Improving Subject Teaching

In many countries, questions are being raised about the quality and value of educational research, and whether educational practice can ever draw upon research evidence as productively as in fields such as medicine. This book explores the relationship between research and practice in education, using the case of science education as an example. It looks at the extent to which current practice could be said to be *informed* by knowledge or ideas generated by research – and at the extent to which the use of current practices or the adoption of new ones are, or could be, supported by research evidence – and so be said to be *evidence-based*. The issues considered are not specific to science, but apply to the teaching and learning of any curriculum subject.

The book draws on the findings of four interrelated research studies, carried out by the Evidence-based Practice in Science Education (EPSE) Research Network. It considers:

- how research might be used to establish greater consensus about curriculum;
- how research can inform the design of assessment tools and teaching interventions;
- teachers' and other science educators' perceptions of the influence of research on their teaching practices and their students' learning;
- the extent to which evidence can show that an educational practice 'works'.

The book is unique in exploring the issues raised by the current debate about educational research within the context of the teaching and learning of a specific curriculum subject. Rather than looking at how research might inform educational practices in the abstract, it looks at how research can lead to improvement in the teaching of specific pieces of knowledge, or specific skills, that we value. The issues it explores are therefore of direct interest and relevance to educational practitioners and policy-makers.

Robin Millar is Professor of Science Education, University of York, UK. **John Leach** is Professor of Science Education, University of Leeds, UK. **Jonathan Osborne** is Professor of Science Education, King's College London, UK. **Mary Ratcliffe** is Professor of Science Education, University of Southampton, UK.

Improving Learning TLRP

Series Editor: Andrew Pollard, Director of the ESRC Teaching and Learning Programme

Improving Learning How to Learn: Classrooms, schools and networks
Mary James, Paul Black, Patrick Carmichael, Mary-Jane Drummond, Alison Fox, Leslie Honour, John MacBeath, Robert McCormick, Bethan Marshall, David Pedder, Richard Procter, Sue Swaffield, Joanna Swann and Dylan Wiliam

Improving Literacy by Teaching Morphemes
Terezinha Nunes and Peter Bryant

Improving Schools, Developing Inclusion
Mel Ainscow, Alan Dyson and Tony Booth

Improving Subject Teaching: Lessons from research in science education
Robin Millar, John Leach, Jonathan Osborne and Mary Radcliffe

Improving Workplace Learning
Karen Evans, Phil Hodkinson, Helen Rainbird and Lorna Unwin

Improving Subject Teaching

Lessons from research in science education

Robin Millar, John Leach, Jonathan Osborne and Mary Ratcliffe

with Jaume Ametller, Hannah Bartholomew, Sue Collins, Rick Duschl, Vicky Hames, Andy Hind, Jenny Lewis and Phil Scott

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2006
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Ave, New York, NY 10016

Routledge is an imprint of the Taylor and Francis Group, an informa business

Typeset in Charter ITC and StoneSans by
Keystroke, Jacaranda Lodge, Wolverhampton
Printed and bound in Great Britain by
MPG Books Ltd, Bodmin

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging in Publication Data
A catalog record for this book has been requested

ISBN10: 0–415–36209–1 (hbk)
ISBN10: 0–415–36210–5 (pbk)

ISBN13: 978–0–415–36209–2 (hbk)
ISBN13: 978–0–415–36210–8 (pbk)

Contents

Illustrations

Figures

Tables

Contributors

Robin Millar is Professor of Science Education in the Department of Educational Studies at the University of York. He was Co-ordinator of the Research Network Towards Evidence-based Practice in Science Education (EPSE). He directed EPSE Project 1 and co-directed Project 4.

John Leach is Professor of Science Education in the Centre for Studies in Science and Mathematics Education (CSSME) at the University of Leeds, and Head of the School of Education. He was a core member of the Research Network Towards Evidence-based Practice in Science Education (EPSE). He directed EPSE Project 2 with Phil Scott, and co-directed Project 4.

Jonathan Osborne holds the Chair of Science Education at King's College London. He was a core member of the Research Network Towards Evidence-based Practice in Science Education (EPSE). He directed EPSE Project 3 with Mary Ratcliffe, and co-directed Project 4.

Mary Ratcliffe is Professor of Science Education in the School of Education at the University of Southampton. She was a core member of the Research Network Towards Evidence-based Practice in Science Education (EPSE). She directed EPSE Project 3 with Jonathan Osborne, and was lead co-director for Project 4.

Jaume Ametller is a Research Fellow in the Centre for Studies in Science and Mathematics Education (CSSME) at the University of Leeds. He worked on EPSE Project 2.

Hannah Bartholomew was a Research Officer in the School of Education at King's College London, working on phase 2 of EPSE

Project 3 and on Project 4. She is now a lecturer in mathematics education at the University of Auckland.

Sue Collins was a Research Officer in the School of Education at King's College London, working on phase 1 of EPSE Project 3. She is now a Senior Research Officer at the Institute of Education University of London.

Rick Duschl was the Chair of Science Education at King's College from 1998 to 2003 and is now a full professor at Rutgers University, New Jersey. He contributed to phase 1 of EPSE Project 3.

Vicky Hames was a Research Officer in the Department of Educational Studies at the University of York, working on EPSE Projects 1 and 4. She now teaches in a primary school in North Yorkshire.

Andy Hind was a Research Officer in the Centre for Studies in Science and Mathematics Education (CSSME) at the University of Leeds, working on EPSE Projects 2 and 4. He is now Head of Science at a comprehensive school in Leeds.

Jenny Lewis is a Senior Lecturer in Science Education in the Centre for Studies in Science and Mathematics Education (CSSME) at the University of Leeds. She worked on EPSE Project 2.

Phil Scott is Professor of Science Education in the Centre for Studies in Science and Mathematics Education (CSSME) at the University of Leeds. He directed EPSE Project 2 with John Leach.

Series editor's preface

The Improving Learning series showcases findings from projects within ESRC's Teaching and Learning Research Programme (TLRP) – the UK's largest ever coordinated educational research initiative.

Books in the Improving Learning series are explicitly designed to support 'evidence-informed' decisions in educational practice and policy-making. In particular, they combine rigorous social and educational science with high awareness of the significance of the issues being researched.

Working closely with practitioners, organisations and agencies covering all educational sectors, the Programme has supported many of the UK's best researchers to work on the direct improvement of policy and practice to support learning. Over sixty projects have been supported, covering many issues across the lifecourse. We are proud to present the results of this work through books in the Improving Learning series.

Each book provides a concise, accessible and definitive *overview* of innovative findings from a TLRP investment. If more advanced information is required, the books may be used as a gateway to academic journals, monographs, websites, etc. On the other hand, shorter summaries and research briefings on key findings are also available via the Programme's website at www.tlrp.org.

We hope that you will find that the analysis and findings presented in this book are helpful to you in your work on improving outcomes for learners.

Andrew Pollard
Director, Teaching and Learning Research Programme
Institute of Education University of London

Preface

The research reported in this book has been carried out at a time when the quality of educational research is under intense scrutiny. Questions have been asked about its value to practitioners and policy makers in guiding them in the choices and decisions that they inevitably have to make. We approach these issues as researchers, but also with many years' experience of teaching, teacher education and curriculum development. We are personally convinced that research in the area we know best (science education) has produced knowledge, in the form of specific research findings and other more analytical and theoretical insights, that could, if more widely and systematically applied, lead to significant improvements in students' learning and attitudes. But equally we know that much practice takes little obvious account of knowledge generated by research, and that the evidence to back up our conviction that research could drive improvement is often less complete or compelling than we would wish. The work reported in this book stems from a desire to understand better the interface between research and practice in the context of the teaching and learning of a core subject of the school curriculum. Understanding better what research can provide, how we might 'engineer' improvements that draw significantly on research and test these in practice, and how potential users of the knowledge and insights generated by research view, understand and access research, are all necessary steps towards 'research evidence-based' or 'research evidence-informed' practice. In this book, we approach these issues from the perspective of one curriculum subject – science. They are not, however, specific to science education, but apply equally to the teaching of any curriculum subject. We therefore hope that all readers with an interest in the role of educational research will be able to see links to issues and questions in which they are interested.

This book is based on the work of the Evidence-based Practice in Science Education (EPSE) Research Network, funded by the UK Economic and Social Research Council (ESRC) in Phase 1 of the Teaching and Learning Research Programme (TLRP). The EPSE Network undertook four interrelated projects. Summaries of these can be found in the Appendix at the end of the book. In the body of the book, we have not set out to provide brief accounts of each project in turn. Instead, our aim is to draw selectively on the work undertaken by the EPSE Research Network to explore a number of general issues and themes about the actual (and potential) role of research in improving the way any subject of the school curriculum is taught and learned.

Part I

What is the issue?

Chapter 1

Research and practice in education

Education research is often criticised for having little impact on practice or policy. This book looks at the influence of research on the teaching of a curriculum subject. It asks what research can offer practitioners and policy makers in making choices and decisions about how a curriculum subject should be taught – and considers how research might need to change to influence practice in subject teaching more strongly. In this opening chapter, we try to disentangle two issues: the role of research in the design of teaching, and the extent to which research can provide a warrant (or justification) for teaching something in one way rather than another. We also summarise some of the key ideas and claims in the recent debate about the role and quality of education research, looking in particular at the suggestion that research in medicine and health provides a model for education research to follow.

Introduction

The past decade has seen a growing demand, in many countries, that education should become more 'evidence-based' – that is, that research should provide evidence to inform choices and decisions about educational matters. This demand has come particularly from those involved in educational policy making, and is part of a wider debate about the quality of public services in general, and the role of research in informing the decisions which practitioners and policy makers must inevitably make (Davies *et al*. 2000). Research, it is argued, provides a

Authored by Robin Millar, John Leach, Jonathan Osborne and Mary Ratcliffe

better basis for choice and action than tradition or 'professional wisdom'. It can challenge ineffective current practices; and provides a means of testing innovations so that only those which are seen to be effective are widely implemented. As David Blunkett, speaking in 2000 as Secretary of State for Education and Employment in the UK, put it:

> We need to be able to rely on . . . social scientists to tell us what works and why and what types of policy initiatives are likely to be most effective.
>
> (Blunkett 2000: 2)

This is very much the policy maker's perspective. How does the research–practice interface look from the viewpoint of the practitioner? Most teachers spend most of their working lives teaching a subject, or group of subjects. Most of the decisions that teachers routinely take are about the detail of how to plan and implement lessons that develop their students' ideas, skills and attitudes within a subject area. If educational research is to have an impact on practice, and be seen to have an impact, it is decisions of this sort that it must influence. Hence, it is on this issue, of the current influence of research on teachers' everyday practice and how it might be increased, that we want to focus in this book. Our reason for starting from the practitioner's perspective is not that we think the role of research in informing policy is less important. Far from it. But research-informed policy can only mean support and encouragement of practices that have themselves been shown to be soundly conceived and effective when implemented. The key questions, then, which we want to explore in this book might be summarised as follows:

- What contribution can research make to improving the teaching and learning of a school curriculum subject?
- Can research help us to improve the way subjects are taught, and the learning that ensues?
- If so, what kind of help can we reasonably expect research to provide: new information, perspectives or insights that we may need to consider and take into account; or clear and compelling evidence that certain interventions 'work' or do not?
- How should research be designed and carried out if it is to provide the kind of knowledge that can lead to practical improvements?

We will approach these questions from the perspective of one curriculum subject – science. The issues which they raise are not, however, specific

to science education. They apply equally to the teaching of any curriculum subject. We have chosen to focus on science because that is the subject we know best. But it is only an example. We hope that readers who are interested in the teaching and learning of other subjects will be able to relate the discussion to their interests.

To argue that the teaching and learning of science (or of any other curriculum subject) should become more 'evidence-based' implies that research is capable of generating knowledge that makes possible the more effective teaching of specific bodies of knowledge and the skills and attitudes associated with them. The aim of this book is to explore critically the extent to which such a view can be sustained, identify the conditions which would have to be met to make it possible, and explore its strengths and limitations as a strategy for improving practice. The book arises from the work of the Evidence-based Practice in Science Education (EPSE) Research Network, funded by the UK Economic and Social Research Council (ESRC) within the Teaching and Learning Research Programme (TLRP). In the original research proposal, the Network's title was 'Towards evidence-based practice in science education'. That 'towards' is important despite its subsequent omission from the Network's shorthand name; indeed the title should perhaps be read with a question mark at the end. The aim of the Network was to improve our understanding of the interface between science education research and practice, and to explore what might be involved in moving school science education towards forms of practice that might be termed more 'evidence-based'. To this end, the Network undertook four projects, which are interrelated and were carried out in parallel rather than sequentially:

Project 1: Exploring the impact on teachers' practices, and student learning, of providing diagnostic tools informed by previous research on student learning.

Project 2: Developing and evaluating short teaching sequences that are explicitly informed by research evidence.

Project 3: Establishing the extent of 'expert' agreement on what people should know about the nature of scientific knowledge and the procedures of scientific enquiry, and exploring the practical implications of trying to teach the ideas seen as most important.

Project 4: Documenting and analysing science education practitioners' views on the actual and potential influence of research on their everyday practices.

Summaries of each of these projects are presented in the Appendix. Our intention in this book, however, is not simply to report and discuss the findings and outcomes of each in turn, but to use them together to explore a number of general issues and themes about the nature of research evidence in education and its relationship to practice.

Research, evidence and practice

To explore systematically the relationships between research and practice, and the ideas of 'evidence-based' or 'evidence-informed' practice, we need to clarify some terms, and develop a framework for exploring systematically the relationship between evidence (or knowledge) and practical choices and decisions. The task of developing and testing a comprehensive model of the complex interrelationships between research and practice is beyond the scope of this book. What we need is a model that is 'good enough' to provide a framework for discussion of the issues.

First, though, a word about the two terms 'evidence' and 'practice'. To call for education to become 'evidence-based' seems to imply that current practices are not based on 'evidence'. This is clearly untenable if taken literally: any rational human activity that involves participants in making choices and decisions clearly requires them to draw on past experience and learning – that is, on evidence. The issue is the kind, and the quality, of the evidence that is used. 'Evidence' in the phrase 'evidence-based education' clearly means 'evidence produced by research', that is, evidence collected through a systematic process of data collection and analysis, open to public scrutiny and, in principle at least, replicable by anyone who chooses to do so.

It is also worth clarifying what we mean by 'practice'. The phrase 'evidence-based practice' treats 'practice' as though it were a single category. But in reality, there are many different kinds of educational 'practices' that we might wish to make more 'evidence-based'. Some concern general aspects of school organisation and procedure (such as decisions about class size, or about expenditure on books compared to that on computers). Others are about teaching approaches that could be applied across all subjects (such as using formative assessment) or across all topics within a subject (such as the use of small-group practical work in science). These are 'practices' that a school or a department might choose to adopt in place of other alternatives. At a more content-specific level, there are the practices that a teacher or a school department follows in teaching specific topics (for example, the teaching scheme and

materials used for teaching elementary electric circuit theory in lower secondary school) or even specific key ideas (such as the model of a chemical reaction as a rearrangement of atoms into new combinations).

A critical feature of all of these kinds of practices except the first (general aspects of school organisation and procedure) is that *content matters*. In other words, the quality of the decisions and choices made depends not only on general principles and criteria (perhaps deriving from some kind of theoretical position or analysis) but also on knowledge of and about the specific content being taught. This is clearly so for content-specific practices but is also true of generic practices that can be applied across subjects or across topics within a subject, such as formative assessment. These can only be implemented in the context of teaching some specific content – and the success or failure of the resulting practice will depend not only on the validity of the generic principle *per se* but also on how it is implemented with the chosen content. Again the decisions and choices involved depend on knowledge of and about that content. It is largely for this reason that we believe it essential to consider issues of 'evidence-based practice' from the perspective of the teaching of a specific subject.

How then might we begin to model the relationships between research evidence and practice? A central element of any such model is the knowledge base on which teachers draw in making decisions and choices. Research and informal experience both suggest that this includes a range of types of knowledge, acquired from a variety of sources (e.g. Brown and McIntyre 1993; Loughran 1999). The categories of teachers' professional knowledge identified in Figure 1.1 are based on the widely-used typology proposed by Shulman (1986, 1987). Many of these categories consist of both explicit (or declarative) knowledge – things a person could tell you that they know – and tacit knowledge, which cannot be fully expressed in words but is implicit in actions.

Teachers' professional knowledge comes from a variety of sources. Some of the more important ones are shown on the right-hand side of Figure 1.1. These are not, of course, as separate or distinct as this might seem to imply. For example, whilst a teacher may have direct knowledge of a specific research finding, the influence of research on his or her thinking may also be indirect. For instance, a teaching session in an initial or in-service teacher education programme might have been based on a specific research finding, or a general perspective that comes from research, as might the treatment of a topic in a textbook or a teaching resource. These may then influence a teacher's knowledge and

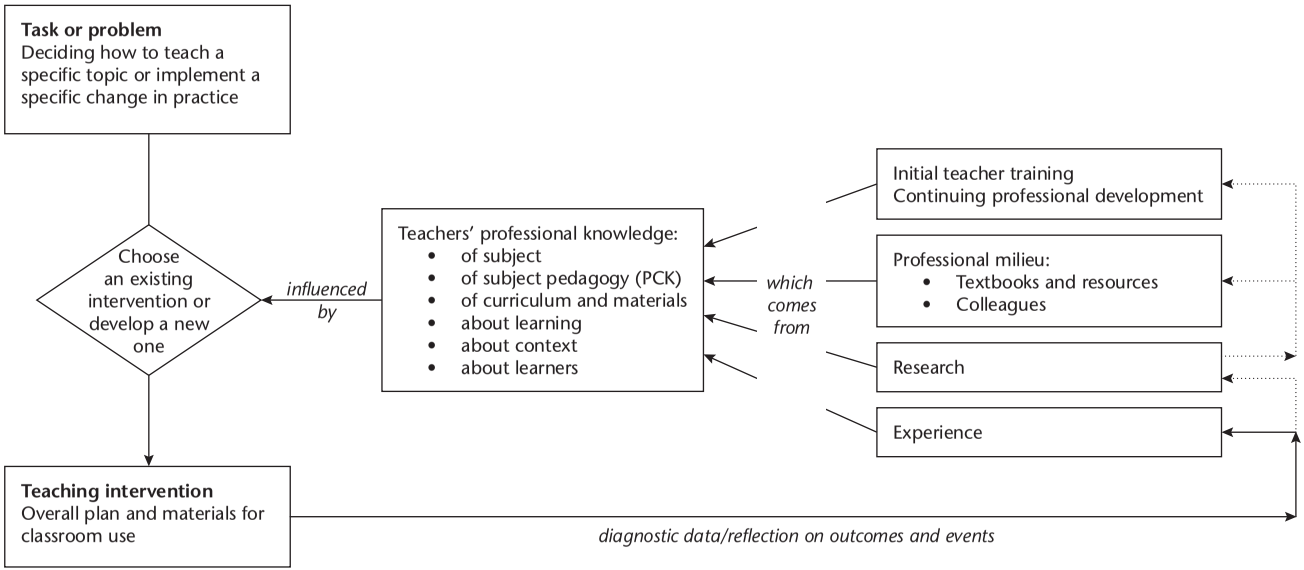

Figure 1.1 Influences on teachers' decision making

ideas, even though he or she may not recognise or acknowledge any specific research input. The dotted arrows on the right of Figure 1.1 try to represent this kind of indirect influence of research.

This complex professional knowledge base then influences the decisions that a teacher takes in deciding how to teach a specific topic or implement a specific change in some aspect of their practice. One possibility is that the teacher identifies an 'off the shelf', ready-made teaching scheme or set of teaching resources, and chooses to adopt this. This might be a published scheme or one that has been developed by a school department. It may be informed to a greater or lesser extent by research evidence. For example, the innovative short course, *Cognitive Acceleration through Science Education (CASE)* (Adey *et al*. 1996), is explicitly based on theoretical perspectives arising from research. On the other hand, recent research in the US would suggest that many school science textbooks are not significantly influenced by research findings (Kesidou and Roseman 2002). The extent of research influence may not be apparent to teachers choosing a particular intervention. And, even if they know that an intervention has been (or claims to be) informed by research, they may not be able to say what the primary research evidence is or how it has influenced the intervention. In the UK, choosing a published science teaching scheme or materials as the basis of a two- or three-year teaching programme has become increasingly common since the introduction of the national curriculum. In other countries, choosing a published 'programme' may also be the commonest way of tackling the kind of task or problem outlined in Figure 1.1.

The other possibility is that the teacher develops his or her own teaching intervention, perhaps selecting elements from a range of published materials and resources, and adapting or augmenting these using ideas from colleagues, or from initial or in-service training courses, or devised by the teacher himself or herself. Again this is quite common in UK schools. In this case, the teacher is drawing on many aspects of his or her professional knowledge, which has in turn come from some or all of the sources identified on the right-hand side of Figure 1.1. He or she may know of, or seek out, specific research findings that bear on the particular task in hand, and make considered use of these. Research findings may also have influenced the form and content of the teaching materials consulted, and used or adapted in developing the intervention – though this may or may not be recognised by the teacher in choosing to use them.

It follows that teaching interventions cannot be easily or sensibly divided into the categories 'research evidence-informed' and 'not

research evidence-informed'. This would become even more apparent were we to go on to consider the implemented intervention rather than just the intended (or planned) intervention. A further raft of choices and decisions is involved in implementing any intervention, many of which have to be taken in the heat of the moment in the classroom. Rather, 'research evidence-informed' is a matter of degree – of the extent of influence of ideas and findings from research, and the 'quality' of this influence. Any given example of practice might therefore be regarded as more or less 'research evidence-informed'.

Finally, there is another, and somewhat different, sense in which a teacher's practice might be said to be 'evidence-informed'. This is where a teacher collects data on students' learning, or responses to an intervention more generally, as it proceeds, and uses this to help make decisions about what to do next. This might be done formally, as a piece of action research, or less formally as part of normal practice. The feedback loop in Figure 1.1 is an attempt to include within the model the kind of knowledge that can come from diagnostic assessment and critical reflection on practice.

Calling a practice 'research evidence-informed' is therefore making a claim about its design. In contrast to this, we propose to call a practice 'research evidence-based' if it has been implemented and evaluated, and there is evidence of some kind, based on data that have been systematically collected and analysed, and are available for public inspection, to support the view that it achieves all (or at least some) of its intended outcomes – and perhaps that it achieves these better than some other approach (Figure 1.2). This parallels usage in medicine and health, where calling a practice 'evidence-based' is also making a claim about its outcomes. In the medical context, such a claim is only likely to be made if the evidence comes from a *randomised controlled trial* (RCT, see p. 13) or, better still, a synthesis of several RCTs. Figure 1.2, however, does not specify that any particular research strategy or research design be used to evaluate a teaching intervention, in order to support the claim that using it is an instance of 'research evidence-based practice'. The research design used to evaluate a particular intervention may, of course, have a very significant bearing on the *strength* of the evidence of its outcomes – and hence on the justification for recommending that it be used more widely. These are issues to which we will return in the concluding chapter.

We are not attempting to claim that our proposed usage of the terms 'evidence-informed' and 'evidence-based' is the 'correct' one. We recognise that these terms have been used with a variety of meanings,

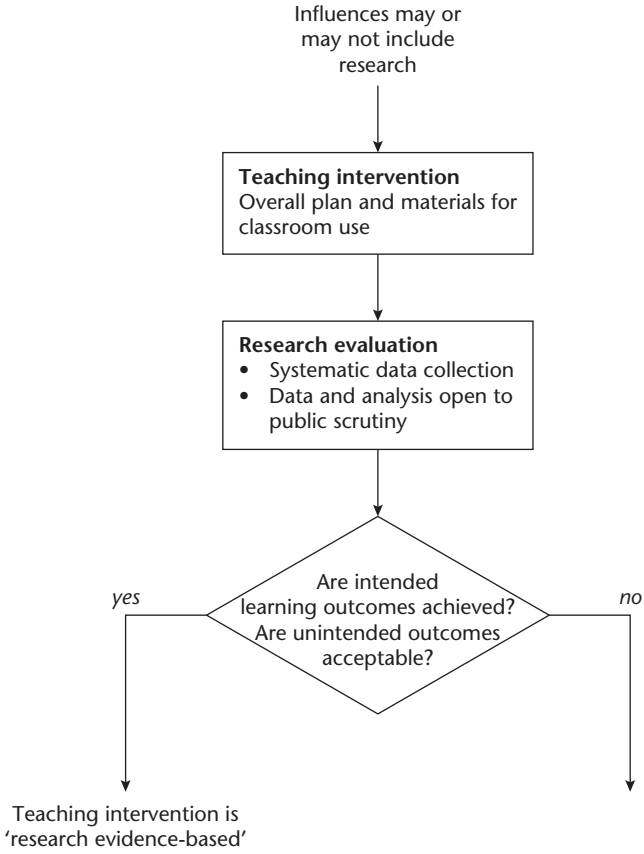

Figure 1.2 Research evidence-based practice

often indeed almost interchangeably. Rather we want to make the point that there are two distinct ideas here, both of which are important. One is about the extent to which the *design* of a teaching intervention is influenced by research findings or ideas; the other is about the extent to which research has provided evidence about the *outcomes* of the intervention, and hence a warrant for choosing to use it. Not only are these distinctly different, they are also independent. It is possible for a teaching intervention that is 'research evidence-informed' not to have been formally evaluated and shown to 'work'. Indeed, it is common. But it is also possible for a teaching intervention that is not to any great

extent informed by research, perhaps because it is based largely on pragmatic or 'commonsense' considerations, to be evaluated systematically and shown to achieve its objectives – and hence to be regarded as an example of 'research evidence-based practice'. An intervention could, of course, be both. By using terms consistently in the way we propose, it may be possible to keep these issues more clearly distinct. In this book we are interested in both.

Improving subject teaching: The role of research

This section explores the more general debate about the nature and value of educational research, and the relationship between educational research and practice – a debate that has been going on over the past decade in the UK and elsewhere. We see this book as a contribution to that debate.

The growing emphasis on 'evidence-based practice' in the public services referred to on the first page of this chapter has emerged at a time when the quality of educational research and its practical usefulness are being seriously questioned in many countries. In the UK, the Teacher Training Agency (TTA) lecture in 1996 by Professor David Hargreaves initiated a debate that still continues. Hargreaves asked if teaching could be said to be a research-based profession, and concluded that it could not. The problem, he argued, lay in the nature and quality of the outcomes of educational research:

> Given the huge amounts of educational research conducted over the past fifty years or more, there are few areas which have yielded a corpus of research evidence regarded as scientifically sound and as a worthwhile resource to guide professional action.
>
> (Hargreaves 1996: 2)

Reviews of educational research for the UK Office for Standards in Education (OfSTED) (Tooley and Darby 1998) and the Department for Education and Employment (DfEE) (Hillage *et al*. 1998) reinforced this view that the outcomes of educational research rarely provide useful guidance for policy makers or practitioners and that, as a result, 'the actions and decisions of policy-makers are insufficiently informed by research' (Hillage *et al*. 1998: xi).

Turning to the question of how the quality of educational research might be improved, Hargreaves drew a comparison with medical

research, and the growing emphasis on evidence-based medicine. First, he noted that:

> In medicine, as in the natural sciences, research has a broadly cumulative character. Research projects seek explicitly to build on earlier research – by confirming or falsifying it, by extending or refining it, by replacing it with better evidence or theory, and so on. Most educational research is, by contrast, non-cumulative. . . . A few small-scale investigations of an issue produce inconclusive and contestable findings of little practical relevance.
>
> (Hargreaves 1996: 2)

A second important difference, Hargreaves claimed, lay in the identity of the people carrying out the research: the involvement of practising doctors in medical research, including setting the research agenda, ensured that the outcomes were relevant to practice. In contrast, educational research is carried out mainly by academics who are no longer directly involved in the educational practices they are researching, so it risks becoming 'a private, esoteric activity, seen as irrelevant by most practitioners' (ibid. p. 3). Third, Hargreaves saw 'a very sharp difference in the way the two professions approach applied research. Much medical research is not itself basic research . . . but a type of applied research which gathers evidence about what works in what circumstances' (ibid. p. 2).

Not surprisingly, many of the points raised in Hargreaves's lecture and in the reviews by Hillage *et al*. and by Tooley and Darby were challenged by other educational researchers (for example, Edwards 1996; Gray 1996; Norris 1996). Hammersley (1997) accepted Hargreaves's criticism that educational research has not led to a cumulative growth of knowledge – as, it seems, do many educational researchers in the UK (Taylor 2002). He questioned, however, whether clinical medicine offers a useful model for educational research to follow, arguing that the problems of generating knowledge about education are considerable and stem from the nature of educational knowledge (and social knowledge more generally). Essentially, the issue is whether law-like generalisations about educational practices, even of a probabilistic nature, exist and can be found by careful research.

Persuasive – and highly influential – arguments for the use of *randomised controlled trials* in educational research have, however, been advanced by others sympathetic to Hargreaves's analysis of the limitations of educational research (e.g. Fitz-Gibbon 2000; Tymms and Fitz-Gibbon 2002). Slavin (2002), for example, argues that:

> Once we have dozens or hundreds of randomized or carefully matched experiments going on each year on all aspects of educational practice, we will begin to make steady, irreversible progress. Until then, we are merely riding the pendulum of educational fashion.
>
> (p. 19)

RCTs can also help to identify and exclude innovations that are worthless or damaging. Torgerson and Torgerson (2001), for example, claim that:

> There are undoubtedly some educational innovations which are harmful to the educational prospects of children. Without testing them in robust trials we simply do not know what is helpful and what is downright detrimental.
>
> (p. 323)

These are not new ideas, but ones with a long history in the field of education (Bain 1879). There are, however, few examples of trials which have had a marked influence on educational practice or policy. And there is also an important counter-argument, with an equally long history, which questions the very notion of a 'science of education', and argues that research in social settings (including education) should not seek universal generalisations, because these do not and cannot exist, but rather to illuminate and interpret events from the perspectives of those involved (Parlett and Hamilton 1972) – as the social context, which is often unique, is a major determinant of individuals' actions.

A debate about the quality and direction of educational research, similar to that in the UK, has also been going on over the past decade in the United States. Here the advocates of evidence-based education, based upon the findings of randomised controlled trials (Boruch 1997; Mosteller and Boruch 2002; Slavin 2002), have made more striking inroads into national policy. The No Child Left Behind Act (US Congress 2001), which currently dominates educational discussion in the United States, mentions 'scientifically based research' 110 times. It defines 'scientifically based research' as 'rigorous, systematic and objective procedures to obtain valid knowledge' which includes research that 'is evaluated using experimental or quasi-experimental designs', preferably with random assignment (Slavin, 2002: 15).

A specific target of the US Department of Education (2002) *Strategic Plan for 2002–2007* is that, from 2004, seventy-five per cent of the

research projects funded by the Department of Education to explore 'causal questions' should use randomised experimental designs. (By 'causal questions', they mean those which ask if a specific factor, such as the use of a specific teaching intervention, correlates with, and therefore possibly causes, a given outcome.) Other actions include setting up the *What Works Clearinghouse* (2003) 'to provide educators, policymakers, researchers and the public with a central and trusted source of scientific evidence of what works in education', and commissioning a 'user friendly guide' on educational research which includes a 'checklist to use in evaluating whether an intervention is backed by rigorous evidence' (US Department of Education 2003). Here 'rigorous' is synonymous with 'obtained from an experimental trial'. The National Research Council report *Scientific Inquiry in Education* (Shavelson and Towne 2002) can be seen as an attempt to defend a broader definition of 'scientific research' in education, which recognises the diversity of educational questions that might be investigated empirically and the range of research methods and designs that might be therefore regarded as 'scientific'.

Research designs for evaluating subject teaching

Given the prevalence of comparisons between educational research and medical research, it may be useful to say a little about the idea of 'evidence-based medicine' and the view of research on which it is based. Sackett *et al*. (1996) define evidence-based medicine as:

> The conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.
>
> (p. 71)

They are careful to make clear that this is not '"cookbook" medicine' (p. 72); professional expertise and judgement have an important role. The aim is that doctors' decisions about the treatment of individual patients should, where possible, be based on evidence that the prescribed treatment is likely to lead to better outcomes than other alternatives. Gray (2002) calls this 'evidence-based decision making' and contrasts it with 'opinion-based decision making', where doctors'

actions follow established or traditional ways of proceeding, or personal preferences based on individual experience.

In many areas of medicine, the accepted method for obtaining trustworthy evidence, on which decisions can safely be based, is the *clinical trial* – an experimental comparison of two (or more) treatments in which patients are allocated randomly to groups, each of which receives one of the treatments. The numbers of successful and unsuccessful outcomes in each group are then compared, using standard statistical methods, to assess the significance of any differences observed. The use of RCTs in medicine has its origins in trials of the efficacy of drug-based treatments. By allocating patients randomly to the treatment and control groups, the influence of other factors (both those we might anticipate and those we might not suspect) is eliminated – so that any observed difference can safely be attributed to the difference in treatment.

The application of this approach to evaluate educational interventions designed to inform or improve subject teaching, however, immediately raises some issues and questions. First, medical treatments are normally applied to individuals, whereas educational ones are typically applied to whole classes. This can be addressed by treating the class as the 'unit' that is allocated randomly to the 'treatment' or control group. One consequence, however, is to increase substantially the number of individuals who have to be involved in a study, in order to obtain clear evidence that an intervention does or does not have an effect. Second, and perhaps more importantly, subject teaching interventions are much more complex than treatments that simply involve taking a tablet (or having an injection) at regular intervals for a specified period of time. They are therefore more difficult to specify completely (to say exactly what the intervention is, so that it can be implemented reasonably consistently by different people in different places), and the outcomes are more difficult to measure. This issue is explored through specific examples in the second section of Part II (Chapters 5–7) of this book.

It should, however, also be acknowledged that many medical interventions are more complex than drug-based treatments. As the RCT has become established in medicine as the 'gold standard' for research evidence, trials have been used or proposed to evaluate a wide range of types of intervention, including strategies for modifying health professionals' behaviour, community-based health programmes, forms of individual and group therapy, and health and sex education programmes. These are referred to in the medical research literature as 'complex interventions' (Campbell *et al.* 2000). If an analogy is to be

drawn between medical research and research designed to inform or improve subject teaching, then in most cases the appropriate parallel is with 'complex interventions' in medicine, rather than with drug-based treatments.

Measurement of outcomes also poses significantly greater problems for subject teaching interventions than for many medical ones. For many medical interventions, the desired outcome is either the patient's survival for a given period, or recovery to 'normal' health. Davies (1999) argues that medical outcomes are not all as simple as this, and are often less straightforward to measure than might be supposed. But his conclusion that there is little difference in this respect between medicine and education is less than persuasive, when considered in the context of research designed to inform or improve subject teaching. Whereas medical interventions have an essentially convergent set of outcomes, aiming to take the individual from a 'deficit state' (illness) towards an agreed and identifiable 'normal state' (health), subject teaching interventions aim to take individuals from a 'normal state' to an 'enhanced state' in which their knowledge, or skills, or attitudes have been extended or developed in some way. An initial temptation is to assume that it is easy to specify enhanced knowledge outcomes that are desirable, but agreeing upon such outcomes and how they might be measured is actually far from straightforward, as we will show in Chapters 2 and 3. Furthermore, subject teaching aims to engender motivation and enjoyment amongst students, and it is hard to argue that an intervention has succeeded if it has a negative effect on motivation, irrespective of knowledge outcomes. There are clearly many forms of 'enhancement' – and correspondingly many ways of attempting to measure and assess the extent to which they are realised. We explore this issue through examples of subject teaching in Chapters 5–7. This difficulty of specifying outcomes calls into question the usefulness of the 'treatment' metaphor when thinking about educational interventions. It also draws attention to the critical importance of a sound and persuasive outcome measure. This is necessary to help clarify and define the outcome we are seeking, as well as to measure the extent to which it is achieved.

The RCT is not, however, the only research design that has been advocated in recent debates about the quality of educational research. Another much discussed approach is the 'design experiment'. This has its origins in the work of Brown (1992) and Collins (1993). Rather than using clinical medicine as a model, an analogy is drawn with engineering practice. An educational intervention (such as a teaching

scheme for addressing a particular area of a subject) is seen as an engineered 'artefact', designed to solve a particular problem or accomplish a particular task. Ultimately it will be judged a success (or a failure) on the basis of how well it works in the contexts in which it is used. Typically, the development process involves iterative improvement. Starting from an initial prototype, the artefact evolves in stages, through a series of investigations which may alternate between field settings (to see how it works *in situ*) and laboratory settings (to test specific aspects under more controlled conditions). The final goal is not a tightly defined, prescriptive intervention, but rather a design 'template' that can be used to generate interventions that 'work' in a variety of settings. In the process, the theoretical ideas that underpinned the development of the prototype are also challenged, refined and extended. So the outcome is not only a design for artefacts that 'work', but also new (or improved) educational knowledge.

One attraction of the design experiment approach is that it is clearly more useful to show that certain design features or criteria 'work', than that any one intervention based upon them does. It also offers a model of the research-practice interface in which professional insight, engagement and reflection have a central role. For these reasons, design experiments are well-suited to studies of the effectiveness of approaches to subject teaching. Indeed, it is difficult to imagine how an RCT could be designed (or resourced) to test the relative effectiveness of two detailed approaches to teaching a particular piece of subject content, as we illustrate in Chapter 5.

However, the precise nature of a design experiment, and the features which characterise the approach, are often unclear in articles written about the approach. Whilst this kind of methodological flexibility has some advantages, it is also a potential weakness. Gorard (2004), for example, comments that:

> many design experiments have been conducted by advocates of particular approaches or artefacts. . . . This makes it an especial concern that so few of their reports make any mention of comparison or control groups. Without these, and without working towards a definitive test, how can they persuade others that they have successfully eliminated rival explanations?
>
> (pp. 106–7)

This brings us back to the issue of outcomes, how we measure them, and how we evaluate and judge the significance of the results we obtain.

Whereas the results of an RCT can be logically compelling, those of a design experiment are likely to be more open to interpretation – and hence may be less likely to persuade people to change practices or policies. These are issues to which we will return later in the book.

Improving school science

Although this book is primarily concerned with general issues about the relationship between educational research on the one hand, and practice and policy on the other, the research which it reports and discusses is on the teaching and learning of science at school level. It may therefore be useful to say briefly what we consider to be the aspects of school science education that are currently a cause of concern and which therefore stand in greatest need of improvement – to set in context the research studies we go on to discuss in the remainder of this book. It might be expected that the issues that concern science educators would vary widely from country to country. In fact, there is a remarkable level of agreement. We would highlight in particular the following:

- In many countries there is a concern about the falling numbers of students who choose to study science beyond the age when it is a compulsory element of the curriculum. This is seen as having social and economic implications for the future.
- The falling enrolment in more advanced science courses is seen as a reflection of the disaffection of many students, during the compulsory phase of education, with science as a subject. Studies in many countries reveal negative attitudes towards the subject, and declining interest in it from early secondary school age onwards. Science is seen as difficult, as offering little scope for imagination or creativity, and as not valuing students' own views and ideas.
- Students in many countries report a disjunction between the science they are taught in schools and the more interesting science that they hear about on the television or read about in newspapers. Many educators recognise a need to orient school science for all pupils more strongly towards 'scientific literacy' – the kind of understanding that citizens require – rather than seeing it as a pre-professional training in the science disciplines.
- In order to deal sensibly with science-related issues that arise in everyday life, it is widely agreed that citizens need an understanding

of the nature of scientific knowledge and the processes of scientific enquiry, so that they may become more astute 'consumers' of scientific information. In most countries, however, learning goals in this area are not well defined or strongly emphasised. In particular, they are often not assessed, and hence are not prominent in teaching programmes.

- There is a growing awareness, in many countries, of the dilemma of trying to provide a science education that is appropriate for all future citizens whilst also catering adequately for the needs of those students who may choose to continue their study of science to more advanced levels. There are increasing doubts amongst science educators about the possibility of addressing adequately, within a single science curriculum, both 'scientific literacy' aims and those appropriate to the pre-professional training of future scientists.
- Research in many countries consistently shows that many students acquire little understanding of fundamental science concepts. Many retain their prior non-scientific ideas, or fail to integrate the taught ideas into a coherent framework. Most is quickly forgotten.
- Tests used to assess science performance, at all levels from schools to international surveys, are criticised for over-emphasising recall of discrete 'bits' of knowledge, rather than understanding of worthwhile ideas and fundamental concepts. Indeed, research has shown that students who score quite well on such tests often cannot give correct answers to questions probing their understanding of basic ideas.

These, then, are the sorts of issues on which science educators might look to research for assistance in solving, or at least ameliorating. Several are about values: what we ought to be trying to teach. These, it might be argued, are essentially non-empirical, though we will discuss in Chapter 2 (and to some extent in Chapter 3) how research might contribute to clarifying them, and helping to work towards better solutions. The other issues are about how best to achieve the goals we have chosen. Here research might reasonably be expected to offer guidance and assistance in improving practices and outcomes.

Overview of book

In this opening chapter, we have set out a framework for discussing and exploring the relationships between research evidence and practice, and explained the sense in which we will use the terms 'research

evidence-informed practice' and 'research evidence-based practice' throughout this book. We have then presented a brief overview of the debate that has raged internationally over the past decade about the quality of educational research, and its usefulness to practitioners and policy makers, making some specific points about how this relates to issues concerning the teaching and learning *of a curriculum subject*.

We return to these issues in the final chapter, Chapter 9, drawing on the experience and outcomes of the four research projects we carried out to address the four questions identified at the beginning of Chapter 1. Chapter 9 considers the extent to which current practice in science education could be termed 'research evidence-informed' and 'research-evidence based', and asks what might be involved in seeking to make it more so. It also explores the implications of the aspiration to have more evidence-based practice: for the design and conduct of educational research, for the ways in which research is mediated to practitioners, and for the professionalism of teachers.

If Chapters 1 and 9 bracket the book, then chapters 2–8 are (to use another metaphor) 'the meat in the sandwich'. These report and discuss in detail some aspects of the four research projects we carried out. We have chosen not to report these projects sequentially – as they were not designed to be carried out in this way and were, in fact, undertaken in parallel. Rather we have tried to use them to draw out some more general issues about the relationships between research and practice, using examples from different projects to illustrate and explore these. Chapters 2–4 look at the role of research in addressing the question: what should we teach? Here the focus is on the role of research in informing the *design* of teaching. Chapter 2 explores how research might contribute more systematically to decisions about curriculum objectives. It discusses the use of the Delphi study method to establish the nature and extent of agreement within and between different 'expert groups'. Chapter 3 then looks at how research might inform the more detailed specification of learning outcomes, by exploring issues which arise in developing and using assessment items to 'measure' attainment of important learning outcomes. Finally, in this Part of the book, Chapter 4 discusses what is involved in designing sequences of lessons that draw clearly and explicitly on research. The focus of these chapters is therefore on the development of examples of 'research evidence-informed practice', as we have characterised it in this opening chapter.

Chapters 5–7 then ask: what can research tell us about the outcomes of teaching? These chapters consider three different examples of implementation and evaluation of specific teaching interventions, exploring

the issues involved in collecting the kind of evidence that might support the claim that an intervention meets some, or all, of its objectives, and which might therefore justify recommending that others adopt it. Chapter 5 reports on the outcomes of a research evaluation of the teaching sequence whose development has been discussed in Chapter 4. Chapter 6 looks at the ways in which teachers' practices were modified and changed by providing them with access to the kinds of assessment tools discussed in Chapter 3 – and at the teachers' views on this, and on its impact on their students' learning. Chapter 7 then considers what might be involved for teachers in trying to teach some of the ideas which emerged as important curriculum objectives from the Delphi study discussed in Chapter 2. Chapters 5–7 therefore revisit the three projects discussed in Chapters 2–4, but now looking at *outcomes* of interventions, rather than issues about their design.

Chapter 8 then asks: what influences the impact of research on practice? In contrast to Chapters 2–7 which look at issues from the researchers' perspective, Chapter 8 considers the research–practice interface from the perspective of the practitioner. The chapter presents and discusses the outcomes of an interview and focus group study of the views of a large sample of teachers and other science education practitioners (such as textbook writers, examiners, and so on) on the influence of research on their day-to-day practice. It identifies the features of research which make it more (or less) persuasive to potential users, and some factors which seem to facilitate or hinder the influence of research on practice in the teaching of a specific subject. Chapter 9, as we have already said, then draws together insights from all four research projects to make some more general points about the research–practice interface.

Summary

Educational research is often seen as having relatively little impact on practice and policy. In exploring the relationship between research and practice, it is useful to distinguish between the notions of 'research evidence-informed practice' (meaning a practice whose design has been influenced by insights or findings from research) and 'research evidence-based practice' (meaning a practice for which there is evidence, from a research evaluation, that it 'works').

Some researchers have argued that research in medicine and health, centring on the use of randomised controlled trials to evaluate treatments and interventions, provides a model that education research should follow. Education differs from health, however, in the complexity of interventions and the clarity with which outcomes can be measured – and so other research designs, such as the design experiment, may be more appropriate. These issues are explored in subsequent chapters, through examples drawn from the four research projects undertaken by the Evidence-based Practice in Science Education (EPSE) Research Network.