we may rethink the impact of knowledge on practice in a very different way from the traditional one. The coexistence of different approaches and models can be extremely fruitful in that respect. The combination of some of them can help us to understand and explain the contrasts and conflicting evidence I referred to in the first section.

My third suggestion deals with the need to relocate educational research in

265

the continuum of instruments used for improving our knowledge base in education. Research is generally considered a significant tool for that purpose, but no longer the only one. The recognition of this new situation specially affects the way in which researchers think about their job. We have already seen that policy making uses a variety of tools to make informed decisions. The new demands made on the producers of knowledge and the expansion of the sources of knowledge imply the need to rethink educational research, not as a separate activity but interconnected with other ones. This also results in a need to stress its objectives and not only its form.

As we look for a clear position for research in the above-mentioned continuum, some new images can be helpful. It has been very common to couch discussion in terms of bottom-up or top-down strategies. But some other images have proved useful in breaking that counter-productive approach. Among them, the images of a web — both the spider's and the world-wide web — and a network seem to offer more possibilities to rethink our task.

A final suggestion consists in accepting the need to integrate the permanent tension among policy orientation, research and practice. There are new avenues here for relocating research and to rethinking its role. In our world, the weight of what has been called the *civil society* is growing. NGOs, foundations, unions, associations play a significant role in many countries in structuring social action. Sometimes they occupy a privileged place in regard to addressing some societal problems.

The possibility of tightly anchoring educational research in this flourishing field could help to give it a new position. If conflict between research and policy making is a long-standing feature, we should try to find some solutions to that situation. The first one could be learning to deal with conflict, integrating it in the way we perceive and interpret our job. The second, complementary one may be to introduce a third player to the field — the civil society — relaxing the exclusive relation with governments. This is probably not an easy way, but it is a possibility for the future. It is our responsibility to explore it.

I know that many of my suggestions will need a long time before they produce some visible outcomes. But the current, fluent situation demands our reaction. Inaction is not an acceptable solution to the current uncertainty and malaise.

References

FULLAN, M. (1993). *Change Forces. Probing the Depths of Educational Reform*. London: The Falmer Press.

HOUSE, E. (1993). *Professional Evaluation. Social Impact and Political Consequences*. Newbury Park, London and New Delhi: SAGE.

OECD (1995). *Educational Research and Development: Trends, Issues and Challenges*. Paris: OECD.

OECD (1996). *Knowledge Bases for Education Policies*. Paris: OECD.

REIMERS, F., McGINN, N. and WILD, K. (1995). *Confronting Future Challenges: Educational Information, Research and Decision-Making*. Paris: UNESCO.

VELAZ DE MEDRANO, C. (1996). Imagen de la ciencia, prácticas y hábitos científicos de los investigadores en Ciencias de la Educación. Unpublished PhD thesis, Universidad Complutense de Madrid.

Closing reflections

Malcolm Skilbeck, Organisation for Economic Cooperation and Development[1]

I warmly congratulate and compliment Seamus and the whole NFER team for the design, the planning and the execution of this splendid event. This has been a conference to enjoy, to learn from and not to forget. It has been an extraordinarily rich occasion: in range of topics, quality of presentations, and breadth of experience and expertise of participants. With the contributions from those chairing sessions, we have had more than 30 presentations on many different aspects of the educational research enterprise. For these three days the NFER has been the IFER, the International Foundation for Educational Research, and in opening itself in this way to the international community it provides a model.

I have been asked to offer some reflections on the conference. We will all have our own and, if yours are like mine, they will require time to assimilate. As Albert Beaton said of Chris Whetton's paper, one needs to read and to ponder the texts. There has scarcely been time to take in all that has been said and written but now publication provides the opportunity for us to do so.

Now, what can we say we have learnt or to have had the opportunity to learn? In the spirit of research and inquiry I will say, first, that whatever we have learnt is still subject to critical self-reflection: the truths that we may have garnered here are provisional, subject to verification. Our insights, even if they have completed Lonergan's circle of knowledge and understanding about which Seamus Hegarty spoke so eloquently, are provisional. As Seamus said on Monday, we have had the luxury of 48 hours of focused reflection on the essentials of our craft. The conference, consciously or not, has pursued the lines opened by Seamus in a searching, wide-ranging introductory address.

The fundamental question we have been wrestling with is what kinds of knowledge do or might underpin educational action and what must we do in order to have confidence in that knowledge. Moreover, knowledge in the domain of education is inseparable from the perceptions and experience of actors. They, and the values and expectations associated with them, must be taken into account in considering just what it is we know and the uses we can make of that knowledge. In other words, communities of users and contexts of use frame our knowledge and are involved in its verification.

Seamus, however, threw down a gauntlet in his opening remarks, or rather two gauntlets which no one picked up, at least not publicly. He told us that the teachers-as-researchers movement is flawed and that, as far as universities are concerned, the von Humboldt ideal of the research basis for teaching is no longer valid if indeed it ever was. He argued for concentration not dispersion of research. Does this sound like special pleading for dedicated research centres or foundations? Or is it the basis of a more powerful policy framework for national research endeavour than now exists?

It is a pity there was not an opportunity to debate this issue in the course of the conference. It is indeed relevant to higher education policy as much as to research policy not only for the UK but for all countries. It certainly deserves closer attention than we have given it, perhaps in another meeting. There is a great deal of confusion and many extravagant claims are made regarding the teaching–research relationship. For example, if research were a necessary basis for university teaching how do we account for the fact that many teachers are not themselves researchers or even part of an active research community? There is, of course, an answer to this question but I have no time to consider it now. But in a comparative review of tertiary education in ten countries that we are conducting in the OECD, we do take up this and related issues, in arguing that universities and colleges need to focus more sharply on the teaching role and to improve the quality and effectiveness of teaching. We do not, however, suggest that research is irrelevant to this.

Whatever the outcomes of the debate over the teaching–research nexus, which can be seen most clearly in higher education, many speakers in these three days have affirmed that research in its many diverse forms — and parenthetically I will add that we need to go still further in analysing the concept and delineating those forms — many speakers have affirmed that research in education needs to pay more attention to what I think of as *applicable* knowledge. This is true whether the research is, in Barry McGaw's terms, basic or applied or experimental/developmental. All three have their uses and applications, and while these differ radically we need to be clear about what they are and better able to communicate what I shall call the uses of knowledge to sceptical audiences.

What we might mean by the 'uses of knowledge' is, as Arnold Spee said, assisted by nationally defined priorities to which different kinds of research will contribute in different ways. This is a reminder that the policy–research relationship calls for yet further analysis, and several speakers dwelt on it. Susan Fuhrman argued that we are not helped when politicians adopt narrow approaches to the uses and impact of research. True, but then how effectively

are we presenting our work in the political arena? One of the richest lines of thought in the conference was the attempt several speakers made to clarify the ways in which research does and can impact on policy and action in the field. On the one hand, Susan Fuhrman pointed out, policy makers need to make the effort to learn the lessons of research. On the other hand, the research community needs to make its own intentions, procedures and results clearer. The most interesting point in this regard, however, is not communication — where there have been substantial improvements — but the emphasis several speakers gave to the time-worn but still crucial concepts of dissemination, impact and utilisation to which I return later. Not that communication is or can be straightforward. Maurice Kogan's excursion into health sciences and David Hargreaves' commentary underlined the conditions under which ideas and conclusions from research gain currency: they include complex sociocultural, political and economic processes.

Contexts must be understood and that means analysis, interpretation and judgement. Gérard Bonnet reminded us that the facts do not speak for themselves: explanatory data on why results or facts are as they are must also be provided. Why is it that many students fail or underperform? What, for that matter, does this apparently simple concept of 'failure' mean? Which contextual factors are important in explaining failure at school is a subject of controversy, but in Gérard's words the problem is still the black box of the classroom with the associated weakness of lack of a culture of self-evaluation. Both he and Suzsan Fuhrman called for a new breed of knowledge entrepreneurs or honest brokers. But who are these people? The answer is that every speaker at this conference is a knowledge entrepreneur. We have had several brilliant examples of the art. Nevertheless, I think we missed something in not being able to discuss exactly what this role is and how it could be better performed, not just in this charming setting on the banks of the Thames but in our everyday working environments.

In the emerging educational market-place of which Ralph Tabberer reminded us, there is scope for new kinds of entrepreneurs. Perhaps we have given rather too much attention to the established institutions — the specialist research agencies and the higher education institutions — and too little to the looser configurations of consultants, consultancy firms and the private sector more generally. There are also roles being played by inspectors, local advisers, key people in schools and the media which merit closer analysis.

Related to the idea of a network or loose groupings or associations of knowledge brokers is the issue of concentration of research expertise, a point I mentioned in commenting on Seamus's paper. Funding policies for research

in market economies may invite more rather than less diversity of providers or at least a shift in our views about the nature of concentration. We need to rethink the concept of critical mass. That mass might just as well be achieved through coalitions including relatively minor players and temporary systems as by the aggregation of staff and resources in large, special-purpose institutions. In looking ahead we shall certainly need to give more thought to structures to sustain research. We need also to study, now, the emerging patterns in other research domains than education. The value of looking over the fence to other public sectors was a point in Wynne Harlen's address and the subject of Maurice Kogan's overview of power–epistemology relations in the health domain. I would go further and extend the comparative analysis to study industrial and commercial research models and practices.

Recent academic studies of new trends in R&D strategies and procedures are also highly instructive in their questioning of linear models, a topic taken up here and there during the past three days. Several speakers provided very significant leads in the analysis of fundamental strategies in research. Barry McGaw's dichotomisation of the natural science and interpretive paradigms gave an initial lead which others took up. Barry's own methodological preference, as befits the head of a large and diverse national research centre, was to be eclectic. Different methods and procedures can cohabit and indeed inform one another. A client orientation, where work depends on contracts, perhaps encourages a readiness to tailor procedures to client specifications and circumstances which may be quite variable. But, how far are we prepared to go with eclecticism? Barry told us about political pressure on a university-based researcher. 'Eclecticism' in this instance is not a matter of paradigms, but whether or not procedures, results, conclusions and their attendant publicity are palatable to policy makers responsible for public education systems. This raises the issue of what I shall call the critical theory of research, namely that research by its nature is sceptical and critical, that it disconfirms as much it may confirm beliefs and practices.

The nature of educational research from the standpoint of continental traditions of critical inquiry was alluded to by Karl-Heinz Gruber in his witty comments on this subject from an Austrian perspective. This reminded me that not only the Viennese Sigmund Freud but also Karl Popper and Ludwig Wittgenstein are part of a culture of critical thought that is all too seldom drawn upon in our discussions of research, even though they remain seminal figures in other, related domains.

I wish to underline the need for research to keep its intellectual roots well nourished. The issue of criticism may also be approached through

consideration of client relationship. Clients are defined principally as policy makers and professionals in the field. We have been reminded that there is also a wider public interest, important if diffuse. Client relationships are well worth examining more closely. They are not reducible to a one-way process of satisfying expressed demand or need. If the fundamental objective of research is the construction of knowledge and its critical reconstruction we need to pay attention to the function of research as the destroyer or challenger of myths and delusions, and the critic at least potentially of policy and practices. It is difficult, when in receipt of grants that sustain jobs, to be seen to bite the feeding hand but I think there is something of a dilemma here: how far are our paying clients and those to whom we are accountable prepared to go in receiving critical knowledge? The dialogue between researchers, policy makers and practitioners that several speakers referred to cannot be bland or uncritical — but it requires great sensitivity and sophistication on all sides to sustain the need for action and the need for a critique of action in some kind of dynamic balance.

In this context, Sally Brown's reference to enlightenment is instructive. We know that Scotland or rather Edinburgh was, in the eighteenth century, one of the great European centres of Enlightenment. I do not think Adam Smith or David Hume would be too pleased to have enlightenment characterised as 'an untested Scottish hypothesis' which is what I think I heard someone say! The Scottish educational system must have contributed to eighteenth century enlightenment as to today's intellectual life in Scotland, and these considerations must have something to contribute to the testing of the hypothesis — with positive results, as several historians have argued.

What, you might ask, is the point of this talk about enlightenment? It was Sally Brown who argued that the impact of research needs to be conceived in terms of enlightenment or the diffusion of ideas rather than as directly instrumental in policy construction. I warmed to this, as to her emphasis on the adventure of ideas and on theory. There is, after all, a rich vein of historical and comparative research in education whose purpose is not to influence policy or improve teaching practice but to contribute to the culture, the global awareness, understanding and the insights of readers including students.

But enlightenment is not of course a gift bestowed on benighted policy makers by highly illuminated researchers. As Neville Postlethwaite informed us, a common language must be sought if communication is to be effective and the questions to be answered are as much those of the policy maker as of the researcher and I would add, as I am sure he would too, there are practitioners' as well as policy makers' questions. This point brings up again the issue of

client relations and the nature of dialogue that is both client-oriented, critical and radical. We come back to the knowledge brokers or perhaps to the idea that the education of policy makers, field practitioners and researchers alike is deficient in that it gives us few insights into communication and little if any practice in the necessary skills.

Chris Whetton's paper caused me to think about communication and, as an element of their own professionalism, the exchange of ideas, at about the point in his address when he discussed the role of the Assessment of Performance Unit in England and Wales. Something of a watershed appears to have occurred in the early '80s as the topic of assessment moved from the technical domain — which he characterised as the psychometric era — into the wider, more diffuse domain of teacher professional practice during a period of wide-ranging reform of schooling. I regret that there was not time to take this very interesting theme further. Chris himself wished to clothe assessment technology with the preoccupations and dynamics of contemporary culture and that is a rather large topic! In his response, Al Beaton wished to focus the debate on validity, although if validity in Cronbach's terms means not only 'looking good' but also '*doing* good', I am not sure that the focus will be such a sharp one after all. That is yet another topic which deserved further discussion and would have benefited from the kind of highly informed debate this audience was capable of giving it. Caroline Gipps started us off by arguing for assessment by teachers as an element of their own professionalism, taking us from a paradigm shift to a seismic shift, a nice metaphorical blend of the social and the natural sciences. We certainly had the making of a first-class debate in the exchange of views on assessment — a debate which it would be particularly helpful to put into the historical, dynamic context to which Chris drew attention in his substantial paper.

It was Karl-Heinz Gruber who challenged the ahistorical, ethnocentric tendency in another area of research inquiry, namely the research review. Wynne Harlen had presented a compelling argument for greater use of this form of collective, integrated knowledge, and David Raffe had agreed with her while challenging Wynne's rigorous classification of the less and the more reliable forms of evidence. But Karl was disappointed that classic texts were being overlooked in the quest for contemporaneous knowledge. Not only that, the English-language researchers are tending to neglect relevant knowledge in other languages. However, David Hargreaves' comments on Maurice Kogan's paper were certainly an exception to that generalisation. The reasons for what may nevertheless be a pervasive ethnocentrism in research reviews include but go beyond lack of familiarity with both the work and the language in which research reports have been written. An international

gathering such as this one would be well placed to analyse this topic which will become increasingly important in a globalised world.

But what seemed to be a more pressing set of issues arose in the presentations from several of the major international organisations. What we were forcibly reminded of is the disturbing contrast between the riches of the richer countries and the poverty of the poorer ones. It must be of great concern to all of us that while the richer countries of the world are investing substantial amounts of money and expertise in constructing what are in effect quite esoteric educational league tables, drawing on large, complex data sets whereby they position themselves in relation to one another, the poorer countries lack the capacity or resources to gather even the most basic information. From a global perspective this is a widening gap and even a dangerous imbalance. Do we lack an ethical and a prudential basis for international educational research policy? I do not mean what is normally meant by the ethics of research, namely procedural principles that govern practice, but a reflective critique of the conditions, world-wide, which facilitate or inhibit the development of an international knowledge base. I do not mean to suggest that this is a simple issue, only that there are dilemmas, contradictions and imbalances that deserve greater attention than they are receiving at present. There is need for the international organisations to get together, not only on technical matters — such as the current revision of the International Standard Classification of Education — but also on the no less important matters of strengthening the capability of those many educational systems which lack systematic knowledge bases. Some steps have been taken in this regard; more are needed.

The conference moved on again without debating this topic, which we must reserve for another occasion. From the last round of presentations and responses, which sustained the high level of analysis of the previous days, I have noted three of the many interesting points that speakers made.

The first is the discussion of relationships between knowledge, power and its agents, sociocultural contexts, and different fields of professional practice. These were revealed to us in the session launched by Maurice Kogan and then in their responses by David Hagreaves and Margaret Wang. In these discussions of quite complex patterns of relationship and a wide repertoire of issues, we were edging well into the domain of applied social philosophy and that certainly strengthened our emerging conceptual map. Further work, with more diverse examples to consider, would be of considerable interest. On this set of issues we should involve colleagues from the other domains of social science and from philosophy.

The second point I noted is the growing awareness, especially in the USA, of the need for systemic reform in which research outcomes and procedures are captured in what Ramsey Seldon called the horizontal and vertical lattice of decision making. Systemic reform is a powerful hypothesis; at one level, it is a matter of common sense to coordinate and seek to integrate different policy measures into coherent strategies. But are there perhaps features of the dynamic and varied structures and decision points that make systemic reform impossible to achieve except marginally? This is a matter requiring further investigation than it has thus far received.

The third of my observations from the concluding session is to welcome Maurice Kogan's and Ulf Lundgren's plea for more theory, more global analysis, more connections among diverse elements in our field of action. Maurice seemed to be seeking a unified action theory to address basic human needs. Ulf pointed to the frustration of policy makers who need insights into the changing world but receive, instead, lengthy reports on research methodology.

This conference was not intended to draw conclusions or to make recommendations. Its value has been of a different kind, to reflect upon directions and needs in educational research through an interchange of ideas among highly experienced and knowledgeable people in the international research community. It is often said that the value of conferences lies in what is achieved in the corridors, in the breaks between the formal sessions. There is truth in this but for me the test is whether or not the conference is generative, whether it set in motion a new dynamic of thought and action. According to that test, this has been a highly successful conference: so many rich lines of inquiry and discussion have been opened up and analysed with great perspicacity. The Foundation can take much satisfaction from having been a far-sighted agency as well as a generous host. Thank you, Seamus, and to all who have contributed to this highly successful occasion.

---

[1] The views expressed are personal; they do not commit the OECD.

List of participants

Christine Agambar, Office for Standards in Education, UK

Hermann Avenarius, German Institute for International Educational Research, Germany

Richard Bartholomew, Department for Education and Employment, UK

* **Albert E. Beaton,** Boston College, USA

* **Jarl Bengtsson,** OECD, France

Virginia Berkeley, Department for Education and Employment, UK

Michael Beveridge, University of Bristol, UK

* **Gérard Bonnet**, Ministry of Education, France

Patricia Broadfoot, University of Bristol, UK

Audrey Brown, Department for Education and Employment, UK

* **Sally Brown,** University of Stirling, UK

Christiane Brusselmans, University of Ghent, Belgium

James Calderhead, University of Bath, UK

* **Bàrtolo Paiva Campos**, Institute of Educational Innovation, Portugal

* **Nancy Cole,** Educational Testing Service, Princeton, USA

Phillipa Cordingley, Teacher Training Agency, UK

Paul Croll, University of Reading, UK

Richard Daugherty, University of Wales, Aberystwyth, UK

Ingemar Fägerlind, Institute of International Education, Sweden

Lindsey Fidler, Economic and Social Research Council, UK

Isaac Friedman, The Henrietta Szold Institute, Israel

* **Susan Fuhrman**, University of Pennsylvania, USA

Maurice Galton, University of Leicester, UK

José Luis García Garrido, National Institute for Quality and Evaluation (INCE), Spain

Caroline Gipps, University of London Institute of Education, UK

* **Marit Granheim,** Ministry of Education, Norway

* **John Gray,** Homerton College, Cambridge, UK

Armin Gretler, Swiss Co-ordination Centre for Research in Education, Switzerland

* Karl-Heinz Gruber, University of Vienna, Austria

* David Hargreaves, University of Cambridge, UK

* Wynne Harlen, Scottish Council for Research in Education, UK

Nicole Harrison, NFER

David Hawker, School Curriculum and Assessment Authority, UK

* Seamus Hegarty, NFER

Georges Henry, University of Liege, Belgium

Tom Kellaghan, Educational Research Centre, Ireland

* Maurice Kogan, Brunel University, UK

Kimmo Leimu, University of Jyväskylä, Finland

* Ulf Lundgren, National Agency of Education, Sweden

Tony Martin, Department for Education and Employment, UK

* Barry McGaw, Australian Council for Educational Research, Australia

* Anne Meade, New Zealand Council for Educational Research, New Zealand

Anthea Millett, Teacher Training Agency, UK

* Peter Mortimore, University of London Institute of Education, UK

Roger Murphy, University of Nottingham, UK

Nikolao Nikandrov, Russian Academy of Education, Russia

* Maris O'Rourke, World Bank, Washington, USA

Francesc Pedró, Open University of Catalonia, Spain

Hans Pelgrum, University of Twente, The Netherlands

Luce Pépin, EURYDICE, Belgium

Maurice Peston, NFER

* Tjeerd Plomp, University of Twente, The Netherlands

* Neville Postlethwaite, University of Hamburg, Germany

Richard Pring, University of Oxford, UK

Michelle Proux, National Institute for Educational Research (INRP), France

Helen Quigley, Nuffield Foundation, UK

* David Raffe, University of Edinburgh, UK

David Reynolds, University of Newcastle, UK

Kathryn Riley, Roehampton Institute, UK

Colin Robinson, School Curriculum and Assessment Authority, UK

Martyn Roebuck, Scottish Office Education and Industry Department, UK

Jim Rose, Office for Standards in Education, UK

Patricia Rowan, *Times Educational Supplement,* UK

Jean Rudduck, Homerton College, Cambridge, UK

Lars-Henrik Schmidt, Danish Institute for Educational Research, Denmark

* Erhard Schulte, DG XII, Commission of the European Communities, Belgium

* Ramsay Selden, Educational Statistics Services Institute, USA

Helen Simons, University of Southampton, UK

* Malcolm Skilbeck, OECD, France

Alan Smithers, Brunel University, UK

John Smyth, UNESCO, France

Bridget Somekh, Scottish Council for Research in Education, UK

* Arnold A.J. Spee, Ministry of Education , The Netherlands

* Sheila Stoney, NFER

* Ralph Tabberer, NFER

Alejandro Tiana, Spanish Open University, Spain

Hans Wagemaker, Ministry of Education, New Zealand

* Margaret C. Wang, Temple University, USA

Ivor Widdison, Council of Local Education Authorities, UK

* Chris Whetton, NFER

Dylan Wiliam, Kings College, UK

---

* Presenter, respondent or chair

12.30-1.30 pm	**Registration** (a light lunch will be served)	

1.45-2.00 pm **Opening**
Lord Peston
NFER
Malcolm Skilbeck
OECD

1 **2.00-3.00 pm** **Educational research in context**
Seamus Hegarty
NFER

3.00-3.30 pm **Tea**

2 **3.30-5.30 pm** **Educational research within national education systems**

Chair
Sheila Stoney
NFER

National case studies
England/Wales - **Ralph Tabberer**
United States - **Susan Fuhrman**
France - **Gérard Bonnet**
Netherlands - **Arnold Spee**

7.00 pm **Reception**

7.45 pm **Dinner**

3 9.00 - 10.30 am   **Nature and function of educational research**
Barry McGaw
Australian Council for Educational Research

Chair
Peter Mortimore
University of London Institute of Education

Respondents
Sally Brown
University of Stirling
Neville Postlethwaite
Hamburg University

10.30 -11.00am   **Coffee**

4 11.00 -12.30 pm   **The psychometric enterprise**
Chris Whetton
NFER

Chair
Nancy Cole
Educational Testing Service

Respondents
Albert Beaton
Boston College
Caroline Gipps
University of London Institute of Education

12.30 -1.45 pm   **Lunch**

5 **1.45-3.15 pm**   **Educational research and educational reform**
Wynne Harlen
Scottish Council for Research in Education

Chair
Anne Meade
New Zealand Council for Educational Research

Respondents
David Raffe
University of Edinburgh
Karl Heinz Gruber
University of Vienna

3.15-3.45 pm   **Tea**

6 **3.45-5.45 pm**   **International educational research**

Chair
Bàrtolo Paiva Campos
Institute for Educational Innovation, Portugal

OECD	- **Jarl Bengtsson**
EU	- **Erhard Schulte**
IEA	- **Tjeerd Plomp**
UNESCO	- **John Smyth**
World Bank	- **Maris O'Rourke**

7.00-7.45 pm   **MAGNIFICAT**

8.00 pm   **Dinner**

7 **9.00-10.30 am**   **Perspectives from adjacent research fields**
Maurice Kogan
Brunel University

Chair
Marit Granheim
Norwegian Ministry of Education

Respondents
David Hargreaves
University of Cambridge
Margaret Wang
Temple University, Philadelphia

10.30-11.00 am   **Coffee**

8 **11.00-12.30 pm**   **Impact and dissemination**
Ramsay Selden
Statistics Service Institute
Washington

Chair
Ulf Lundgren
National Agency for Education
Sweden

Respondents
John Gray
Homerton College, Cambridge
Alejandro Tiana
University of Madrid

9 **12.30-1.00 pm**   **Closing reflections**
Malcolm Skilbeck
OECD

1.00 pm   **Lunch**

Conference ends

nfer

The role of research in mature education systems

The classroom has taken over from the factory floor as the place where the prosperity of nations is determined. In a knowledge-based society, educating the young – and everybody else – is a vital concern.

What role does educational research have to play in this? How can it contribute to educational reform? What lessons can it learn from practice in adjacent research fields? Are there key concerns shared across countries?

These and related questions are addressed here. Many of the leading researchers from Britain and around the world came together at Windsor to take stock and look to the future on the occasion of the NFER's 50th Anniversary. This book distils the conference presentations and debates into a cogent statement on educational research in our time. Challenging examples, clear-sighted analysis and persuasive vision combine to produce an authoritative and provocative text.

Researchers, academics, educational planners and all those concerned with maximising the return from educational investment will find this a stimulating and productive read.