

**'value added' measurement
of school effectiveness:
a critical review**

Lesley Saunders

nfer

‘value added’ measurement of school effectiveness: a critical review

Lesley Saunders

nfer



INVESTOR IN PEOPLE

Published in January 1999
by the National Foundation for Educational Research,
The Mere, Upton Park, Slough, Berkshire SL1 2DQ

© National Foundation for Educational Research
Registered Charity No. 313392
ISBN 0 7005 1530 5

CONTENTS

| | page |
|--|-------------|
| Acknowledgements | i |
| List of Acronyms | ii |
| 1. INTRODUCTION: THE CHANGING CONTEXT FOR EDUCATIONAL 'VALUE ADDED' | 1 |
| 1.1 The Changing Context for Value Added | 1 |
| 1.2 Aims of the Review | 2 |
| 1.3 A Note on the Bibliographical Search | 2 |
| 1.4 The NFER's Work in Value Added | 4 |
| 2. A BRIEF HISTORY OF VALUE ADDED IN EDUCATION | 6 |
| 2.1 Introduction: An Idea Whose Time Has Come? | 6 |
| 2.2 Origins and Meanings | 7 |
| 2.3 Value Added in Education | 10 |
| 2.4 The Value in Value Added? | 19 |
| 3. ANGELS ON A PINHEAD? REVIEW OF THE ACADEMIC LITERATURE | 21 |
| 3.1 Introduction | 21 |
| 3.2 The Development of Value Added Principles in Educational Research in the Compulsory Sector | 22 |
| 3.3 Charting the Relationship Between 'Value Added' and School Effectiveness | 27 |
| 3.4 Summary of Analytical Models and Methodological Strategies | 42 |
| 3.5 What Can Value Added Reveal about 'Effectiveness' and Educational Quality? | 46 |
| 3.6 What Further Questions and Queries does Value Added Raise? | 48 |
| 4. A SHARPER STICK TO BEAT THEM WITH? REVIEW OF THE POLICY LITERATURE | 52 |
| 4.1 Changes in Official Policy on Value Added | 52 |
| 4.2 The Value Added National Project | 53 |
| 4.3 Reconciling Academic Criteria with Policy Requirements | 59 |
| 5. GETTING BEYOND THE NUMBERS GAME: ETHICAL AND PRACTICAL CONSIDERATIONS IN USING VALUE ADDED | 64 |
| 5.1 Introduction | 64 |
| 5.2 Issues in Schools' Use of Value Added | 64 |
| 5.3 The Need for Further Research | 67 |
| 5.4 Fruitful Qualitative Research Issues | 68 |
| 5.5 Using Value Added for School Improvement? | 71 |

| | | |
|-----------|---|-----------|
| 6. | CONCLUSIONS: THE FUTURE FOR VALUE ADDED IN EDUCATION | 73 |
| 6.1 | Identifying and Addressing Under-Achievement | 74 |
| 6.2 | Institutional Development | 75 |
| 6.3 | Partnership between Schools and Local Education Authorities | 77 |
| 6.4 | Final Thoughts | 78 |

APPENDICES:

| | | |
|-----------|---|-----------|
| A. | BIBLIOGRAPHY OF THE NFER'S WORK IN VALUE ADDED | i |
| B. | REFERENCES | iv |

ACKNOWLEDGEMENTS

This review would not have been possible without the help of Pauline Benefield, NFER Research Librarian, who always seemed to know what was wanted before it was asked for. Christine Taylor, also a member of the Library staff, gave generously of her time to correct and update the references. Julia Rose prepared the text for publication. The author is indebted to many colleagues in several institutions for their ideas and support; but this text has particularly benefited from the constructive criticism of NFER colleagues, especially Dougal Hutchison, Chief Statistician and Marian Sainsbury, Principal Research Officer, Assessment and Measurement. Thanks are also due to Geoffrey Walford (Reader in Education Policy, Department of Educational Studies, Oxford University) and Malcolm Rigg (Director of Social Research at the British Market Research Bureau) for their insightful comments on early drafts. Any errors of fact or interpretation which remain are the responsibility of the author.

LIST OF ACRONYMS

| | |
|--------|--|
| ALIS | A-Level Information System |
| CEM | Curriculum, Evaluation and Management Centre |
| CNAA | Council for National Academic Awards |
| DES | Department of Education and Science |
| DFE | Department for Education |
| DfEE | Department for Education and Employment |
| ERA | Education Reform Act (of 1988) |
| ESL | English as a second language |
| F/HE | Further/Higher Education |
| FSM | Free school meals |
| GCSE | General Certificate of (Secondary) Education |
| GEST | Grants for Education Support and Training |
| (G)NVQ | (General) National Vocational Qualification |
| KS | Key Stage (of National Curriculum) |
| LEA | Local education authority |
| NC | National Curriculum |
| NCER | National Consortium on Examination Results |
| NFER | National Foundation for Educational Research |
| OFSTED | Office for Standards in Education |
| PCFC | Polytechnics and Colleges Funding Council |
| PI | Performance Indicator |
| QCA | Qualifications and Curriculum Authority |
| QUASE | Quantitative Analysis for Self-Evaluation |
| SAT | Standard Assessment Task |
| SDP | School Development Plan |
| SE(S) | Socio-economic (status) |
| SEAC | School Examination and Assessment Council |
| SEN | Special Educational Need |
| SEO | Society of Education Officers |
| SEU | Standards and Effectiveness Unit |
| SMT | Senior management team |
| SPSS | Statistical Package for the Social Sciences |
| TGAT | Task Group on Assessment and Testing |
| UGC | University Grants Committee |
| WO | Welsh Office |

1. INTRODUCTION: THE CHANGING CONTEXT FOR EDUCATIONAL 'VALUE ADDED'

1.1 The Changing Context for Value Added

A critical review of value added in education is particularly relevant in the late 1990s. The introduction of the national value added system from autumn 1998 can be seen as the culmination of a decade of sustained and public argument about how to measure the performance of pupils in the nation's schools in a way which sheds light on progress as well as standards. It might even seem that common sense – albeit supported by complicated statistical techniques and qualified by methodological reservations – has won a brilliant victory. At the very least, the academic debate on school effectiveness and how to measure it is now integrally linked with the national political agenda for educational quality. This means that the range of parties interested in value added extends from politicians to school senior managers, and from academic researchers to lay governors. There is consequently a need for continuing discussion not only of the main technical aspects of how to assess value added, but also of how the different perspectives of the various 'stakeholders' have resulted in different expectations and requirements of value added.

It is interesting to look back and see how dramatically the context and rationale for calculating value added measurements of performance have changed since the 1980s, when 'value added' was regarded as a quasi-technical idea which had strayed into education from economics and whose practical application to the debate on standards was not immediately obvious. And it took much longer for national government to be convinced: even when the principle of value added had been much more clearly understood and its practical possibilities advanced to the point of several systems and services being widely available, the Secretary of State for Education in 1993-94 was adamant that results which were anything other than 'raw' must be 'cooked'. He refused to give official support to what he apparently construed as another trick in the educational mafia's book. To both scholars and educational managers value added continued to hold out important possibilities, however, and methodological expertise developed alongside the growing political recognition that better ways were needed of assessing schools' effectiveness than performance tables based on figures which took no account of pupils' and schools' different starting points.

Perhaps not surprisingly, a consensus about the best statistical procedures and analytical models for measuring value added has been hard to reach. Much of the

debate has been concerned with how to strike the right balance between what is ideal, analytically speaking, and what is feasible. It has focused on such basic issues as the pros and cons of using National Curriculum assessments (in the form of 'levels', i.e. integers, or of finer versions of levels) instead of or as well as nationally standardised cognitive ability or skills tests as outcome and/or input measures, and whether or not to adjust outcome indicators by taking into account the socio-economic characteristics of the pupils and/or school-level contextual information. There has also been a lively argument about whether or not multi-level modelling is the preferred statistical technique given its lack of transparency to the layperson on the one hand and its unrivalled accuracy on the other.

Even so, by the late 1990s the term 'value added' is part of common parlance, used by journalists, Government politicians, local policy-makers, school managers and teachers, as well as academics, to mean something like 'a fairer or more accurate measurement of pupils' performance and therefore of the quality of their education'. Something like, indeed: therein lies a major difficulty. It is part of the purpose of this review to tease out the diverse and sometimes conflicting or unclear meanings contained within the range of usages of the term, and to suggest that common sense may be misleading.

1.2 Aims of the Review

The main aim of the review is to contribute to the policy debate on the *interpretation and best use* of value added measures of school effectiveness. The review accordingly discusses the following questions:

- ♦ what is meant by value added in an educational context;
- ♦ the role of school effectiveness research in the conceptual and methodological development of value added;
- ♦ the impact of the changes in the policy context for value added;
- ♦ the implications of value added for the school improvement agenda;
- ♦ what remains to be further developed or resolved in the refinement of value added.

1.3 A Note on the Bibliographical Search

There are several pre-existing and very extensive research fields whose connection with value added – even when this is only implicit or must be inferred – is crucial for a full understanding of the term's educational import. These include the assessment and measurement of performance, the development of performance indicators, the

definition and measurement of school effectiveness, the identification of factors affecting performance, the economics of education and ultimately the issue of using performance data for improving pupils' attainment. There is also a large and diverse body of literature, ranging from seminal academic reports and articles to ephemeral press features, explicitly concerned with defining, measuring and/or problematising 'value added'. It is therefore only sensible to impose some boundaries on the present study. We have adopted the strategy of concentrating attention on the following kinds of item:

- ♦ for tracing the historical lineage, articles etc. published before the mid-1990s (when value added became part of mainstream debate) and limited to items in bibliographical searches whose titles or key words contain the term 'value added';
- ♦ for understanding the mutual influence of school effectiveness and value added research, articles etc. which, whilst they may not allude to 'value added' as such are clearly methodologically and/or conceptually part of the clarification of what value added has come to mean (this is mainly an exercise in hindsight);
- ♦ for examining the later development of value added, articles etc. which provide one or more of the following:
 - first-time published use, or clarification, of key concepts and assumptions;
 - important differences from the consensus or conventional view;
 - description of techniques for measuring value added;
 - critical overview of value added approaches in the context of the broader research issues involved;
- ♦ for understanding the changing policy context, reports and press releases emanating from national, especially government, agencies.

A full list of references is given in Appendix B. Because this is first and foremost a review of the meanings and uses of value added, the more technical aspects of statistical modelling and analysis are merely touched on. Sufficient references are given, however, to enable the interested reader to follow up key sources.

Where to stop? The literature is being added to almost weekly, and the implementation of the full national value added system over the next few years will no doubt create the need and opportunity for even more discussion. We have chosen to construe the introduction of that system, however, as imposing a logical kind of stopping place for a review like this.

1.4 The NFER's Work in Value Added

The NFER has been a major contributor to the development of value added research, primarily in the fields of empirical investigations of independent variables associated with performance and of the practical application of value added analyses to schools' and LEAs' development planning. The range of NFER's work is outlined below.

There are currently two NFER services available which provide high quality quantitative data to inform school/LEA improvement:

- ♦ ***QUASE (Quantitative Analysis for Self-Evaluation)***

This is a subscription-based service direct to secondary schools in all sectors or entire LEAs. Analyses of overall performance at GCSE are provided, together with fifteen GCSE subjects, attendance, and differential effectiveness for girls/boys and lower/higher attainers. Further options – such as pupil and parent attitudinal surveys, summary reports for teachers/governors and brokerage with other QUASE schools – are available.

- ♦ ***Examination Results in Context***

LEAs need sensitive and accurate measures of school performance, both for purposes of accountability and for targeting limited resources more effectively. Several metropolitan LEAs have been participating in a programme run by NFER for a consortium under the aegis of the Local Government Association. The NFER provides value added analyses, consisting of LEA- and school-level tables, of pupils' total and average GCSE scores as well as results in mathematics, English and science.

Each LEA or school receives a confidential package containing a detailed analysis of its performance in the specified outcome areas, given in both graphical and tabular forms and set against appropriate national norms. Interpretative follow-up seminars are also offered.

In order to continue to inform these services as well as to contribute to the academic debate on value added, the NFER is undertaking further research and development projects in the area.

- ♦ ***Schools' Use of Value Added Data: an Exploratory Study***

This study is examining the issue of support and training for schools in managing and using value added data. Better guidance and protocols are needed, based on empirical evidence of what is happening in schools. The aims of the project are to investigate how far and under what circumstances value added analyses have a role to play in school improvement and the raising of pupils' attainment in different institutional contexts; and to draft some guidelines for good practice.

♦ ***Secondary Research and Analysis of Value Added Data***

The main focus of this project is to carry out secondary analysis and research on the QUASE and Examination Results in Context data in order to address a number of important issues, including the following:

- How are GCSE results related to background factors, and have these relationships changed over time?
- What can be said about overall trends in GCSE performance over time, taking account of background factors?
- To what extent are school-level effects stable over time, or over different outcome measures?
- Are there apparently different relationships for schools in inner-city LEAs and those distributed across the nation?

Other work is being carried out at NFER which has a 'value added' dimension, including:

♦ ***Testing and Evaluation for the National Literacy and Numeracy Projects***

♦ ***Baseline Assessment***

- ♦ Project to collect pupil-level test results on the optional national QCA tests in Years 3, 4 and 5, in order to monitor progress from the end of KS1 through KS2.
- ♦ Project to collect individual pupil-level KS1 data from 1994 to 1997 and match to subsequent KS2 data.

Each of these has its own Research Information Sheet available from NFER.

A bibliography of articles and reports by NFER research staff is given in Appendix A. Parts of this review have been written up as academic and press articles: see, for example, Saunders (1997), Saunders (forthcoming a-e)), Saunders and Thomas (1998).

A summary report is available from the NFER Dissemination Unit (SAUNDERS, L. (1998). *'Value added' measurement of school effectiveness: an overview*. Slough: NFER.

2. A BRIEF HISTORY OF VALUE ADDED IN EDUCATION

2.1 Introduction: An Idea Whose Time Has Come?

The review starts, in this section, by giving an account of how and – as importantly – why the term 'value added' has come to be used in an educational context. It goes on in Sections 3 and 4 below to outline the diverse ways in which the term value added is understood and applied by both academics and policy-makers; and to discuss the variety of methods used to arrive at measurements and/or analyses of value added. In doing so, some evaluative judgements are made of what should and should not count as workable definitions of value added, and consequently of how best to arrive at robust and rigorous analyses which purport to measure value added.

It is an important part of this account to demonstrate what are called 'functional ambiguities' in the term 'value added'; that is to say, ambiguities which cannot be altogether eliminated because they are necessary to how the term is made to function. It argues that these ambiguities were in turn inherited from the economic context where the term originated.

In the educational context, 'value added' is one of those terms which comes with an agenda already attached, the agenda in this case being the political preoccupation in the UK (as in other western developed nations) with standards and quality in education, or rather the lack of quality and/or of deterioration in standards therein and the justifiable wish of politicians to get better value for public expenditure. For some time, this agenda has been explicitly attached to the issue of global competitiveness, and the consensus belief that education and training are important levers for economic competitive advantage. At the same time, many practitioners have been concerned to demonstrate that standards (particularly in their own institutions) have not been slipping, that the profession is giving value for money. To put it another way, from the vantage point of the late 1990s value added seems in many ways to have been an idea waiting for its time.

This preoccupation with standards and quality both fed and was sustained by what has been called 'an industry' (Woodhead, 1997, p. 3) in academic research into school effectiveness. From the late 1980s onwards the debate about measures of performance which reveal the 'increment on performance of each individual child that goes to [the school]' – in Smith and Tomlinson's (1989, p. 301) phrase – became vociferous and many-faceted. It led on the one hand to highly technical expositions of particular

methodological approaches and in this the achievements of the research community were extensive and important. The decade of the 1990s saw an impressive development in identifying, and solving, problems of statistical methodology so as to enable both individual pupil and school academic achievements to be more rigorously analysed and more acutely interpreted. This work is discussed, in broad terms, in Section 3 below.

On the other hand, the complexity of the issues involved in assessing schools' effectiveness, both relatively and absolutely, became inescapable. It is worth remembering that when value-added measurements began to be talked about outside the restricted circle of academic researchers and statisticians, the then Secretary of State for Education was scathing about what he was convinced were ever more sophisticated ways of 'cooking' schools' results: he wanted to stick with 'raw' ones that were simple to compile and understand. By 1995, however, the Secretary of State for Education and Employment (the Departments had been merged in the meantime, thereby confirming the assumed link between education and the economy) who had replaced him had been won round to the extent that a research contract was awarded to the Curriculum Evaluation and Management Centre (CEM, then based at the University of Newcastle) by the Schools Curriculum and Assessment Authority (SCAA) to design and pilot national systems for value added measures (Fitz-Gibbon, 1997). The policy concerns surrounding value added are addressed in Section 4 below.

2.2 Origins and Meanings

Before exploring the measurement and range of possible applications of value added, however, it is necessary to spend a little time unravelling some of its meanings since, as Spours and Hodgson (1996, p. 7) point out, 'value-added is used as a short-hand or as an umbrella term to describe a diverse field of practices and projects'. In acknowledging that the term originates in economics, they attempt an economic definition (as many educational commentators do not), as follows:

The term 'value-added' originates in economics when referring to the difference between inputs, such as raw materials, and energy and final outputs represented by the value of sales... In recent years, the economic concept of 'value-added' has taken on an additional meaning when associated with the need for advanced Western economies to focus on high quality manufacturing as a way of competing successfully with rapidly growing economies... The term 'value added' in this context is applied to goods whose quality and value is increased

by high levels of technology and skill in the manufacturing process
(p. 5).¹

Implicit in this definition is clearly some principle of transformational activity on the part of human agency, and this should be a signal that there is perhaps more to the economic meanings of the term than meets the eye.

Clearly it is not feasible to undertake a comprehensive review of economic theory for this review; but some minor annotation of what seem to be the difficulties in arriving at an economic definition of value added is relevant. One might wish to explore further, for example, the apparently naive question, 'value added to what, by whom and for whose benefit?' This modest task is not without problems of understanding and interpretation, however, because a seemingly inherent characteristic of the term value added is its ambiguity. Is it, for instance, an absolute or a relative concept? And with descriptive or evaluative force? Thus, does 'adding value' simply describe what all companies do in transforming inputs into outputs; or does it refer to what efficient companies do in transforming inputs into outputs with a lower cost-benefit ratio than their competitors?

The answer seems to be, it depends. Mainly it depends, on the one hand, on a prior definition of 'value', which itself depends on the purposes for which it is being defined and, usually in an economic context, (ac)counted. The other critical factor is the level of abstraction within the economic system at which the definition is made. So, whilst it may be well-known that the term 'value-added' has been borrowed from economics, what is not so well-known (at least amongst educationists) is that its economic definitions and usages are as supple and multiple as the educational ones.

To take an everyday usage, 'value added tax (VAT)' is the tax paid on the difference between the buying price and the selling price². This meaning of value added is

¹ Intriguingly, this discussion by Spours and Hodgson leads one to speculate whether there has not also been some feedback in the opposite direction, that is into the economic context from the educational one, since the discourse within which what one might call this evolved definition of value added is situated again concerns the relationship between the (national) economy and education/training provision; and to what extent education and training can be demonstrated, or is only presumed, to be integral to national economic health. Whilst this is not the place to discuss that discourse and its own long history, it is worth saying that some doubt has recently been cast on whether the quality of education/training should be taken as the key variable in the success of an economy (see, e.g., Hutton, 1996, quoting Bluestone, 1994, Freeman and Katz, 1993, and Nickell and Bell, 1995).

² It should be noted, of course, that in the case of VAT, the 'value' or revenue accrues to central government; more usually, it is shareholders in private companies who are identified as the key beneficiaries from value added.

consonant with the definition given by Spours and Hodgson (quoted above), which enables them to go on to say that:

Using the same input/output notion as in economics, value-added has been used to describe the difference between the state of knowledge or qualifications of a pupil on course entry and her/his state on exit (p. 5).

Common-parlance usages of the term are actually premised, however, on complex theoretical investigations of economic value, which in turn have been shaped by substantive ideological assumptions. The key question of agency – of who adds value to what and for whose benefit – which is often obscured in classical economic theory, was comprehensively treated, of course, by Marx in his published work from the Paris Manuscripts of 1832 to *Das Kapital* (first published in 1867). In Marxist theory, value added is construed as part of the expropriation of labour-power by capitalists, as in this extract from Mandel (1962 p. 305):

Labour-power, as we know, has the dual function of conserving the value of constant capital (the stock of machines, raw materials, buildings) by transferring part of this value to currently produced commodities, and of producing all the new value available to society. The first-mentioned property of labour-power makes it possible to conserve the accumulated stock of social wealth and instruments of labour, which determines the average level of the productivity of labour and the material civilisation of the given society. The second makes it possible to create an income – a 'value added' – which in capitalist society is divided between income of labour (wages) and income of capital (surplus-value).

The object of quoting this excerpt is not to enter into a debate on Marxist versus classical economic theory, but to underline the point that, when educationists say that value added is a term borrowed from economics, this does not serve well as a derivation of its meaning. The conclusion to be drawn is that there is no definitive definition, no *ur*-meaning to value added: it is in essence a conflicted locution. Paradoxically, perhaps, this will be of help as the educational applications are explored. The issue of advantage or benefit is a particularly important one to pick up on: value added, even in an economic context, has an evaluative significance even if it is not actively present in all uses of the term. In an educational context, we would argue, the evaluative *potential* located within the term has been converted into an evaluative *purpose* in all its applications.

It is not easy to find the definitively first use of the term value added in an educational context in the UK. An article written by two researchers based in the USA was published in the *Journal of Educational Statistics* in 1976 (Bryk and Weisberg, 1976); it used the term (which appears in the title) to mean 'the growth increment [provided by the educational intervention or program] over and above natural maturation' (p. 150). By 1990 the term was in sufficiently common use in the UK to appear, without warranting elucidation, in the title of a key document published by the Council for National Academic Awards, *The Measurement of Value Added in Higher Education* (McGeevor *et al.*, 1990). An article discussing the work behind this document (Gallagher, 1991, p. 20) then helped to map the value added territory more explicitly, though conceptually and methodologically the suggested approach was inchoate. Gallagher herself points out that 'the term value added appeared in post-school education at least as long ago as 1978 [in Pratt *et al.*, 1978]'.

Certainly the early 1990s were, to judge from anecdotal reports and first-hand experience, a time when many people in educational policy circles were talking with varying degrees of confidence about the need to understand and analyse the 'value added' by institutions to their pupils' academic performance. By 1992, the National Commission on Education considered the term well enough understood even by people outside the education profession to entitle one of its widely disseminated Briefing Papers *Measuring Added Value in Schools* (McPherson, 1992), which is one of the clearest and most accessible accounts available to a lay audience of the key technical and ethical issues. (Incidentally, one indication of the rapid proliferation of interest in value added is the fact that there were twice as many bibliographical references on the topic of value-added in 1995 as in 1990.)

2.3 Value Added in Education

It will now be helpful to consider the provenance of the term value added in education and the route by which it came to be used as a term connected with the measurement of quality. It should be noted, therefore, that the remainder of Section 2 is concerned with tracing *historical* connections: the review of the literature is brought more or less up to date in Sections 3 and 4 below. The discussion here explores some of the history of the term value added as applied to the economics of education and traces two distinct strands in its development. In doing so, it permits some criteria to emerge for making evaluative judgements of what should and should not count as meaningful definitions of value added.

The previous sub-section noted specific dates by which the term value added was evidently in common usage. But this is not to imply that the ideas which lay behind the introduction of the notion of value added to education were not already very well established by then. In fact, the development of value added analysis has two rather different albeit intertwined strands, which – in addition to the existing difficulties in the economic derivation outlined in the previous section – may well account for some of the continuing confusion in its definition and application.

As an explicit *term*, its first appearance seems to have been in the further/higher education (F/HE) sector in the context of the development of institutional performance indicators. As a critical *issue*, however, it originated from the school effectiveness research agenda which was particularly at home in compulsory secondary education. In both cases, there has been an associated methodological inquiry which has become a dimension of educational research in its own right.

What incontrovertibly gave both sectors a rationale – some would say an outright requirement – for being concerned with value added in education were the centrally imposed funding-and-accountability mechanisms introduced into the education system in the late 1980s and early 1990s. It is interesting to note that a key document in the justification of how these mechanisms should and could work was compiled by a firm of accountants (Coopers & Lybrand, 1988), an event which was surely a message about the way things were to be from now on. The ensuing legislation included the Education Reform Act of 1988 affecting primary and secondary schools in England and Wales, and the various statutory changes affecting the funding, control and status of institutions of further and higher education (F/HE) in those countries.

A search of the literature of that time makes it possible to discern how exponents in the different sectors developed rather different kinds of argument about value added measurement. Essentially, the F/HE sector was particularly concerned with assessing efficiency and effectiveness *within* educational institutions, whilst the school sector was more pre-occupied with ways of making fairer comparisons *between* educational institutions. And again this distinction of emphasis should be no surprise, given the wider educational policy context. At the same time, one should not make too much of the difference, since what the two conceptualisations have in common is the underpinning question: 'How can student/pupil progress be measured in such a way as to throw light on the performance of institutions?' This question is probably the key to understanding the methodological principles of value added, and will be returned to in Section 3.

However, it suits the purpose of this section to pursue the historical perspective a little further and temporarily to treat the two strands as separate developments. So far as the F/HE sector is concerned, the response of the Council for National Academic Awards (CNAA) and the Polytechnics and Colleges Funding Council (PCFC) (McGeevor *et al.*, 1990) to the 1988 Government directive on efficiency in higher education, gives some strong clues. This document defined value added as:

a measure of pupil achievement which takes into account the effects of differential pupil inputs (i.e. entry qualifications). Pupil achievements thus conceived can be used to contribute to the evaluation of institutional quality in the teaching and learning functions (p. 1).

This essentially economic notion of inputs (and outputs) can be traced back to the argument developed over a decade earlier in Pratt *et al.* (1978). The authors of this thought-provoking book undertook some ground-clearing discussion in Chapter 3, 'Economic issues: the economics of education', when they asked: 'Input and output – can there be efficiency in education?' Their view was that:

The concepts of input and output are central to economics. They underpin notions of efficiency and productivity and calculations of *value added* and of costs and benefits. Yet they seem to have caused economists little but trouble in their application to education. There are indeed some economists who, when faced with an education institution or service, cannot tell the difference between inputs and outputs (p. 154). [Our emphasis.]

In a critique of Government cost studies, the authors expand on the question of what should count as inputs and outputs in an educational context. In budgetary terms, of course, a major input measure must be that of financial cost; in a complex service such as education, the main issue might well be seen – and was so seen by the then Department of Education and Science (DES) – as one of how to define costs. But the authors argue that:

It is not a matter of a satisfactory definition of costs. It is a matter of deciding what costs to calculate, and the decision will be different for different purposes. *The question of what to count as input is not a general or definitional question*: it is a question that has meaning only when related to both the output and the process involved. It will differ, for the same institution or service, in different circumstances (p. 156). [Our emphasis.]

The sentence which has been highlighted is arguably a crucial piece of conceptual clarification which serves to underline a similar point made in the previous sub-section. It is a principle to which it will be necessary to return.

The question of how educational 'outputs' were to be defined was apparently even more tricky. It is salutary to be reminded by Pratt and his colleagues that in 1978 it apparently could not be assumed that the DES, still less the University Grants Committee (UGC), would think of outputs in terms of pupil performance or, more globally, educational standards. The authors characterise official thinking on outputs as 'unsatisfactory' and 'muddle[d]', in that factors such as 'pupil/teacher ratios' and 'quality of teaching' as well as 'graduates' themselves, are instanced as examples of output or product. They go on to argue firmly for a 'simple but serious' application of economic concepts to education; and therefore for a view of inputs which would comprise all resources, including staff, and of outputs which would comprise 'educated pupils'. But, they say:

This may seem a simple and obvious conclusion; in fact it overturns all current practice in costing, budgeting and accounting; both administrative and academic... In education we ... need to be able to discuss what mix of inputs is most apt for producing desired outputs...
(p. 159)

These extracts should serve to indicate how the discussion of efficiency and quality in F/HE was closely tied into a concern with funding, and therefore how 'educational value added' started out as a truly economic pre-occupation, at least in this sector. The chapter in question goes on to propose an explicit definition of value added:

The question of quality will then relate to the difference which the educational institution or service has made to the entrants... The course's contribution, the value added by it, is precisely the difference it makes to the pupil (pp. 160-3).

What follows in the remainder of that chapter is a discussion of the problems such a definition raises, such as the inappropriateness both of economic measures like graduate salaries and of existing examination and assessment systems to measure the difference courses make, combined with the tendency of educators to make unsupportable claims about the contribution of their courses to pupils' maturity and so forth. The authors conclude:

We have discussed various economic ways of measuring the [difference which the college makes to the uneducated pupil] and have concluded that these are not apt... It is for the educators to create measures which, apt in principle, are also convincing in practice (p. 166).

It might be argued that this is the challenge which the value added 'project' – in so far as the various individual academic studies, research and development reports and government policy statements add up to such a thing – has in principle accepted; except for the fact that the arguments and conclusions presented in this book seem to have sunk almost without trace. What has happened in practice is that educators have defaulted to using proxy variables which already exist and are measurable, for understandably pragmatic reasons – but this is to pre-empt the discussion in the following section of this review.

So far as it is possible to check, the work was last cited, in the context of discussing educational value added, in Gallagher (1991, p. 20), although 'the advice and assistance' of one of the authors was acknowledged in the CNAA document (McGeevor *et al.*, 1990) cited above. It is not proposed to undertake a critique here of the computational bases for value added presented in the latter document, important though they undoubtedly have been for the work of university and college managers. What is more at issue for this review is the way in which Pratt *et al.*'s clear though challenging definition of value added was rendered imprecise by a series of assumptions and elisions made in subsequent documents such as this, which thus allowed a basic confusion to be perpetrated and, as it has turned out, perpetuated.

First of all, as the excerpt quoted earlier in this section reveals, the authors of the CNAA/PCFC document seem to have pre-empted the discussion of what should constitute 'input' and 'output', since they implicitly equate 'output' with 'a measure of pupil achievement' and explicitly define 'input' as, and thereby restrict its meaning to, 'entry qualifications'. The imputed relationship of input to output is rather vaguely glossed as a matter of 'taking into account'. What the measurement of value added therefore comes down to is something 'based on a comparison of entry and exit qualifications', which is not at all the same thing as a computation of how efficiently inputs in the normally understood meaning of the word were deployed to achieve desired outputs.

Further on in the same document, the following note is appended to a table purporting to show 'value added scores for male/female sociology graduates':

all [methods of calculating gender differences in value added] indicate that women *obtain greater value added* than men (p. 23) [Our emphasis.]

A little syntactical deconstruction is called for. It will be remarked that this phrasing contains an ambiguity about the agency to whom the capacity to add value is being attributed, in that 'obtain' could have an active or a passive meaning (being equivalent either to 'women achieve greater value added than men' or to 'the results for women show a higher level of value added than do those for men' respectively). Now this vagueness of verbal expression might be regarded as a trivial matter – pointing it out at all might be dismissed as sophistic nit-picking; but its implications are decidedly non-trivial. For the ambiguity amounts to a conceptual equivocation (not necessarily intended by the authors, it must be said) about what sort of activity, in what category, 'adding value' is. To make this as plain as possible, we are arguing that the ambiguity, when elided with the notion that measuring value added is something to do with 'a comparison of entry and exit qualifications', results in a blurring of boundaries between two logically and politically disparate concepts: namely, on the one hand, the progress made by an individual between, say, A-levels and degree and, on the other, the amount – if any – of difference to that individual's progress which can be attributed to what has been done by the institution s/he attended.

As a rider to this discussion, it may be noted that the objectives of the CNA/PCFC project were to test different approaches to the measurement of value added:

'... *relevant to the operational needs* of the PCFC and higher education institutions' (p. 1) [Our emphasis.]

It may be speculated that those needs had more leverage on the project than a strict attention to what could be construed as a question of semantics merely.

The distinction made above, between individuals and institutions, is material to the application of, and the methodology associated with assessing, value added in education. It is an argument of the present review that the former concept, whilst integral to both the discussion and the measurement of value added, is not value added in itself. Adding value, certainly according to Pratt *et al.*, is what establishments (colleges, they would say, as well as companies) do, by virtue of being economic systems with resources at their disposal and decision-making structures for disposing in a variety of alternative ways of those resources; adding value is, by definition, not

something which is applicable to what individuals can do for themselves by way of making progress.

What is striking is that this conceptual ambiguity has continued right into the most recent literature, where it is the case not only that different authors mean different things by the same term but also that the same authors use the term to mean these two rather different things (see, for example, Spours and Hodgson, 1996; Fitz-Gibbon, 1997). As was said earlier, it is quite hard sometimes to spot such instances or to get to grips with the issue at all, however, precisely because it is concerned at least as much with things which are *not* said or clarified as with things which are.

Turning now to the Gallagher article (1991), which, as was earlier noted, is a useful critical commentary on the CNA/PCFC document and the work which formed its basis, we can see that some of these problematic issues were adumbrated therein. Gallagher defines value added as [a performance indicator which] 'is intended to indicate the contribution of an institution to a pupil's achievement, to determine what an institution adds to a pupil as she/he passes through it.' She glosses this as follows: 'although addressing largely qualitative issues, if value added is to be adopted as a performance indicator it requires measurement in quantitative terms.' (p. 19) Each of these points, one could say, constitutes a key principle to bear in mind when considering value added, namely: the need for *standardisation* of measurements; the notion that institutions '*make a difference*' to pupil performance which can be assessed; and the acknowledgement that what is being looked at is a matter of qualitative judgement – or, as we might prefer to spell out, involves *making value judgements* (whether or not these have been made explicit and subject to scrutiny).

But a major finding to be highlighted at this point is that educational value added in F/HE is usually coterminous with 'progress made' or 'distance travelled' (two commonly-occurring phrases in the literature); and whilst the idea of 'distance travelled' clearly entails a starting and an end point, it does not require that there must be a known correlation between the two; still less that beginnings and ends must be defined in precisely the same way in different institutions. Spours and Hodgson's (1996) definition – repeated here for ease of reference – succinctly sums up this view:

Using the same input/output notion as in economics, value-added has been used to describe the difference between the state of knowledge or qualifications of a pupil on course entry and her/his state on exit. By taking the starting point of the learner into account, it is possible to consider the extent to which both a course and an institution have been responsible for any progress made (p. 5).

Such a definition, whilst not incorrect, conceals several conceptual and technical difficulties. Gallagher (1991) was moving towards a more rigorous view of measurement when she said that 'some [additional] notion of relativity is required if judgements are to be made about the institution and comparisons made between institutions.' In other words, what Gallagher calls *comparative* value added needs to be adduced, which will create 'a level playing field' and one which is based, not on an arbitrary score, but on 'an empirically derived expected value'. (Gallagher, 1991, p. 22, quoting McGeevor *et al.*, 1990). As Gallagher points out: 'using this approach, it is no longer up to institutions to decide the value of entry and exit qualifications.'

This is the point at which the discussion of value added in F/HE begins to bear strong similarities to the applications developed in the compulsory sector. Since this review is ultimately more concerned with schools than with colleges and universities, the F/HE narrative will finish here without trying to do justice to the subsequent attempts to explore value added in the post-compulsory sector: a useful reference to follow up is Barnard and Dixon (1998), which contains a relevant bibliography.

Let us now turn to the second strand in the development of educational value added: how value added has come to be understood in secondary as distinct from tertiary education and in particular what it owes to the school effectiveness debate.

If, without further argument for the moment, one accepts that there is a different meaning of value added in the compulsory sector and therefore takes the value added question to be pretty much as Sweetman (1997, p. 77) states it: 'at its simplest, how much better do pupils in a good school perform than they might have been expected to do from their performances on their first day?', then it is possible to track this back through the accumulating research into school effectiveness, which already by the early 1990s had a long and strong pedigree. This research is summarised and discussed, so far as it is relevant to the present review, in Section 3 below, but it is perhaps worth risking some over-simplification here to say that at the core of the research was the demonstration that some schools are measurably more effective than others (for example, Rutter *et al.*, 1979; Smith and Tomlinson, 1989). Rutter *et al.*'s research arguably contributed to the shift of focus of concern away from national monitoring of aggregate statistics towards a scrutiny of the performance of individual schools. The conclusions reached by the end of the 1980s in this area could be represented thus:

Research and analysis showed that schooling is not effective as a means of reducing individual inequality. It did not show that schooling has little effect on whether or not children can read, write and do

arithmetic... The result of going to an effective school can be seen as an increment on the performance of each individual child that goes to it. This increment may be large enough to be very important for its effect on what each individual is actually capable of, yet small in comparison with the differences between individuals.’ (Smith and Tomlinson, 1989 p. 301)

It was inevitable that this new emphasis on performance (as distinct, say, from other legitimate concerns like curricular process or learning experience) should have concentrated attention on how best to *measure* performance or ‘achievement outcomes’ in Mortimore and Stone’s phrase (1990, p. 73), in ways which would show how far schools had actually contributed to the achievements of their pupils. (Hardly surprisingly, then, the introduction of value-added analyses has in turn – as is argued later in the review – had an impact on the direction and scope of school effectiveness research.)

The evidence from, and methods of, this body of research were able to be pressed into the service of attempting to make fairer comparisons between institutions, which has continued to be a huge pre-occupation amongst head teachers and others concerned with the effect on schools of ‘league tables’, the public reporting of schools’ academic performance according to results obtained in national curriculum tests and public examinations, introduced by Government in 1992. (Schools were already enjoined by the 1980 Education Act to publish public examinations results in their school prospectuses.) Secondary schools and their managers were most visibly and immediately affected by league tables, because of the presumed role of such tables in informing parental choice of school (extending parental choice had been another avowed intention of the Education Reform Act), and also because GCSE examination results for pupils at the end of Key Stage 4 of the national curriculum – unlike national curriculum tests for pupils at the end of Key Stages 1, 2 or 3 at that point in time – were already a standard means of assessment and accreditation which could readily be deployed for drawing up school-based lists in the public domain. One could go so far as to say that making comparisons between schools has been the major purpose of measuring, and *a fortiori* of defining, value added in the compulsory sector.

Technical and conceptual complexities of the kind alluded to earlier in this section were therefore squarely in the frame. We may note, for example, the publication of ‘The search for a fairer way of comparing schools’ examination results’ (Gray *et al.*, 1986) more than a decade ago. In posing the question: ‘under what circumstances might one be justified in claiming that [school A] has done ‘better’ than [school B]?’ (p. 91), the authors could be said to have encapsulated the aim of (comparative) value

added analysis *avant la lettre*. The findings from this study – for example, that ‘differences between schools did not appear to be a very important source of differences in the results achieved by individual pupils of differing ability/attainment levels attending different schools’ (p. 118); and that there was ‘competing evidence on the question of whether individual schools were differentially effective with pupils whose abilities/attainments differed at intake’ (p. 118) – also pre-figured a very great deal of the evidence on relative measurable effectiveness which has since been accumulated.

As the authors argue, methodologically speaking two things follow from the intention to compare schools’ performance: first, some concept of ‘contextualisation’ must be introduced, to allow for the fact that schools are dealing with pupil populations which differ considerably in terms of prior achievement and other key variables (the issue of ‘input’). Secondly, since schools are ‘part of an educational system which has natural hierarchies or levels of ‘nesting’ [e.g. pupils within classes within year groups within schools]’ (p. 117), the statistical model adopted to analyse performance data should reflect this structure. These arguments are taken up in Section 3.

2.4 The Value in Value Added?

The evidence presented so far indicates that, although by the early-to-mid 1990s it seems fair to say that the combination of rapidly developing school effectiveness studies together with a government-led focus on the value-for-money of further/higher education had fundamentally changed the attitude of most educational managers towards sophisticated quantitative data and its relevance to the everyday life of schools and colleges, the acceptance of the term value added in an educational context should not be taken to mean that everyone understands the same thing by it (nor, indeed, understands it *tout court*).

Moreover, the evidence suggests that no final definition could ever be arrived at, since value added is a term which contains within itself an evaluative function: its meaning and therefore its definition (as well as its application and associated methods of measurement) will depend on the purpose for which it is to be deployed in any given context.

It is no accident that education has borrowed the term from economics, since there is a consensus in most countries of the world that education makes a direct contribution to national economic performance and is thus a sub-system of the economic system;

'Value added': a critical review

however, no convincing model of how that contribution can truly be measured has yet been developed, and so educators and economists alike, and economists of education, have continued to use proxy variables such as qualifications, often without question. Even so, it seems to be possible to say that the ideas contained within the usages of 'value added' in an educational context contain the following items (variously combined and/or emphasised) of assumption and evidence:

- ♦ educational performance – as a measure of standards and quality – matters not only to the individual but, more importantly, to the health of the national economy;
- ♦ measures of standards and quality, and hence of the contribution of education to the economy, are necessary; it is not enough to rely on the assertions of educationists;
- ♦ value added is a meaningful term in an educational context and can be deployed in the assessment of standards and quality;
- ♦ institutions make a difference to pupils' educational performance;
- ♦ it is possible, and from many points of view desirable, to make comparative assessments of institutional performance; in consequence, some institutions can be shown to be measurably more effective than others in terms of specified pupil outcomes;
- ♦ it is possible to identify and account for factors beyond an institution's immediate control which are implicated in its pupils' performance;
- ♦ it is possible to calculate expected levels of performance based on these factors, and then to assess individual institutions' results against these.

It is now appropriate to scrutinise the academic literature more thoroughly, for a further exploration of the ambiguities identified in this section, for a discussion of the methodological developments in measuring value added and for an understanding of how far value added has a useful contribution to make to the understanding of educational quality.

3. ANGELS ON A PINHEAD? REVIEW OF THE ACADEMIC LITERATURE

3.1 Introduction

This section reviews key texts in the academic literature on value added in the educational sphere since the mid-1980s, with the aim, first, of exploring further the argument developed in the previous section about the principles of, and ways of doing, value added analysis; and secondly of identifying the strengths and weaknesses of the approach in the context of assessing 'school effectiveness'. In doing so, it shows how value added has shifted in function: starting as a contested methodology concerned with making better (fairer and more valid) comparisons between schools and 'owned' by researchers on behalf, so to say, of educational stakeholders generally, value added has now become a management tool concerned with raising pupils' attainment and internal school improvement, owned by national government, LEA advisers and school senior managers on behalf of an education-management culture. (That this latter development has been extant for some years in at least some states in the USA is demonstrated in Webster *et al.*, 1994.)

The line of inquiry it seems desirable to pursue can be posed as a series of questions, thus:

- ♦ what meaning(s) of value added seem to have been pressed into the service of educational effectiveness discourse?
- ♦ what kinds of analytical models and methodological strategies have been developed as a consequence?
- ♦ in turn, what impact have these models and strategies had on the way effectiveness has been understood in practice?
- ♦ what have value added analyses revealed about 'effectiveness' and educational quality? Are these findings valid? Are they interesting and important? To whom?
- ♦ what further questions and queries have value added analyses raised? In particular, what do value added analyses leave out of account and what serious weaknesses, if any, do their models contain?

The section discusses these questions in some detail. Section 3.2 investigates the provenance of value added principles in the context of research in the compulsory education sector; section 3.3 comments, in a historically organised narrative, on the

relationship between school effectiveness literature and the notion of value added. Sections 3.4 and 3.5 each offer a brief discursus on value added methodology and on the contribution of value added to school effectiveness evidence respectively. Section 3.6 concludes the section by recording the limitations of value added approaches and then revisiting the proposition made at the beginning of the section, that value added has shifted its functional meaning over time.

3.2 The Development of Value Added Principles in Educational Research in the Compulsory Sector

It was claimed in the previous section that the calculation of 'value added' measures in the school effectiveness context was initially and primarily concerned with making fairer and more valid comparisons between schools in terms of their pupils' academic performance, the impetus for which came from the Government's decision to publish comparative tables of school performance. If the relevant literature is examined more closely, ample evidence can be found that this is indeed so, although – to understand the issues properly – the distinct issues integral to that general theme must be identified. These may be thought of as follows:

- ♦ the measurement of educational quality and the development of educational performance indicators (see Mortimore and Stone, 1990);
- ♦ the problems associated with attempting to compare schools' effectiveness on the basis of published examination/test results (see Goldstein and Cuttance, 1988);
- ♦ the interpretation of schools' test/examination results, and especially of the differences between schools in this respect (see Gray *et al.*, 1990);
- ♦ 'disentangling the variety of influences on a pupil's examination performance' (Education, 1992).

It is plausible to construe the Education Reform Act of 1988, and in particular the recommendations of the Task Group on Assessment and Testing (TGAT) (GB. DES and WO, 1988), as providing both a focus and a momentum – 'the moment' – for research which up to that point had comprised innovative but isolated projects on what was characterised as 'the school effect'. Much of this work had been overtly concerned with an equal opportunities agenda, with specific reference to selective versus non-selective schools (for example, Gray *et al.*, 1984; Maughan and Rutter, 1987) or to the achievements of minority ethnic group pupils and/or pupils in inner city schools (for example, Maughan and Rutter, 1986; Smith and Tomlinson, 1989).

Some work was also already being done on developing systems for comparative performance analysis (see Gray *et al.*, 1986).

It was due to the strength of the conceptual and methodological frameworks thus established – supported by hierarchical modelling techniques (see Raudenbusch and Bryk, 1986; Goldstein, 1987) – that the implications of the Government's radical proposals in 1988 for a national system of assessment and testing, and its intention to publish the results school-by-school to assist the extension of parental choice of school, could be challenged so quickly and cogently by academics.

Thus before the end of that same year a short paper by Goldstein and Cuttance (1988) was published in which the authors explicitly forewarned of, and tried to forestall, what they called the 'handicapping system' (p. 200) which would in all likelihood do an injustice both to schools as would-be providers and to parents as would-be selectors of educational quality. The authors argue that if the proposed system were to make comparisons between schools on the basis of pupils' attainment in circumstances where there was variation, either social and/or academic, in pupil intakes to schools – and it could be assumed that in the real world there *would* normally be such variation – this would mask the true extent of the progress made by pupils in different schools, as distinct from the standards reached by them.

The basis for their argument was the 'vast research literature' demonstrating that 'children from socially disadvantaged areas tend to have lower exam scores and test results than those from more socially advantaged areas' (p. 197) and, even more importantly, that 'the attainment of ... children when they first enter the school is the single most important determinant of subsequent achievement' (also p. 197). If these factors were not taken into account, the performance of schools could be neither fairly assessed nor validly compared, as TGAT themselves intimated. But, according to Goldstein and Cuttance, TGAT could be accused of wanting it both ways: on the one hand, failing actually to mention the school effectiveness research so far published whilst on the other – by recommending that local socio-economic (SE) factors should be published 'alongside' schools' results – passing the buck to parents and LEAs for making the necessary adjustments to schools' results in order to get a truer picture of their effectiveness. The authors conclude (p. 201) that:

the system of school evaluation proposed in the Report [of TGAT: Black, 1988] would be quite misleading, and school boards, parents and teachers would be right to protest loudly at the use of such poorly constructed performance indicators.

Smith and Tomlinson (1989, p. 303) who must have been writing their conclusions in the same year, protest in a similar vein as follows:

[Our findings] show that a comparison of the raw test results between schools would be highly misleading... If the results of the tests are to be made publicly available, it is essential that they should be analysed by methods akin to those used in the present study [i.e. using multivariate analysis]...

They too go on to say:

There is no support in the findings of this study for the idea that increasing parental choice will improve standards... it is quite clear that parents cannot identify the schools that are doing well in terms of pupil progress. This is hardly surprising, since it takes a complex analysis to identify those schools.

There are other key points made by these authors which must be noted in an historico-narrative account such as this. Presciently, Goldstein and Cuttance introduce their paper by warning that 'publication [of comparative school results] could well become the most visible and perhaps most important aspect of the whole system, irrespective of the actual content of the assessment or indeed of the national Curriculum itself.' (p. 197.)

On more substantive matters, Goldstein and Cuttance point out that there are flaws in the system at a basic level, in that a comparison of school averages:

tells us nothing about the relative achievements of different types of pupils within the schools... Consequently schools which perform well relative to other schools for the average pupil in the population may perform less well for disadvantaged *or* advantaged pupils. (p. 198) [Our emphasis.]

This issue – of differential effectiveness – if found to be generally applicable would obviously have consequences for both the public reporting of school comparisons and for parents' choice of school for their individual children. Concern about the urgent need for 'Fair and comprehensible ways of presenting performance indicators in context, and to reveal differences between sub-groups of pupils,' fuelled another influential paper of around the same time (Nuttall *et al.*, 1989, p. 769). Based on the findings from a multilevel analysis of a large dataset held about secondary schools in inner London, it concluded that:

...school effectiveness varies in terms of the relative performance of different sub-groups. To attempt to summarise school differences, even after adjusting for intake, sex and ethnic background of the pupils and fixed characteristics of the schools, in a single quantity is misleading... [T]he concept of overall effectiveness is not useful.' (pp. 775-6)

On a second crucial issue, Goldstein and Cuttance point out (p. 198) that ranking schools in order of results – whether raw or adjusted for social background and/or prior attainment – is in any case highly problematic because the method used for arriving at estimates of school performance partly determines the ensuing rank order. We may assume that 'method' in this instance covers both the issue of what outcomes are used – particularly those which can differentiate between different groups of pupils (see above) – and the question of what statistical modelling techniques are employed. These matters are discussed in more detail in the sub-section following.

Thirdly, these authors say that 'the more background factors which are taken into account, the more unstable and unreliable become the resulting comparisons'. This is an important principle to keep in mind, because there have been calls – understandably from the point of view of those seeking to manage and 'improve' schools – for a model which would include everything which might be thought of as a factor affecting schools' performance. But actually the underlying principle of most value added work is an eliminative rather than an accumulative one, relatively simple in conceptual terms though complex statistically. Its relationship with school effectiveness is therefore indirect: its main role has been not so much to identify a range of effectiveness correlates but to discern different kinds of 'noise' or extraneous information in the analyses of effectiveness, and to get rid of it as far as possible.

Finally, a point made by both Smith and Tomlinson and Goldstein and Cuttance (and which can be easily inferred from Nuttall *et al.*) is one that follows inexorably from the ability to make fairer and more valid comparisons between schools on the basis of pupils' progress, but which tends to be obscured by the term 'value added'. Smith and Tomlinson put it like this:

A school having a low balance of intake could appear to be doing badly, when in fact it was doing well; while a school with a high intake balance could be flattered by the raw test results (p. 303).

This relates to another principle which it is important to bear in mind: that, since value added analyses are norm-referenced in the way they are calculated, there will inevitably

be 'losers' as well as 'winners'. This implication is not always understood in the cut-and-thrust of getting schools' public relations messages across: head teachers have been heard to say things like: 'we just want to be able to prove what we already know, that all our teachers are adding value'.

Nuttall *et al.* mention in passing another 'major' issue for investigators of school effectiveness, that of the stability of school effects over time (p. 770). This has since become a key area of research and analysis.

Two rather later articles are also worth citing in this context, Paterson (1991) and Schagen (1991). Paterson's detailed paper (cited by McPherson, 1992, below), whose findings are based on a multilevel model of school effects and their association with both socio-economic status (SES) and prior attainment of pupils, concludes:

Any system of monitoring schools requires attention to be paid to SES if it is to inform the practice of education. Equally, though, the measurement of SES has to incorporate studies of the detailed ways in which the social and economic circumstances of school pupils influence their opportunities and incentives to learn. Simply knowing that SES is associated with attainment – even in the relatively subtle ways which we have used here – is merely a first step towards understanding why it is so, and towards developing practicable ways of overcoming its harmful effects (p. 117).

In other words, the fact that socio-economic status is empirically associated with attainment does not mean we are entitled to infer an *a priori* association, still less a causative one. Moreover, Paterson suggests that there is perhaps an even more fundamental conceptual issue at stake:

... one way of viewing this problem of membership [of neighbourhood, occupation and/or labour market] would be as an opportunity to test various theories about the ways in which individuals can be said to belong to groups... [But] it is likely that some of the multi-dimensionality has appeared because a cross-section of pupils captures parents who are at different points in different types of occupational or family careers. The multi-level nature of social processes has the consequence that time itself is multi-level (p. 117).

This hints that the interplay of educational and social factors is really a complex narrative being played out through time at the individual, family, school, local and national levels simultaneously is a very useful reminder of the challenges posed by the

social sciences to a reductionist 'common sense' view of issues like school performance and effectiveness.

Schagen's (1991) article by contrast is brief and practical. The reason for including it here is that it aims, by giving clear worked examples, to 'show that, given suitable data and the use of appropriate statistical techniques, it is possible to use numerical outcomes (e.g. test results) to obtain, valid and helpful results about schools and their effects on children.' (p. 216) Premised on the claim that 'league tables of schools based on a single output measure are statistically dubious' (p. 222), it discusses the operation, and the strengths and weaknesses, of both multilevel modelling and data envelopment analysis.

A key point to be made is that none of the five sources discussed above once uses the term value added. And yet everything in the development of value added analyses since then confirms the truth of their insights painstakingly arrived at a decade ago. In unravelling the history of value added, one is therefore more or less impelled to 'read back' into their research concerns and strategies later understandings and usages of value added.

3.3 Charting the Relationship Between 'Value Added' and School Effectiveness

It seems to be in the post-ERA research literature from 1991 onwards that we need to search for evidence of how 'value added' has become part of the discourse on educational effectiveness, and on what terms. Of the many such texts authored by researchers and others, only those which, first, specifically allude to 'value added' approaches, data and/or analyses in their titles or abstracts and, secondly, provide one or more of the following, have been selected for discussion in this sub-section:

- ♦ first-time published use, or clarification, of key concepts and assumptions;
- ♦ important differences from the consensus or conventional view;
- ♦ description of techniques for measuring value added;
- ♦ critical overview of value added approaches in the context of the broader research issues involved.

The key texts selected on these criteria are, in chronological order of publication³, Nuttall, 1991; Kennedy, 1991; Jesson, 1992; McPherson, 1992; Nuttall *et al.*, 1992; Wiliam, 1992; Birnbaum, 1993; Fitz-Gibbon and Tymms, 1993; Cornall and Lofthouse, 1993; Gray, 1993; Mortimore *et al.*, 1994; Schagen, 1994; Fitz-Gibbon, 1995; Goldstein and Thomas, 1995; Gray, 1996; Jesson, 1996; Sammons, 1996; Thomas and Mortimore, 1996; Thomas *et al.*, 1997. Some other articles and papers which use the term 'value added' (and are often interesting enough in their own right) amount, for current purposes, to glosses on earlier or more detailed texts and so we have chosen not to discuss them at this point, although references are sometimes made in passing. Yet further articles are important later on in the discussion of the application of value added findings to school improvement, and are dealt with in the appropriate section(s).

Nuttall's (1991) article in the Times Educational Supplement is a plea, couched in lay-person's language but attempting to summarise the research intelligence so far, that educationists should quickly demonstrate the feasibility of establishing systems for analysing pupil performance in terms of their relative progress, as opposed to accepting the principle of raw league tables. He says:

For nearly 30 years, research on school effectiveness has used the *progress* made by pupils from their level of performance on entry to their level of performance at the time they leave, rather than just their 'raw' results at the time of leaving. This approach, well illuminated by the metaphor 'value added', makes intuitive sense and is readily comprehensible (p. 22).

As well as this insistence on the need for measures of progress, the article makes several other points about the requirements of the value added 'principle' or 'model', viz:

- ♦ the need for a range of outcome measures (not merely five or more A-C grades at GCSE, for example);
- ♦ the need to allow for SE variables;
- ♦ the need for analyses of differential effectiveness;
- ♦ the need to allow for the possibility that schools' results are not stable over time;
- ♦ the need to use multi-level modelling as the statistical technique.

³ The order of publication may not always, of course, reflect the order of composition.

One may take it that these principles are based on the analysis, and its implications, of the 1990 GCSE examination results undertaken by Nuttall *et al.* for the Association of Metropolitan Authorities, although not published until 1992. Kennedy's (1991) article for *Managing Schools Today* is less clear about the prerequisites for calculating value added; but this may not be altogether surprising given that he was writing from the perspective of the Audit Commission which at that time was engaged in exploring approaches to value added in the 16-19 sector – a sector that has remained to this day intractable in terms of establishing appropriate baselines and inputs for calculating value added. As we saw in the previous section, the definition of and methodology for assessing value added in F/HE tended to be connected with a less problematic notion of 'progress made' than that which was emerging from the longer and stronger tradition of school effectiveness research. Thus Kennedy, in taking issue with league tables of raw results, claims that 'a sounder basis for comparing institutions is the progress made by pupils, which will reflect the "value added" by their education.' (p. 13) It is evident from what he goes on to say that 'progress made' is equivalent to the 'distance travelled' by different pupils in different institutions between, say, their GCSE grades and their A-level grades. He notes, however, that 'no generally accepted robust methods for measuring the value added have yet been established' and that 'a number of key technical questions' would need to be resolved before the extent of the likely differences between the performance of individual schools could be gauged. He makes no mention of the well-known work on GCSE to A-level comparisons piloted and developed through the ALIS (A-level Information System) service since the early 1980s by Fitz-Gibbon, then at the Curriculum, Evaluation and Management Centre (CEM) at the University of Newcastle.

McPherson's (1992) briefing paper for the National Commission on Education (chosen as the first in a long and generally impressive series of Briefings for a general audience, and mentioned in Section 2 above) is more sanguinely entitled *Measuring Added Value in Schools*; it does a good job of summarising existing research in order to make the case for a more balanced and complex view of school comparisons. Whilst noting that there is a place for raw outcome measures as a reflection of actual attainment, his main argument is that 'a bad indicator system', such as that based solely on raw test or examination results, carries the hidden costs of 'mistaken judgements, needless anxieties and fruitless "further investigations"... triggered by false signals'. The measurement of added value – explicitly defined as the 'calculation of the contribution schools make to pupils' progress' – is central to developing a good indicator system. Such measurement must be predicated on 'an explicit theory of good standing' which is open to scrutiny and refinement, and includes such features as:

- ♦ pupils' prior attainment;
- ♦ the longitudinal nature of progress;
- ♦ the multilevel nature of schools;
- ♦ the multivariate nature of the factors involved, especially 'non-school factors that boost or retard progress', such as pupils' socio-economic background;
- ♦ differential effectiveness for different groups of pupils.

This is a more careful and weighty operational definition than that given by Kennedy, and shows an understanding that some of the 'technical questions' alluded to by him were capable of resolution at least in principle, even if the practical collection of necessary data was still problematic. McPherson provides a diagram – based on Paterson's (1991) research – showing how schools' relative performance becomes more similar, and how their relative positions change, the more adjustments for non-school factors are made. Moreover, McPherson seems to be the first writer to acknowledge the role that value judgements will always play in arriving at a 'good' information system on schools' results; and that what is useful information to parents in choosing schools may be less useful for teachers in setting about raising attainment, and *vice versa*. But 'useful' information is not inevitably simple information, even – or especially – where parents are concerned:

...complexity is not in itself an argument against aiming for the best possible indicator system. Any attempt to improve schooling by means of informing choice presupposes that parents are capable of understanding at least the complexity of an adjusted outcome score. To reject that possibility is to reject the possibility of informing parents (p. 4).

Three of the Digests published by the journal *Education* in the early 1990s dealt with value added for a non-specialist but professional audience, mainly LEA and school managers. The first of these (Jesson, 1992) is clearly presented and argued, and uses analyses of GCSE results in different ways to illustrate the importance of the following principles:

- ♦ the need to base calculations on individual (not aggregate) data;
- ♦ the need to take account of prior attainment;
- ♦ the need to take account of gender;
- ♦ the need to use multilevel modelling;

- ♦ the need to use the model 'not simply as a means of generating a more appropriate league table of "effectiveness" or "value added" but as a resource to prompt genuine questions of comparison between practices and procedures in contrastingly effective schools.' (p. ii)

Jesson seems to discount the use of a social deprivation variable, on the grounds that if used 'as the only context within which examination results are evaluated, ... it can become an excuse for not expecting high achievement.' (pp. i-ii.) Jesson is not alone in expressing this concern (see, for example, Wiliam, 1992, cited below). But Jesson also supports the use of multilevel modelling as being 'essential to do justice both to the underlying relationships in the whole population of pupils as well as to the particular contribution made by each school'. The point about multilevel modelling is precisely that it can include several background variables simultaneously and assess their relative importance in accounting for outcome variables. In any case, we saw that at least one commentator (Paterson, 1991, cited above) appears to have recognised that empirical relationships such as that between attainment and SES may be construed, not as evidence of an unchangeable fact, but as support for the necessity to change educational and/or social policy.

The anonymous author of the second *Education Digest* was not so convinced as Nuttall, McPherson and Jesson of the transparency of value added approaches. The article starts by warning that the interpretation of value added results 'depends on a number of important assumptions and qualifications which are not particularly easy to understand' and justified the emphasis on technicalities in the article by saying: 'However, if these techniques are going to be the main way we statistically analyse pupil achievement in the future, then it is vital that everyone in education understands what they can and cannot do'. (p. i) The article is based largely on a discussion of procedures used in one outer London borough on one year's GCSE data, however, rather than on a critical overview of research evidence; and although the commentary makes some generally useful points about the importance of statistical validity and the role of the null hypothesis, as well as the varied influence of socio-economic factors, the text is too dense and laborious to fulfil its stated purpose. In concluding, the author takes up a position on the use of the most up-to-date techniques (i.e. multilevel modelling) which contradicts that of McPherson and Jesson:

'[The multilevel approach] is the most difficult to understand for the non-statistical reader... One should not underestimate the importance of this... Even the simpler approaches to value added can appear arid and remote... For this reason it may be that the use of multilevel analysis is not the best place to start.' (p. iv)

(Rather oddly, this Digest turns up, in slightly modified form, sixteen months later as yet a third Digest, authored by Birnbaum (1993), though without reference to the earlier version.)

The fact that Nuttall's, Kennedy's, Jesson's, McPherson's and Birnbaum's pieces – all using the term value added as if it were the *mot juste* that still required a little elucidation – were written in the early 1990s for a non-research audience is worth considering. As has been remarked, the relative lateness of the usage in compulsory education as compared with its established quasi-technical use in HE (discussed in Section 2 above) contrasts with the maturity of the research discipline which lay behind it. This suggests that the term 'value-added' was lifted from HE to describe a set of issues now of critical concern in the compulsory (and immediately post-compulsory) sector, but that it was initially reserved for generalist rather than specialist discussion because it seemed to be an idea whose common sense and common justice people could 'intuitively' (Nuttall's word) grasp.

So far as it is possible to tell, then, the term appears to have come into generalist use in work relating to the compulsory sector *before* its adoption in more scholarly discourse. This is hard to establish with complete certainty because of the time that often elapses between the writing and the publication of articles in academic journals. In 1992, Wiliam's article for the *British Educational Research Journal* was addressing 'some of the technical difficulties in operationalising the notion of "value-added" in [the published results of aggregated school performance] (p. 329)' particularly as they related to National Curriculum assessments. (Wiliams gives a detailed discussion of the nature and structure of these assessments, which, whilst conceptually helpful about the limitations of the assessment framework, does not directly concern the present discussion.)

Wiliams uses quotation marks throughout to indicate the provisionality of the term 'value added', which he does not explicitly define himself but for which – since he cites McPherson (*op. cit.*) – he can be taken to have adopted McPherson's definition. He agrees, on certain conditions, with McPherson's statement that socio-economic factors as well as prior attainment may influence pupils' performance; but goes on to argue on theoretical and ethical grounds that SES data should not be included in a 'value added' analysis of performance. Although McPherson argues for 'an explicit theory of good standing', he says, no such theory at present exists with regard to the type of SES indices that should be included. He argues that different studies have used a wide range of different variables, with differing results in terms of the correlations found.

Furthermore, Wiliam considers that the act of computing socio-economic correlations known may have social consequences of the 'self-fulfilling prophecy' type or, in Wiliam's words: 'it is extremely doubtful that the use of socio-economic factors in value-added models will do anything to combat under-achievement by pupils, and may serve to consolidate it.' (p. 338.) Given an adequate explanation of what value added analyses do *not* reveal as well as what they do, this should be a misplaced anxiety, albeit one shared by other commentators such as Jesson (1992). Wiliam also disagrees with the suggestion made by McPherson and Nuttall that differential effectiveness is important to report on, because: 'the use of multiple indices is likely to cloud, rather than clarify, the picture' (p. 339). Whilst one may in retrospect therefore disagree with some of Wiliam's argument, his major contribution in this article was to raise the taxing practical issues – which the School Examination and Assessment Council (SEAC) had so far avoided – associated with using the new National Curriculum assessment framework for comparing pupils' progress in different schools. Most of the school effectiveness research at that time was by contrast based on public examinations and/or standardised test results.

The two Discussion Papers on value added issued by the Centre for the Study of Comprehensive Schools in 1993 (Fitz-Gibbon and Tymms, 1993; Cornall and Lofthouse, 1993) are chiefly interesting for the fact that, again, they are written for a non-specialist audience and with the weaknesses of raw league tables very much to the fore. It is perhaps unfortunate that the papers do not relate well to each other, and indeed on one or two questions – such as whether SES factors should be taken into account in calculating value added or if individual schools can 'do' value added analyses for themselves – offer conflicting answers. On the specific issue addressed by Wiliam (1992, see above), Fitz-Gibbon and Tymms are sarcastically dismissive of the possibility of using National Curriculum standard assessment task (SAT) data for calculating value added; Cornall and Lofthouse write abstractly about the use of 'pupils' test scores'. The latter paper is really only a polemical statement of position and the authors do not attempt to tackle seriously the conceptual and technical problems raised by other commentators. The reason for inclusion here is the context, i.e. briefings for schools which might be thought to be particularly vulnerable to raw league table results.

In his report for the Scottish Office, Gray (1993) states the purpose of the value added approach as 'comparing like with like' (p. 2): 'if there are two pupils who are identical in every respect how much difference does it make to their educational progress if they attend different schools?' With the benefit of work piloted and developed over a

number of years with different LEAs, Gray gives the basic requirements for building a value-added approach, whose similarities with criteria offered by Jesson (quoted above) are without doubt due to their collaboration over a number of years. Gray's text, slightly abbreviated from the original, runs as follows:

- ♦ data on individual pupils rather than aggregate data, including:
 - a measure of outcome for each pupil which reflects all levels of pupil performance
 - a measure of each pupil's attainment, preferably one which is finely differentiated, plus one or two other items of information about pupils' background
 - in the absence of prior attainment, several items of information about each pupil's background (ideally including a measure of social advantage as well as social disadvantage);

- ♦ analysis of data using multilevel modelling.

The report, though short, is the first publication (as far as we are aware) to cover issues and challenges in the *implementation* of value added approaches, as well as some discussion of basic technicalities. Amongst the sixteen questions he identifies for schools and LEAs to consider, for example, Gray points out that:

the most important issue to be addressed relates to the overall purposes of introducing value-added approaches. Is the main purpose to hold schools directly accountable for their performance? Is it to identify schools that are markedly under-performing and may even be 'at risk'? Is it to help schools justify their existing levels of performance? Or is it to help them to understand more about their own performance and how to improve it? If it is to be about several of these, are they all equally compatible? (p. 9)

The significance of this report lies not merely in the clarity and usefulness of the content but in its audience: the work was commissioned by the Audit Unit in the Scottish Office Education Department in order to inform the school development planning initiative in Scotland at a time when the English Department for Education was still acting as if afraid that anything other than raw results would be too complicated for public consumption and would encourage schools to make excuses for poor standards.

But value added was a persuasive idea and government agencies were slowly convinced of the need to have better measures of school performance than raw results could provide: the political watershed was marked in 1994 with the publication by the

School Curriculum and Assessment Authority of *Value-added Performance Indicators for Schools* (discussed in its own right in the following section). As the familiarity of the idea grew from 1994 onwards, the term, together with its functions and supposed advantages, became more or less taken for granted. Instead, as the literature shows, attention turned to other underlying and possibly intractable issues which needed further explication. Paradoxically – in view of the arguments initially advanced for value added by many of the commentators discussed above – one of those issues was the question of whether value added analyses could really be used for public accountability systems.

The immediate occasion for the 1994 paper by Mortimore *et al.* was the death of Desmond Nuttall: the paper was intended to be a tribute to his work and to that end the paper offers a summary of the main issues relating to school effectiveness and the measurement of value added, Nuttall's theoretical and practical contributions to which had become internationally recognised. This paper is explicitly situated in the school effectiveness tradition of academic research, and is important for the overview it gives of this research (albeit with a particular emphasis on Nuttall's work). It is also interesting, in this scholarly context, for its unremarked use of the term value added in the title and throughout. A definition of value added is not attempted (presumably because by now not felt to be necessary), but the term is used with a positive connotation, in the context of the provision of public information, thus:

Raw results tell parents the grades their children have obtained but they can say nothing about how well the school attended has performed. In contrast, value added results tell parents how effective their children's school is in promoting achievement (p. 319).

The paper is exemplary for the way it foregrounds the areas where there are as yet unresolved issues and provides examples of conflicting evidence – in relation to stability of school effects over time, differential effects for different groups of pupils and the effects of whole-school contextual factors over and above individual pupil characteristics – to demonstrate 'the varied and unpredictable impacts of schooling.' (p. 329). It calls both for further research into these impacts and for a broader range of outcomes to be analysed beyond the narrowly academic.

A contrasting note is sounded by Saunders and Schagen (1994) who take issue with the way value added has become a new 'buzz-word' in education:

[value added] has become a rallying standard to signify opposition to crude comparisons between institutions without taking any account of the pupils with whom or the circumstances within which they are working. In the process, however, the meaning of the term has become understandably rather blurred (p. 1).

They say that 'much useful analysis is more properly classified under the general title of "school effectiveness". The rest of the paper is a description and evaluation of how the ideas and techniques in their earlier work (Schagen, 1991) have developed operationally, through the piloting of the QUASE (Quantitative Analysis for Self-Evaluation) Service for secondary schools, sponsored by the National Foundation for Educational Research (NFER). This service was set up to operate on the same principles of 'good practice' as ALIS (mentioned above) which are that:

- ♦ it provides a confidential service to schools in which schools are given only their own results and no league tables are formed;
- ♦ results are fed back 'in a digestible form which can influence educational practice' (p. 1).

Saunders and Schagen record the results of the QUASE feasibility study as follows (p. 7):

- ♦ it was possible to combine results from disparate tests at age 11 or 12 to give a measure of pupils' prior attainment on entry to secondary school;
- ♦ multilevel models using such intake measures, plus school context data, can be used to provide schools with sensible performance information;
- ♦ the feedback from the QUASE pilot project was appreciated by schools and frequently tied in with information from other sources;
- ♦ some useful feedback can be provided to schools with no intake data, but this is necessarily more tentative.

They do not mention here that QUASE uses a series of separate departmental indicators as well as seven generic ones like total and average GCSE scores, thus operationalising, at least in part, the notion of differential effectiveness. Saunders and Schagen conclude that:

The quantitative results should not be seen as ends in themselves but as a contribution to the process of school improvement. There is a need to work interactively with schools to help them integrate

quantitative information of this type into their school development planning (p. 8).

Overall, the paper shows, in practical terms, how the criteria and principles agreed by people like Goldstein, Gray, Jesson and Nuttall may be operationalised.

Fitz-Gibbon and Tymms (1995) is another milestone, in being the first of the many reports emanating from the so-called Value Added National Project. (The final report and accompanying technical reports are considered in Section 4 below, as belonging at least as much to the policy as to the academic arena.) In it, this definition of value added is given: 'Measures of ... relative progress have come to be called "Value Added".' (p. 2). This is elaborated on as: 'If some pupils make greater progress than other, similar pupils, the difference is referred to as the Value Added.' (p. 7) As has been argued in Section 2 above, this is not the most rigorous definition, because it fails to make explicit the need to identify *schools'* contribution to pupils' relative progress; but it has come to inform most of government thinking since. To summarise a substantial piece of work, it may be noted here that the report considers issues of variables, coverage, analytical method, reporting and interpretation for each of the different approaches selected for comment, and where appropriate discusses the pros and cons involved. It concludes that: 'the statistical trialling undertaken to date has lent assurance to the view that simple, understandable and statistically acceptable Value Added indicators can be made available when suitable input and output measures are in place over reasonable timespans.' (p. 29) One of its findings – at odds with what was said in the earlier SCAA report (1994, p. 84) – was that the National Curriculum assessments could provide such measures. Given the *caveats* made by the CEM team in the past about over-reliance on performance data and especially National Curriculum assessments for school evaluation purposes, it is noteworthy that – with the exception of school data used for the pilot studies about to commence, which was to be kept confidential – the purpose of a national value added system was at least in part to publish the results for 'parents and the public' (p. 19).

The view that value added data could and should be in the public domain was not shared by Goldstein and Thomas, 1995, who by the middle of the decade had arrived at the position that even the 'private' use of value added data by individual schools may be misleading. They point first to the inevitably historical nature of performance data by the time test or examination results are available to be used as outcome variables, such that schools may have completely changed their policy or practice in the meantime; and then to a much more serious problem which they state like this:

While we can study the factors associated with pupil performance and come to conclusions about which of them appear to be associated with 'success', yet it seems to be very difficult to identify precisely which schools are doing well or badly... In other words, research into school effectiveness is a useful activity in our attempts to obtain knowledge about the process of education, but a very poor tool for holding schools to account (p. 37).

Gray's (1996) concern is similar to that of Goldstein and Thomas, and 'is premised on the assumption that ... comparisons [between schools] will continue and that the most helpful service can be rendered by reminding those who would make judgements of some of the problems to be encountered.' (p. 121) In effect, the chapter is an elegant and concise summary of the technical and conceptual issues discussed elsewhere, and thus requires no further description. Its significance lies more in the fact that it is another example of the phenomenon that commentators who appeared to have much in common in the early 1990s are, by the latter half of the decade, taking up opposing positions.

Jesson's (1996) report performs the same sort of function for the DfEE as Fitz-Gibbon's did for SCAA, that is, it reports on the feasibility of using current assessment data, in this case Key Stage 3 Assessments, 'for evaluating differences in performances by pupils and their schools' (p. 12) at GCSE. Broadly, there was found to be sufficient correlation to provide 'a considerable opportunity for comparative evaluation of institutional performance using a common "starting point"' (p. 12), although the author made it clear that he did not yet regard this as a reliable and consistent guide to value added assessments of school performance.

Thomas and Mortimore's 1996 study involved 'considering and evaluating a variety of different models for measuring school effectiveness using sophisticated statistical techniques (multilevel modelling).' (p. 6) They particularly contrasted the results of controlling for prior attainment with those of controlling for SES factors. Amongst the findings, they report that: 'when prior attainment data are available no school context factors are significant and the fit of the model is substantially improved.' Most but not all studies, however, have shown some negative correlation between low SES and academic performance: Paterson (1991) has a thorough discussion of statistical analyses purporting to show that 'SES is an important correlate of attainment over and above ability, and that the SES mix of the school has an effect on attainment over and above individual or family SES' (p. 97). Sammons (1996) found that: 'by controlling only for prior attainment... a misleading picture of effectiveness was obtained for some individual institutions' (p. 119), and Schagen (e.g. 1996) has found consistently that

although prior attainment data explained a great deal of the variance in schools' GCSE results, allowing for school context factors – including parental attendance at parents' evenings as well as overall proportions of FSM – explained even more of the variance.

But, over and above the question of whether there is any statistically explanatory power in doing so, Thomas and Mortimore offer another version of the conceptual rationale for excluding context factors. They too argue that school socio-economic context may influence teacher expectations of pupils, with a resultant impact on pupil outcomes. This argument is addressed in Section 3.4 below.

Usefully, Thomas' and Mortimore's study also confirmed the likelihood in most schools of differential effectiveness, for different subjects and different groups of pupils. They conclude with a call for further research to explore 'the relationship between negative value added results and measures of school processes, in particular, the quality of teaching and learning.' (p. 28)

Thomas' *et al.* (1997) report on research undertaken, using what they call 'a value added approach' (i.e. multilevel modelling to control for pupil intake measures) into the stability and consistency of schools' GCSE results. Their conclusion is that 'effectiveness is best seen as a feature which is outcome and time specific' (p. 194). They summarise the current position on school effectiveness as showing that any evaluation of performance needs 'to address three key questions: (1) Effective in promoting which outcomes? (2) Effective over what time period? (3) Effective for whom?' (p. 194). As Sammons (1996, p. 143) has it, 'school effectiveness is perhaps best seen as a relative term which is dependent upon time, outcome and student group'.

By the late 1990s, then – the point at which the government had chosen to espouse the value added principle – there appears to be considerable agreement in principle about why the value added approach was found necessary in the first place. There is somewhat less consensus on what information of the kind that could loosely be called value added would most help teachers in improving, and parents in choosing, a school and what it consequently means in terms of methodology. The initial reason given by commentators for undertaking value added analyses seems primarily to add to, modify or otherwise elaborate on the published tables of performance based on raw results. The market principle that parents could and should be making informed choices about schools for their children thus seems to have been accepted by researchers including many of those discussed above. So we can say unequivocally that one of the intended

functions of the value added approach was to render public information about schools' performance more accurate in a whole variety of ways for that purpose.

Another related, and for some head teachers an over-riding, function was to enable schools to mitigate the message given by their raw results, to counteract bad press. As Buck (1993) put it: 'This exercise [a method for contextualising GCSE results in one LEA]... provides a fast response to potential condemnation in the public domain.' (p. 92) So far, then, there was an acceptance that the role of value added analyses was directly related to the reconstruction of published 'league tables'. Since the conversion of the then Secretary of State for Education under the Conservative administration, Gillian Shephard, to the principle of value added, this has continued to be the government's position up to and including the 1997 proposals to establish a national value added system.

But, as was suggested above, it is not a position held by everyone. Amongst others, Tymms (1990), Fitz-Gibbon (Fitz-Gibbon and Tymms, 1993; Fitz-Gibbon, 1994), Saunders (1997), Saunders and Schagen (1994), Schagen (1991 and 1997) and particularly Goldstein (see, for example, Goldstein and Thomas, 1995, Thomas and Goldstein, 1995, and, most recently, Goldstein, 1997a, 1997b) have argued against this public accountability use of value added analyses. The dilemma can be stated like this, that on the one hand the *principle* of value added does seem to have been readily comprehensible, in Nuttall's (1991) words, an idea whose time had self-evidently come. On the other hand, the more rigorous the actual value added *analysis* becomes, in terms of underpinning theory and appropriate modelling techniques, the closer we get, in Thomas and Goldstein's (1995) words, to comprehending 'the complex, multi-faceted and often inconsistent nature of school effectiveness'.

To put it at an extreme, the value added task began by appearing to promise better information for public consumption, but instead turned out to demonstrate that 'better information' and 'public consumption' were incompatible, if the latter depended on there being 'simple and straightforward' measures of progress (SCAA, 1994). As Thomas and Goldstein (*op. cit.*) remark, 'research emphatically demonstrates that the measurement of progress or value added... is neither simple nor straightforward'. It is important also to record that Fitz-Gibbon drew attention, in the Final Report on the national value added project (Fitz-Gibbon, 1997), to grave concerns about its applications in practice. (Such concerns are understandable: scholars may wish to urge caution when they see policy-makers and the public expecting too much from an approach they themselves have pioneered.)

There are two ways of going on from here: one could make the case along with Thomas and Goldstein for further centrally-funded and 'well-designed, long-term research into the processes which shape the achievements of different kinds of pupils in different types and organisation of schools'; or one could follow Tymms and Fitz-Gibbon (Tymms, 1990; Fitz-Gibbon and Tymms, 1993; Fitz-Gibbon, 1995), at least in their earlier articles, Jesson (1996) and Saunders and Schagen (Saunders and Schagen 1994; Saunders, 1996, 1997) in arguing for the installation of 'feedback' systems designed to enable schools to make use of value added information for autonomous development and improvement. (This was the official position of the Scottish Office Education Department as far back as 1993: see foreword to Gray, 1993.)

These alternatives are not mutually exclusive, of course; but at the moment stalemate seems to have been reached, in that – despite the weight of evidence against it – the Labour government is if anything even more determined than the previous administration to pursue the notion of value added league tables, or at least to give prominence to numerical measures of progress and improvement within the public accountability system, without necessarily being committed to funding the more sophisticated methodology which would give such measures more validity and reliability.

The political climate has seemingly led to a situation in which practitioners of value added tend to fall into one of two groups: the pragmatists or the idealists. The pragmatists are those who are inclined to say that the idea of value added is here to stay and the sooner we develop a system which is workable and comprehensible the better. Credibility with teachers and parents and above all user-friendliness for school managers – now charged with making year-on-year measurable improvements to their pupils' performance – are more important criteria than theoretical completeness or statistical purity. The government and its agencies are firmly in this category. Idealists see themselves as keepers of the flame of school effectiveness research, which long pre-dates current obsessions with public accountability and the workings of the educational market-place. They would argue that using procedures which are not as accurate and rigorous as state-of-the-art modelling techniques permit will bring the whole idea of value added into disrepute. The country's education system will end up with a method for making judgements about educational quality on the basis of adjusted performance data which is even more unreliable and disinformative than raw data. For a good-humoured but incisive version of this position, see Gray, 1994. Some, like Goldstein and Thomas (1995), would apparently go even further and almost say that value added, as an assessment of any individual school's performance,

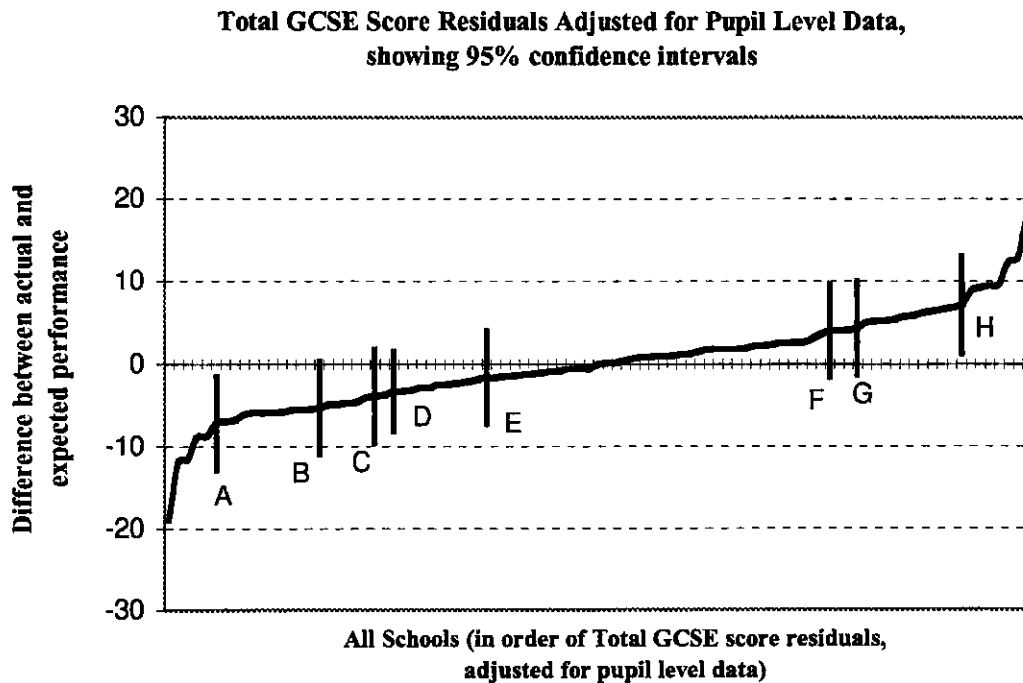
is inherently impossible, because of what they call 'a kind of uncertainty principle' (p. 37). So the wheel seems to have come full circle, with the government supporting value added approaches and some very respected academics arguing against them.

3.4 Summary of Analytical Models and Methodological Strategies

It may be helpful at this point to stand back a little from the historical narrative and summarise briefly what can be said in this about the analytical models and review methodological strategies associated with the measurement of value added. The review does not presume to be technically oriented, however, and what follows is no more than a superficial gloss – mainly for clarificatory purposes – on an extensive body of academic work.

First of all, it must be understood that, as a way of investigating school effectiveness, the application of what have come to be known as value added models has a peculiarity not readily obvious to the classroom teacher or statistical layperson. Unlike other forms of research, which rely on the steady accumulation of evidence, value added has an essentially *eliminative* basis. That is to say, as the author of an Education Digest (1992) pointed out, value added is an attempt 'to make some estimate of the total influence of these ... factors [i.e. factors outside the school's direct control, such as pupils' prior attainment] by a process of elimination.' Value added is the statistical discrepancy – the residual – between the actual score and the score calculated when factors known to have a relationship with performance have been accounted for. Or rather, value *added* is represented by any residual above zero, i.e. with a positive value: since it is part of the modelling process to relate each school's score to an average or expected value derived from the sample population as a whole, some schools or departments within the population are bound to end up with a negative residual. It is now normal good practice to state the degree of certainty with which the residual can be assigned, using some statistical index of significance, usually at the 95 per cent confidence limit. For the most part, school residuals arrived at by the above method are not significantly different, statistically speaking, from the average, that is, from what would be expected (see Chart 1 below for one example; the diagrammatic representation is commonly known as a *Goldstein plot*, for reasons which must be self-evident).

Chart 1 Diagram showing eight secondary schools in an LEA ranked in rank order of their total GCSE scores adjusted for pupil-level data



Only two schools, A and H, have total scores which are significantly lower (A) or higher (H) than would be predicted.

Source: National Foundation for Educational Research, QUASE analysis 1997.

Two or three things need to be said about this, from the point of view of interpreting analyses:

- ♦ Assigning value added to a school's results is equivalent to the theological notion of 'the God of the gaps'. The residuals demonstrate only that there is a positive (or negative) difference from what would be expected; they do not identify what is responsible for that difference.
- ♦ Usually, statistical residuals are quite small. Even so 'in terms of differences between schools in pupils' outcomes they can be highly significant both educationally and statistically... differences... have significant implications for pupils' subsequent education and employment prospects' (Sammons, 1996, p. 120).
- ♦ Since the residuals for most schools are not statistically significant, it is very hard to argue that in the majority of cases schools are performing much better or worse than would have been expected. That this is true does not help the cause of league tables, whose *raison d'être* is to rank schools as if there were a real difference between them.

The second area which needs some further comment beyond what has already been said in this section has to do with whether or not socio-economic (SE) variables should be included in the analysis, provided that there is i) a reasonable proxy measure for SES which can be used and ii) that there are good statistical reasons to believe that low SES is correlated with depressed performance. To deal, rather cursorily, with the first proviso: eligibility for free school meals (FSM) has long been taken as a proxy for low SES, although with great reservations often expressed by those who do so. The major reason for using it is its relative ease of collection. More evidence is needed to assess whether having more detailed and accurate information on SES at the pupil level gives results significantly different from those arrived at by using the FSM indicator. Evidence in support of point ii) has already been presented above such that it would seem mistaken to conclude that there is no effect on pupils' and schools' results of poverty and other types of disadvantage.

The serious objections to including SES in the assessment of schools' performance therefore seem to relate more to questions of ethics and politics than of statistics. They can be summed up in the way the then Minister for Schools (Stephen Byers) responded in 1997 when tackled about the continuing out-performance by schools in leafy suburbs of those in inner cities: 'poverty is no excuse'. Two distinct issues need to be separated out of this single soundbite, one concerned with the assessment of schools' performance and the other with managing the motivation of young people. On the first issue, since there continues to be an empirical association between pupils' academic performance and their SES, it would seem intellectually dishonest not to take account of such information when attempting to assess the relative effectiveness of schools serving widely contrasting catchment areas. Such assessment is, of its nature, historical: it has little to say about what might be the case in some differently constituted future (as has been said, the correlation is empirical not *a priori*) and even less about what ought to be the case (how far public educational resources should now be targeted on schools in the poorest areas is a political question which the empirical data clearly poses). For a cogently argued position paper on this issue, the 1997 paper by Mortimore and Whitty would be hard to better (Mortimore and Whitty, 1997).

In fact, one could argue – as Saunders (1997) has done – that the *different* value added scores which a school may obtain depending on whether prior attainment only or prior attainment plus school context data is included can be used to suggest to school staff whether or not their expectations might have been influential in pupil outcomes. This suggests that the potential importance of using more than one model *in the same analysis* needs to be more fully explored. Taking a slightly different but related tack,

Hutchison (1993) discusses the difference between 'pupil progress' and 'school effectiveness' residuals. And of course this issue must be related back to the question posed in the previous section, about the multiple and conflicting purposes of undertaking value added analyses in the first place.

On the different question of managing pupils' motivation, it is easiest to argue from the statistical evidence itself. Although FSM is a significant variable correlating with performance at the school-level, most studies find that it has little or no explanatory, i.e. predictive, capacity at the individual pupil level. From a statistical point of view, what matters more than any other variable at this level is prior attainment. If teachers stick to what the data tells them, therefore, there should be no question of 'poverty being an excuse', in terms of withholding encouragement, support and challenge for any pupil. The problem, it seems to us, is not so much that some teachers misuse statistical, including SES, data but that they place much more store by subjective impressions. Sometimes they do so to the extent of (unconsciously) entertaining lower expectations of pupils from, for example, particular housing estates when the data – if used and understood – would tell a different story.

Jesson (1992, p. i) goes even further in putting health warnings on the analysis of individual pupil performance. Even prior attainment is not to be trusted as a predictor: 'individual pupils are not pre-ordained to score at any particular level at GCSE by virtue of assessments of their "ability" made at some earlier stage in their career.'

A key lesson to be learnt, then, is statistical analyses can be very powerful when used to assess the past performance of aggregates; but weak and possibly misleading if used to predict the future performance of individuals.

Finally, there is also a need to recap on what some might see as a fine distinction between 'value added analysis' and 'contextualisation'. TGAT conflated these two approaches, according to Buck (1993), and others have done so since. Contextualisation properly refers to the allowance made for SES factors in the computation of schools' results. Probably the best known example of this is to be found in the work of the National Consortium on Examination Results (NCER), which has for many years been providing GCSE results to LEAs (and thence to schools) on this basis. In this approach, the data is collected and analysed at school level, and regression analysis is the method used. Because of the demonstrable negative relationship between academic performance and low SES, the outcomes of this analysis were perceived, and welcomed, by LEAs and schools as more realistic

assessments of relative performance than reports based on raw results alone. But there are two main objections to equating this approach with an analysis of 'value added'. If we turn back to Jesson's and Gray's (1992/3) criteria, three of the five technical criteria remain unfulfilled, namely, the analysis of data at the individual pupil level, the inclusion of a prior attainment score and the use of multilevel modelling as the statistical technique. This last stipulation is made because, in trying to organise evidence on the performance of pupils for the purpose of measuring the relative effectiveness of schools, we need to deal with what is known as hierarchical data. Goldstein (1987) puts the argument like this:

... the existence of hierarchically organised data implies that we should take that hierarchy into account when we analyse data... the failure to account for hierarchies may lead us into trouble... If we were interested in the factors which influence pupil attributes, say their educational achievements, then among those factors we will generally wish to include the characteristics of the pupils themselves, their classes, and their schools (p. 3).

There are now very few serious scholars in the field of school effectiveness who would argue against the use of multilevel modelling. Those few who do, like the commentators reviewed in section 3.2, tend to do so on pragmatic grounds, in that multilevel modelling is held to be less transparent than regression analysis and needs to be done by trained professional statisticians. But the point is surely that the method of analysis needs to be consonant with the complexity of the data and the issues it is attempting to elucidate rather than tailored to the limitations of one particular audience or end-user. It then follows that there is an interpretative as well as an analytical task to be done by the scholars (and others) providing value added analyses. This has ethical as well as practical implications, which are discussed in Section 5.

A useful comparative critique of various value added models is to be found in Thomas and Mortimore (1996).

3.5 What Can Value Added Reveal about 'Effectiveness' and Educational Quality?

Value added analyses, as we have explored them earlier in this section, have provided information on the following key issues, and – perhaps even more importantly – have made us able to think about those issues in evidential rather than just common-sense terms:

- ♦ the size of the school effect, relative to other effects;

'Value added': a critical review

- ♦ factors beyond schools' control correlated with academic performance, in particular the difference made by the overall composition of a school, in academic and/or socio-economic terms;
- ♦ evidence of 'differential effectiveness' in the same school (e.g. positive results for some pupils but not others, related to their gender, ethnicity, prior attainment, SES);
- ♦ evidence of variable effectiveness (e.g. between departments in the same school);
- ♦ evidence of lack of stability in schools' results over time.

A clear and comprehensive overview of what the research into value added measures of school effectiveness has achieved is given by Thomas (1998).

Further questions which value added statistical approaches are now beginning to explore include:

- ♦ What can be said about social as distinct from academic effectiveness (as measured by attitudinal data and/or attendance)? Are the independent variables correlated with social effectiveness the same as for academic effectiveness? What evidence is there of stability/consistency/improvement in terms of attitudinal factors in individual schools?
- ♦ How far is academic effectiveness correlated with social/motivational effectiveness within the same institution?
- ♦ How might the correlates of *changes* in school effectiveness (e.g. towards improvement) be identified?
- ♦ Given that research suggests that the greatest variance is within-school rather than between-school (see, e.g., Hill and Rowe, 1996), how might the relative effectiveness of teaching groups/classes, etc. be explored? What *caveats* would need to be made on such research from a statistical point of view?
- ♦ Is it possible to investigate 'how the variation [in effectiveness by... ability group, by social class, by ethnicity and by subject department] is maximised or minimised within different types of "effective" or "ineffective" schools?' (Reynolds, 1995.)
- ♦ Conversely, institutional 'effectiveness' might *prima facie* be thought to be correlated with the level and type of LEA or other external support. How could this be explored using multilevel approaches?
- ♦ Going on from here, what other important factors – notably i. the effect of previous institutional impact on pupils' performance; ii. the relational or micropolitical aspects of schools; and iii. 'partnership' models of education which

utilise agencies to work in or alongside schools – should now be brought into the analysis of school effectiveness? How could this be done validly and manageably?

The concept of value added, together with more sophisticated ways of treating data, seem here to stay, although the purposes and functions of value added continue to evolve in interesting and politically responsive ways.

What further or deeper questions are raised by value added analyses – in other words, what value added analyses as currently understood leave out of account as well as what they reveal not only about schools but about social and political processes in education – are worth summarising as a conclusion to this section.

3.6 What Further Questions and Queries does Value Added Raise?

Whilst Fitz-Gibbon's 1996 definition of value added as 'the statistic of choice' may not commend itself to everyone, it is a clear indication that the value added approach is in the late 1990s seen as both less problematic – in terms of how acceptable the principles are to, say, policy-makers and school senior managers – and less problematised, i.e. less likely to be questioned even in the most sceptical of staff rooms.

As we have seen in this and the previous section, value added started its educational life as a contested educational concept, in which questions of value judgement were explicit: was it right, and feasible, to measure schools' effectiveness by some other means than absolute academic standards? The rightness was accepted without hesitation by school senior managers and LEA personnel who were experiencing the fall-out from 'raw' league tables of performance. Questions about feasibility, and thus about statistical method, then became paramount and the mid-90s saw a proliferation of discussion papers and articles about how to do it. The more the methodological sophistication was developed, however, the more complex an undertaking value added was revealed to be, and *pari passu* the less appropriate for making simple summary judgements about 'better' and 'worse' schools.

But the public purposes of value-added have continued to be multiple, espoused by policy-makers as well as by school senior managers; broadly speaking, they can be grouped as follows:

- ♦ the development of better and fairer ways in which to assess school effectiveness, overall and by institution: 'effectiveness' has therefore come to mean academic progress made as well as standards reached;

- ♦ the deployment by schools and LEAs of 'finer grids' with which to examine local and departmental differences, so as to establish action points for school improvement.

That these purposes do not sit comfortably with one another seems to be recognised at present only by academics in the field, and not universally by them. In any case, there are limits to any kind of quantitative data and its application, *qua* value added, to the field of school effectiveness and improvement which need to be recognised. These can be summarised as:

- ♦ There is not yet an agreement that the right measures are being used (e.g. for measuring prior attainment and/or SES) nor that the right models are being applied (Sammons argues that the existence of a 'primary effect' has not been taken into account in most secondary school analyses and that these may therefore be 'seriously mis-specified' (1996, p. 131). There is probably some way to go before scholars consider this research agenda to be completed.
- ♦ Moreover, the theoretical body of knowledge about the ways schools influence pupils' outcomes is not well developed: research is needed both into the generalisability of models of educational effectiveness and into the links between class/teacher and school effects (Sammons, 1996, p. 142) as Goldstein (1997c) says:

... little of the true potential of school effectiveness research has yet been realised. Even the minimum requirements for valid inference are demanding ones. They imply considerable expenditure of thought and resources as well as the long term follow-up of large numbers of pupils, schools and classrooms. They need to be replicated across educational systems, phases and types of schooling. It follows that those engaged in school effectiveness research need to convey this modesty of knowledge when they make claims to understanding (p. 394).

- ♦ Quantitative data represents only one aspect of schools' performance. Not all desirable outcomes are easily measurable. (Nor is all that can be measured equally important.)
- ♦ Quantitative data is only one instrument of evaluation: it reveals correlations, not causes, and it deals in aggregates, not individuals. It needs to be used in dialogue with qualitative evidence and professional judgement.
- ♦ Because it is norm-referenced, value added analysis does not set new levels of goals/targets which may be needed.
- ♦ What is a good or effective or successful school? The existence of value added analyses has not done away with the need for establishing a public consensus on this question; and probably the need to redefine 'effectiveness' will constantly

recur as changing labour and social markets continue to change our perceptions of what education is for.

Well, at this point – around the time of the introduction of a national value added system – ‘effectiveness’, ‘improvement’ and ‘raising standards’ seem to be almost solely concerned in practice with test and examination results: value added has done nothing to undermine what one might call ‘the tyranny of the measurable’, and has even perhaps slowed down developments in more qualitative research on effectiveness (as undertaken by e.g. MacBeath *et al.*, 1996). Furthermore, the problems which most seem to concern policy-makers are those of methodological procedure. Even for the relatively restricted type of outcome measure which value added analyses currently provide, not many commentators have explicitly raised the issue of ethical considerations, including whether a protocol for the use of value added analyses is needed (but see Saunders, 1997); whereas this may turn out to be the area where most effort is needed.

At an even more basic level of criticism, would people like school governors and parents – as deeply concerned about the quality of education as any researcher but unlikely to have the time or inclination to read the scholarly arguments and their implications – be right to entertain a suspicion that researchers have been engaged in an exercise no more meaningful or helpful than counting angels on a pinhead? Has the whole value added agenda been obfuscating rather than illuminating, as John Patten (then Secretary of State) once intimated? How far are the complexities which have been introduced to the measurement of educational quality justified in what they have contributed to the possibility of raising standards? Whose professional and career interests have been well served by the introduction of value added? There are no easy answers to these questions, but those of us engaged in the value added business need to keep them on our list.

Notwithstanding these strictures, this section has provided sufficient evidence, we think, that value added has contributed enormously to understandings of the measurable aspects of school effectiveness; it has also served to show, perhaps, that, in its educational context at least, it ought to be seen as one of those ‘metaphors and analogies, which [scientific thought] *must both use and suspect.*’ (Byatt, 1997, p. 251; our emphasis.) The potential must somehow be preserved for ‘generat[ing] possibilities for things to happen that are closed off by the epistemologies of certainty.’ (Stronach and MacLure, 1997, p. 5).

'Value added': a critical review

It is in the nature of governments to deal in epistemologies of certainty, however, and it is to the two major areas of current policy concern mentioned above – measuring relative progress and assisting with school improvement – that the next section is devoted.

4. A SHARPER STICK TO BEAT THEM WITH? REVIEW OF THE POLICY LITERATURE

This section briefly discusses the key policy literature on educational value added during the 1990s. This is done primarily in order to understand how successive administrations applied the notion of 'value added' to evidence about standards, progress and effectiveness; and to attempt some evaluation of how far the national value added system, in which the debate has in one sense culminated, is likely to improve the quality of that evidence. The section concludes by suggesting that, whilst the introduction of a national value added system has much to recommend it, there may still be crucial work for policy-makers to encourage and support if value added indicators are to have a beneficent impact on school processes.

4.1 Changes in Official Policy on Value Added

It is well known that the Conservative administration took some time to come to terms with the idea that it might be helpful to take account of background factors in assessing how effective schools were. This is the then Parliamentary Under-Secretary (Michael Fallon) (Fallon, 1991) responding, in a radio broadcast in 1991, to issues raised by the analysis of 1990 GCSE examination results for the Association of Metropolitan Authorities (Nuttall *et al.*, 1992):

[Interviewer: People would say that gender and social class and one-parent families and so forth, these are very important backgrounds and should be considered when school tables are published.]

Well I'm sure parents will consider them but what we are not going to allow all the educational experts to do is to dress up these tables with a whole lot of sociology so that parents can't actually make valid comparisons from one school to another... All these issues will be debated I'm quite sure as the Bill progresses in Parliament, but we don't underestimate parents. We think the important thing is that parents should be the interpreters of this information and the last thing we are going to do is let the education industry in on the act to start glossing up all this clean data with all kinds of educational sociology which we simply don't think is valid.

The Secretary of State in 1993-94 (John Patten) remained adamant that he wanted no truck with what he called 'cooked' as distinct from 'raw' data on schools' results. He was replaced in July 1994 by Gillian Shephard. A signal that the government might

now want to take notice of 'cooked' data was contained in a press notice launching the government's new initiatives on teacher training; one of the innovations was to be the placement of pupil teachers in urban schools which were doing well despite appearances. Over the next few months, there were accumulating signs that the Secretary of State was attracted to what value added analyses seemed to offer: a fairer way of comparing schools' performance.

4.2 The Value Added National Project

In fact, such a change of policy had already been well prepared for by Sir Ron Dearing's recommendation, contained in the interim report on the National Curriculum and its Assessment (Dearing, 1993), that work be commissioned into value added performance indicators. This work was undertaken during 1994 by a working party convened by the then School Curriculum and Assessment Authority (SCAA) and chaired by John Marks. The SCAA working party produced its report and recommendations in October 1994 (SCAA, 1994). Sir Ron Dearing, in commending the report to the Secretary of State, was explicitly ambitious about its implications: '... it should be possible for us to make proposals for the development of a national system.'

The 25 detailed recommendations include the principal recommendation that three simple models for value added analyses should be developed, 'the essential feature of these models [being] that they provide a basis for making like-with-like comparisons'. The summary concludes (p. 7) that:

the use of value-added indicators in developing policies for school improvement should be encouraged on a national basis by developing a culture of openness in connection with performance indicators and methods for school improvement, by using the inspection process to encourage schools to pursue further the use of quantitative performance indicators, and by making available to all schools, via the DFE, OFSTED and SCAA, information concerning methods of analysing performance indicators which have been found to be helpful.

One of the tasks of the Secretary of State during her first six months in office was to welcome the report; in so doing she demonstrated the 180 degree turn-around in Department policy (Shephard, 1994):

I am firmly committed to the development of robust national measures of the value-added by schools to children's education... It

is imperative to ensure that any measures of value-added are straightforward and are intelligible to parents... I commission SCAA to take this work forward...

An important feature of her response was to reiterate Dearing's acknowledgement of the useful work – much of which, as we saw, had been funded by local education authorities and other influential bodies – already being done outside national government circles:

I also welcome the impetus which the report gives to the use of value-added measures by schools and LEAs as indicators of their own effectiveness. Some useful work of this kind is, I know, already going on... The wider use of such indicators will be facilitated if SCAA, OFSTED and the Department are as open as possible about the development of work in this area... I shall expect the Authority for its part to have regard to the approaches being tried out locally in taking forward its work. (*ibid.*)

It then took little time for a national value added project, supported and funded by SCAA, to be initiated and a contract was set out in early 1995 to run from March of that year to December 1996. The objective of the project was *'to advise the Secretary of State on the development of a national system of value added reporting for schools based on prior attainment, which will be statistically valid and readily understood.'* (Fitz-Gibbon and Tymms, 1995, p. 31.) The focus was to be on results at Key Stages 2 and 4, that is, broadly speaking, the outcomes from primary and from secondary schooling respectively. The contract was awarded to the Curriculum, Evaluation and Management Centre, then housed at Newcastle University and later relocated to Durham University. A series of technical and overview reports was produced, and the final report appeared in 1997. (See chart overleaf for a timetable of developments in the national value added system.)

The change of administration from Conservative to New Labour in May 1997 made no difference to the aims and momentum of the project; the policy transition was seamless. As the Qualifications and Curriculum Authority put it, *'the White Paper Excellence in Schools stressed the need to provide parents and others with information about the progress pupils make relative to other pupils. These are usually referred to as "value added" measures.'* (QCA, 1998, p. 1). Currently, the Standards and Effectiveness Unit (SEU) at the Department for Education and Employment and the Office for Standards in Education, as well as the Qualifications and Curriculum Authority, are taking a lead in the production of national value added and other contextual data on school and local education authority performance.

In part, this is intended to aid the setting of 'challenging but realistic targets' for schools and LEAs (see DfEE Circular 11/98). One might even argue that it is this statutory requirement to set targets which will bring value added analyses closer to the everyday life of schools and teachers.

EVOLUTION OF THE NATIONAL VALUE ADDED SYSTEM

- 1993 Sir Ron Dearing's interim report on the National Curriculum and its Assessment recommends that work be commissioned into value added performance indicators for schools
- 1994 SCAA working party, chaired by John Marks, reports to Secretary of State with 25 recommendations for taking value added measurements forward based on public examination and National Curriculum assessments
- 1995 Contract for a feasibility study of the procedures necessary to set up a national Value Added System awarded to the Curriculum Evaluation and Management Centre, University of Durham: statistical studies on available datasets to be completed by September 1995 and two new pilot projects – KS1-2 and KS3-4 – to run from September 1995 to December 1996
- 1997 Series of project reports published, advising that *'it is possible and desirable to set up a national system to provide schools with value added indicators of their performance'* and providing detailed findings from the project
- 1997-8 Consultations take place with schools and other interested bodies on the best ways of calculating and presenting value added indicators for schools and value added measures in school performance tables
- 1998 The first national value added analyses, covering KS1-2, KS2-3 and KS3-4, published in autumn term in the form of i) lines representing the achievement of pupils at the median and quartiles of the distribution of outcomes for each possible input score, and ii) of chances graphs showing the percentages of pupils achieving different outcomes from the same starting points
- Secondary school performance tables contain 'progress measures' (as distinct from true value added)
- (NOTE: This strategy changed somewhat because of operational difficulties encountered)*
- 2000 Secondary school performance tables to include value added measures
- 2002 Primary school performance tables to include value added measures

As for Shephard's wish that SCAA should take due note of developments already happening, it is noteworthy that SCAA's successor, the Qualifications and Curriculum Authority, set up a series of meetings with value added service providers – of which the NFER was one – at which updates on the progress of the national value added

system were given, opinions taken from those already involved in value added and problems frankly aired.

In the meantime, various other reports on value added prepared for or on behalf of a policy interest appeared, including the following:

- ♦ Assessing School Effectiveness: (Summary of a research study on) Developing Measures to put School Performance in Context (Sammons *et al.*, 1994, for OFSTED)
- ♦ Value Added in Education: A Briefing Paper from the Department for Education (GB. DFE, 1995a)
- ♦ GCSE to GCE A/AS Value Added: Briefing for Schools and Colleges (GB. DFE, 1995b)
- ♦ Value Added and School Improvement (Society of Education Officers, 1995)
- ♦ Adding Value and Improving Performance (Elkin, 1995, for the City Technology Colleges Trust)
- ♦ Value Added Measures of School GCSE Performance (Jesson, 1996, for the DfEE).

Undoubtedly these publications helped to establish and promote the debate amongst a wider public concerned with education than would normally have access to academic journals and technical papers. The reports written by Sammons *et al.*, Jesson and the Society of Education Officers (SEO) are particularly clear and helpful, in taking on some difficult technical and statistical issues but at the same time always relating these to the educational purpose(s) of attempting to measure value added. As well as discussions of the relative merits of regression and hierarchical (i.e. multilevel) models, and of prior attainment and socio-economic data as 'predictors' of performance, for example (which are also aired by Sammons *et al.*), the SEO report contains sections concerned with what might be called value added in practice: for example, 'making sense of the outcomes', 'the link with school improvement', 'confidentiality versus sharing the outcomes', 'the role of the education authority' and 'the case for local value added initiatives'. It concludes with the observation that:

For most education officers and schools, the analysis is not an end in itself. Much work remains to be done to develop the insight it provides. Collaboration of this kind [a partnership between LEAs and schools, individually and collectively] probably offers the best prospect of school improvement currently available in the service of all children and their parents. (Society of Education Officers, 1995, pp. 38-9.)

Of course, one might say, this is a predictable position for the report's authors to take, but the general proposition remains true, as is argued in Section 5 below. Some of the other material, by contrast, served to perpetuate unhelpful ambiguities and misunderstandings. We would want to take issue with several of the definitions given of value added, for example, as being unenlightening or inaccurate or both; but since this review began with the argument that the term was dependent on 'functional ambiguities', it would be tedious to labour the point with further examples. More specifically, in her section discussing factors other than prior attainment which may affect performance, Elkin (1995, pp. 30-1) apparently does not distinguish between those over which schools have little or no control (such as percentage of students with a statement of special educational need) and those which are directly the result of school policy and management (such as amount of homework). This suggests that she has not understood the 'eliminative' nature of the conceptual model used by all the quantitative approaches she comments on. Furthermore, in discussing socio-economic factors, Elkin claims that:

the idea of "making allowances" is a seductive one. Yet... it is only too easy to say that because "our" school has large numbers of students from low income families, or whatever, that the pupils cannot be expected to achieve much (p. 32).

She thus makes the common mistake (and one which perhaps has an ideological component) of confusing 'making allowances for' with 'allowing or controlling for (certain specified and empirically established factors)'. (Both of these general issues were discussed in Section 3 above.)

A comparable confusion is to be found in the DFE Briefing Paper on value added (GB. DFE, 1995a), the main purpose of which was to give some background, from the Department's perspective, to the national value added project and to describe briefly the work in hand. The paper points out (p. 2) the 'major technical difficulties in assessing the scale of any such effect [of adverse socio-economic factors]' and goes on to claim that 'adjustments to value added measures on these grounds which were not sufficiently rigorous could justify poor performance or legitimise low expectations.' Again, this appears to have some political rhetoric at the back of it. Adjustments for socio-economic factors should indeed be based on rigorous and empirically replicable data; but problems of non-rigorous interpretation may still remain. Legitimising poor performance is a failure of understanding, interpretation and, ultimately, imagination rather than a result of poor data. One way round this issue is to present school-level residuals at three stages of adjustment: unadjusted, adjusted for prior attainment and

adjusted for prior attainment together with socio-economic context, in order that insights can be obtained from examining the *differences* between a school's results at successive stages of analysis. This was also discussed in Section 3 above.

The DFE paper also, rightly, discusses how to arrive at the right balance between paying attention to 'technical niceties' and enabling 'intuitively attractive' summary measures to be derived. However, the ensuing recommendation (p. 8) that '*summary measures, e.g. for use in comparative performance tables, need to be few in number, and must be consistent and fair between institutions... [whereas] value added measures for use by school and college management will typically be more disaggregated*' seems either to duck the issue or else to imply that summary measures of value added which are to be used for publicly comparing schools can afford to forgo (some of – which?) the technical niceties. Whereas, on the basis of evidence presented in Section 3 above, many academics think that the more 'high stakes' the value added analyses are – and public comparisons of institutional performance are pretty high stakes – the more crucial it is that the technical niceties are observed.

Naturally, much of the debate has been concerned with how to strike the right balance between what is ideal, analytically speaking, and what is feasible. The remarks of the SCAA working party report (SCAA, 1994) are worth quoting, even though they are hidden away in Appendix 9:

... decisions about which variables to include inevitably depend on views about what is important in the real-world situation we are attempting to model... Such decisions are therefore, almost inevitably, open to question and debate... It therefore needs to be recognised that there is no single correct method of analysing a complex social situation... It also needs to be recognised that all corrections and adjustments to primary data are approximate and that there will always be arguments about their validity (p. 85).

The paragraphs in the DFE document could be regarded as no more nor less than a gloss on what SCAA had argued, but something has been lost in the translation, as they say.

Interestingly, the SCAA report goes on immediately to say:

There is thus a strong case for using simple methods of analysis alongside complex ones and for presenting analyses in ways which retain as much contact as possible with the primary data... (*ibid.*)

- which could be read as a belated argument in support of raw data.

But both the DfEE and QCA recognise that value added analyses may be put to a variety of different uses; a basic distinction can be drawn between accountability requirements and the contribution of performance data to school improvement. Both uses were apparently supported by a majority of respondents to the two QCA consultation exercises, although in the final project report (Fitz-Gibbon, 1997, pp. 25-9), the sample population of headteachers were found to be overwhelmingly in favour of value added data for internal use but very negative towards the idea of publication of such data. Are there methodological or technical implications of trying to put the data to different uses? The DFE's position on has been outlined on the previous page; according to QCA (1998, p. 1), the distinction means only that *'the information and measures are both arrived at by the same method, but their purpose, although complementary, is very different.'*

Well, in one sense, the important thing is that schools and LEAs are no longer operating in a policy void as far as value added is concerned: the salient facts to bear in mind are, first, that one of the main objections to performance tables as originally conceived, that they gave no indication of relative progress in differing pupil populations, has been met at least in principle. Secondly, all schools, whether or not they choose to supplement national data with other sources of value added analysis, will have access to quantitative performance analyses which could in theory be a substantial contribution to management development, institutional review and target-setting.

Why, then, qualify these desirable and desiderated moves with phrases like 'in theory' and 'in principle'?

4.3 Reconciling Academic Criteria with Policy Requirements

In constructing a brief for the national value added project, SCAA specified that a national system would need to be:

- ♦ readily understandable;
- ♦ statistically valid;
- ♦ not an undue burden on schools;
- ♦ cost-effective.

(Fitz-Gibbon and Tymms, 1995, p. 2)

More specifically, value added measures included in published tables of performance need to be:

- ♦ applied consistently across all schools;
- ♦ based on robust statistical methods;
- ♦ sufficiently straightforward for parents and others to understand, and for schools to check.

(QCA, 1998, p. 1)

As policy requirements, these are perfectly legitimate, one could say minimal. To what extent are they capable of being fulfilled? One problem may be, given the evidence marshalled in Section 3 from the academic literature, that it will not be possible to have value added measures which are both statistically valid and readily understandable (*pace* Fitz-Gibbon, 1997, p. 14: even with her 'very simple and accessible models', one wonders with what sort of sample of ordinary people the graphs and tables were trialled), and also give better insights into schools' performance than raw scores.

In Section 3 above, we noted a general consensus amongst writers in the school effectiveness field that measuring value added with accuracy depended on the following criteria:

- ♦ using individual pupil data;
- ♦ taking account of prior attainment;
- ♦ taking account of other pupil background factors (gender, ethnicity, SES);
- ♦ including outcome measures which reflect the whole range of achievement;
- ♦ analysing the data with a multilevel model.

Less commonly stated criteria which may still be taken as having cogency include:

- ♦ the need to take account of school context (as well as individual pupil level) factors;
- ♦ the need to use finely differentiated input and outcome measures, not merely broad bands or levels of achievement;
- ♦ the need to avoid 'high stakes' measures as input variables; a good example of this would be using the publicly available outcomes from a previous stage of education such as national curriculum test scores as the measure of prior attainment or 'input' for the next stage.

These criteria are concerned with questions of accuracy and robustness – although problems of cost effectiveness and comprehensibility for a lay audience have not been

entirely ignored in academic literature. (The extent to which widespread comprehensibility should continue to be a preoccupation may be a moot point, although it has superficial appeal. How far parents use performance tables as the determining criterion in choosing a school has yet to be definitively established, but seems to be dependent on both parental class and local circumstance.) A reading of the QCA reports on the national value added project shows that they contain as thorough a presentation of methodological and technical problems as could reasonably be expected of a principal agency of a government committed to introducing a national system of value added as soon as possible. In the final report, there is discussion of, *inter alia*:

- ♦ the need to base calculations on individual pupil data;
- ♦ the need to have more finely differentiated performance data than that provided by assessments according to national curriculum levels;
- ♦ the need to use three year rolling averages;
- ♦ whether and how to account for factors like student mobility and attendance;
- ♦ the relevance of socio-economic factors;
- ♦ relative merits of different statistical models;
- ♦ presentation of results;
- ♦ questions of quality control and data-checking by schools;
- ♦ training and support for school staff in understanding value added data;
- ♦ the potential contribution of value added data to school improvement;
- ♦ the possibly harmful effects of a national system.

This last issue is discussed in the report's penultimate and thought-provoking chapter (pp. 87-93), about which not a great deal has been said (aside from Kellner, 1997). Fitz-Gibbon quotes Smith (1995) in identifying unintended and presumably unwelcome behavioural consequences of publishing performance data. '*Eight problems associated with non-effective or counter-productive systems*' are listed, as follows:

- ♦ tunnel vision
- ♦ sub-optimisation
- ♦ myopia
- ♦ measure fixation
- ♦ misinterpretation
- ♦ misrepresentation
- ♦ gaming

- ♦ ossification.

Fitz-Gibbon comments on this list:

With the exception of ossification every one of these possibilities was commented on by headteachers in open-ended items in the questionnaires. These are not theoretical but actual, already-perceived problems.

As we saw in Section 3, a reputable school of thought is now arguing that value added measurements – far from making the public reporting of comparative performance more accurate and informative – underline the fact that performance tables for use in the public domain are impossible to construct with validity *and* simplicity. Fitz-Gibbon's critique strongly suggests that the usefulness of published value added measurements 'in attempts to improve efficiency and equity' may also be compromised. This is the nub of a public relations problem for New Labour: a commitment to continuing the Conservative policy of publishing performance data has led inexorably to their trying – like the Conservatives – to make that data fairer and fuller, and value added measures offer the most plausible approach. The alternative, of ceasing to produce performance tables, seems unthinkable at this point. An independent evaluation of the effects over time on schools and pupils of the national system is surely called for. In the mean time, the evidence in Section 3 seems to suggest that value added data in the public domain is most useful as a screening device for identifying those relatively few schools which are 'outliers' in statistical terms, rather than for providing definitive assessments of all schools.

What of the use of value added for institutional review and school improvement? Well, some similar strictures must apply; that is to say, if it is true that value added measures:

- ♦ are only as good as the data they are based on;
- ♦ deal in correlations, not causes;
- ♦ contain an irreducible degree of statistical uncertainty;
- ♦ are based on a normative and retrospective model which may not provide any helpful information about levels of performance in future;

then it needs to be said that value added measures function as only one instrument of evaluation.

Even so, the information on relative performance provided by value added analyses is more robust than that characteristically given by other methods in the past. A commonly-heard phrase in schools these days is: *There's no hiding place now*. Most importantly, explanations of poor performance which rely on assertions or assumptions about the inadequacy of the pupils can more readily be exposed. Value added measures make it a great deal easier to identify the root causes, as well as the underlying trends and patterns, of under-achievement.

So it seems reasonable to claim that, given the right culture, value added can help to pose better and more focused questions about the way a school or LEA has performed with its pupils and to stimulate more informed discussion amongst school staff about the way they organise and deliver their teaching. How far 'the right culture' exists, what needs to be done to nurture it and whether the introduction of the national value added system will help its growth are key questions to which national policy makers will need to seek answers. The following section discusses ways in which some aspects of these questions can be explored.

5. GETTING BEYOND THE NUMBERS GAME: ETHICAL AND PRACTICAL CONSIDERATIONS IN USING VALUE ADDED

5.1 Introduction

The implication of the discussion in previous sections is that, whilst it is true that value added measures of performance have been well developed in methodological and analytical terms, the other end of the process, so to speak – of support for schools in managing and using such data – is still relatively under-researched. The material presented above has shown just how challenging the technical and conceptual issues are. Indeed, it is arguable that support for schools is more than ever necessary, as increasingly sophisticated data is generated by the DfEE, the QCA, LEAs and schools themselves. First-hand experience suggests that, whilst some school senior managers are highly numerate and well aware of what different kinds of so-called value added analyses can tell them (and, just as importantly, what they cannot reveal), it is likely that there are many in the profession for whom guidance in interpreting and using such data is crucial. This is arguably even more true in respect of value added data used for school review and improvement than data used for external accountability.

That this is an area which colleagues in the value added field are keen to explore and develop was demonstrated by the interest shown in a seminar hosted by the National Foundation for Educational Research (NFER) in early 1998. Key figures in value added and school effectiveness research attended a meeting for the purpose of discussing possible protocols for interpreting and using value added: as a result of this meeting, a joint article was written (Saunders and Thomas, 1998). Ultimately, however, the challenge to those providing value added analyses must be: Is the effort worthwhile? Do value-added approaches help to improve education in practice? This is obviously a crucial question, to which there are as yet no clear-cut answers. We first need to know, on the basis of empirical evidence, how value added data is actually used in schools.

5.2 Issues in Schools' Use of Value Added

In fact, the evidence on the ways in which schools use value added data is so far quite modest, partly of course because it is only very recently that the majority of schools had access to anything that could remotely be called value added. Few research studies have focused specifically on this aspect of school practice, moreover: it is only

a slight overstatement to say that the broad tendency in school improvement research has in the past been to assume that value added data is of limited practical interest outwith the debate on relative school effectiveness. (Several books published in the last couple of years have taken a different view of this issue, however, in attempts to integrate the school effectiveness and school improvement traditions. See, for example, Gray *et al.*, 1996; Reynolds *et al.*, 1996; White and Barber, 1997; Gray *et al.*, forthcoming)

A decade ago, the very small amount of evidence gathered was pretty negative: 'The picture that emerged [from a study of schools receiving ALIS-type data] was of a very peripheral level of awareness, a tendency not to attend meetings, to read reports sparingly and to take no action on reports once read.' (Williamson and Fitz-Gibbon, 1990.) Some years later, another small-scale study (Harris *et al.*, 1997) used NFER QUASE data to link quantitative and qualitative evidence at the departmental level. This study postulated a range of common characteristics in departments identified as 'adding value', which can be summarised as follows:

- ♦ Good organisation/management, from schemes of work to planning and utilisation of resources;
- ♦ Systematic scrutiny of exam and test results;
- ♦ Sense of vision (especially of subject discipline) conveyed by department head;
- ♦ Change and innovation accepted, if in accordance with departmental vision;
- ♦ Effective internal communication within department;
- ♦ Delegation of tasks to colleagues (based on sense of professional trust);
- ♦ Active involvement of pupils in review of and reflection upon their own work;
- ♦ Consistent approach to assessment which motivates, reinforces and builds confidence;
- ♦ Clear structure for lessons which also integrates feedback to/from students;
- ♦ Consistent homework policy;
- ♦ Use of reward, in preference to punishment, to modify behaviour;
- ♦ Normal range of experience, capability, motivation, etc., amongst departmental staff, but within a context of low staff turnover.

It was intended that the findings should be used as an aid to self-evaluation and internal improvement within the cluster of schools which had participated in the study. However, the follow-up study conducted by Wikeley (1998) revealed problems in disseminating and sharing so-called 'value added', which can be expressed as follows:

'Value added': a critical review

- ◆ staff felt the dissemination of the 'value added' findings had been imposed from above by their senior management team/an external agency;
- ◆ staff in departments which had 'added value' felt that being identified in this way was divisive;
- ◆ some claimed that the research 'told us nothing new';
- ◆ senior management's perceptions about effective departments seemed to differ from other staff's;
- ◆ there was a tendency to distrust external quantitative data as opposed to instinctive judgement; people tended to be much more critical of the basis for the statistical analysis than they were of the basis for so-called professional judgement;
- ◆ there was, on the other hand, a lack of other robust kinds of data to confirm or refute the effectiveness findings;
- ◆ there was a lack of follow-up to the work *via* self-evaluation in departments; staff tended to focus on the need for further resources instead.

But actually these findings are probably no longer a reflection of the situation in most schools. There is no doubt that the tightening-up of the school improvement agenda by the DfEE's Standards and Effectiveness Unit has focused attention on the respective roles and responsibilities of LEA officers and advisers, school managers, governors and teachers in a new and unprecedented way. Taking a 'lies, damned lies and statistics – I trust my gut instincts' view of performance data is no longer an option for any of them.

The most recent piece of evidence we have encountered⁴ is a written case study of a secondary school which has used QUASE data (West and Moore, 1998). The authors claim several benefits arising from the school's use of value added data and the developments to which this has led in respect of monitoring pupils' performance and progress; these can be summarised as follows:

- ◆ Staff attitudes have changed from a 'minimalist understanding' and even a feeling of unease to a situation where 'measurement data is seen more as a tool to aid increased pupil achievement.'
- ◆ 'There is more openness surrounding the audience for the data and far more discussion about the issues that the data highlights. Many staff now use data to inform their curriculum planning at both departmental and classroom level.' Specific developments have been introduced to enhance the quality of learning.

⁴ In autumn 1997, the Qualifications and Curriculum Authority commissioned a study of the use of value added data for school improvement; at the time of writing, no findings from the study have been published.

- Pupils have benefited from a growing self-awareness about their progress (or lack of it).
- 'The most valuable outcome, paradoxically, probably belongs to that aspect of school effectiveness that remains elusive, defying quantification. It is to be found in the increased levels of awareness of the teachers, the day-to-day adjustments, the small experiments with new approaches, the time devoted to reflection or just thinking about the relationships between teaching and learning.'

It has to be said, however, that aspects of this paper are written from an implicitly management perspective:

Some curriculum areas/departments are now identified as better performing than others and this is known throughout the school. This could be interpreted as creating a difficulty, but it is rather seen as an opportunity for differentiated support to be provided for those departments/areas that need it. This enables the effective targeting of resources for improvement.

This sounds like a management development aim rather than a description of institutional reality. It may sometimes be important to get beneath the rhetoric, not in order to put a negative slant on managers' intentions nor to suggest that progress is not being made, but in order to find out how best to understand and address the difficulties which may arise. As Brown *et al.* (1995) argue, the notion of a top-down, management-oriented approach to school improvement based on school effectiveness data may be 'simple-minded' if it fails to address the question of how the findings are, in real-life, to be integrated into the thinking and practice of those who are seen as needing to improve. Furthermore, because such findings are in any case based on policy-makers' managers' and/or researchers' own constructs, they may be insufficiently valid and/or replicable.

5.3 The Need for Further Research

A prerequisite for further development is therefore a description of how schools – or rather the various staff within them – are already using value added data, for what purposes, and with what outcomes, both intended and unintended, for management, staff and pupils. At the time of writing, the NFER is funding a small-scale but indicative empirical exploration of these issues. Given the agenda to which most schools are now working, a particular focus of the study is how far and under what circumstances value added analyses have a role to play in raising pupils' attainment in different institutional contexts.

An understanding of organisational theory and the 'micro-political' processes at work in schools must inform this agenda. Recent field-based projects commissioned by DfEE and by QCA on schools' use of data for raising standards have been premised on untested assumptions about, for example, the transparency of data and the internal organisation and ethos of schools. As was noted above, much has recently been written by leading academics about the failure of school effectiveness studies to relate directly to the ways in which schools can, in practice, change and improve – and about the need to undertake better-specified research which will close the gap (see, e.g., Reynolds *et al.* 1993; Brown *et al.*, 1995; Reynolds *et al.*, 1996; Gray *et al.*, 1996; White and Barber, 1997; Lauder *et al.*, 1998; Mortimore, 1998).

In the context of value added, it is necessary to gain a better idea of:

- ♦ what the common needs and gaps in technical knowledge are, especially for senior and middle managers: *value added as an innovation*;
- ♦ the uses to which value added data is actually being put in schools: *value added as a lever for change*;
- ♦ the ethical and managerial issues this raises: *value added as a management instrument*;
- ♦ the kinds of school culture or ethos most likely to make use of value added for improvement: *value added as a 'climate indicator'*.

Furthermore, it is difficult in the government-sponsored projects to make room for the exploration of ethical issues, which we know (albeit from anecdotal evidence) preoccupy staff at different levels within school hierarchies: we are aware that value added data is being used as a lever for change in ethically dubious ways. All of this needs to be explored, primarily through qualitative fieldwork mediated by rigorous conceptual frameworks, for its impact on the raising attainment agenda.

5.4 Fruitful Qualitative Research Issues

A number of research issues suggest themselves for such fieldwork, as follows:

- ♦ how do school senior managers understand and respond to value added data? How do they integrate it with review and development planning processes? With whom do they share value added information and for what purpose? Can a range of functions and uses of value added data as a management tool be identified? What are the main constraints and barriers to understanding and use?

'Value added': a critical review

- ♦ how do middle managers understand and respond to value added data? How do they integrate it with their review and target-setting procedures? What are the main constraints and barriers to their understanding and use?
- ♦ how do teachers understand and respond to value added data? How do they integrate it with their diagnostic, assessment and curriculum planning procedures? What are the main constraints and barriers to understanding and use at this level?
- ♦ are staff in schools with poor value added results surprised to learn this, or does it confirm what they knew? If the latter, what reasons – managerial, pedagogical, curricular, resource-related – would they advance for the results?
- ♦ do poor value added results on one or some key performance indicators mean that any given school has lower-than-expected results on all performance indicators? Have the staff identified precisely where their strengths and weaknesses lay? Are there any pupil groups or subject areas in these schools whose performance has been satisfactory or even higher-than-expected?
- ♦ turning to schools with good value added results, again, does this confirm what staff feel they knew? What reasons would they advance for the results? Are there any pupil groups or subject areas whose performance has not been so good, in value-added terms? What evidence, if any, is there of consistent improvement over time?
- ♦ how well are school staff able to use performance data interactively with their own professional judgements about effective teaching and learning, at departmental/year group, at classroom and at individual pupil levels? How do school managers and teachers use their results to set realistic targets for individuals and for groups?
- ♦ how far do staff at different levels consider value added data helps to raise pupils' attainment? What evidence can they demonstrate for this?
- ♦ how well are the new expectations and new uncertainties which value added approaches bring with them being managed?
- ♦ what are the main or common unintended consequences – including ethical and professional ones – of introducing and using value added analyses?

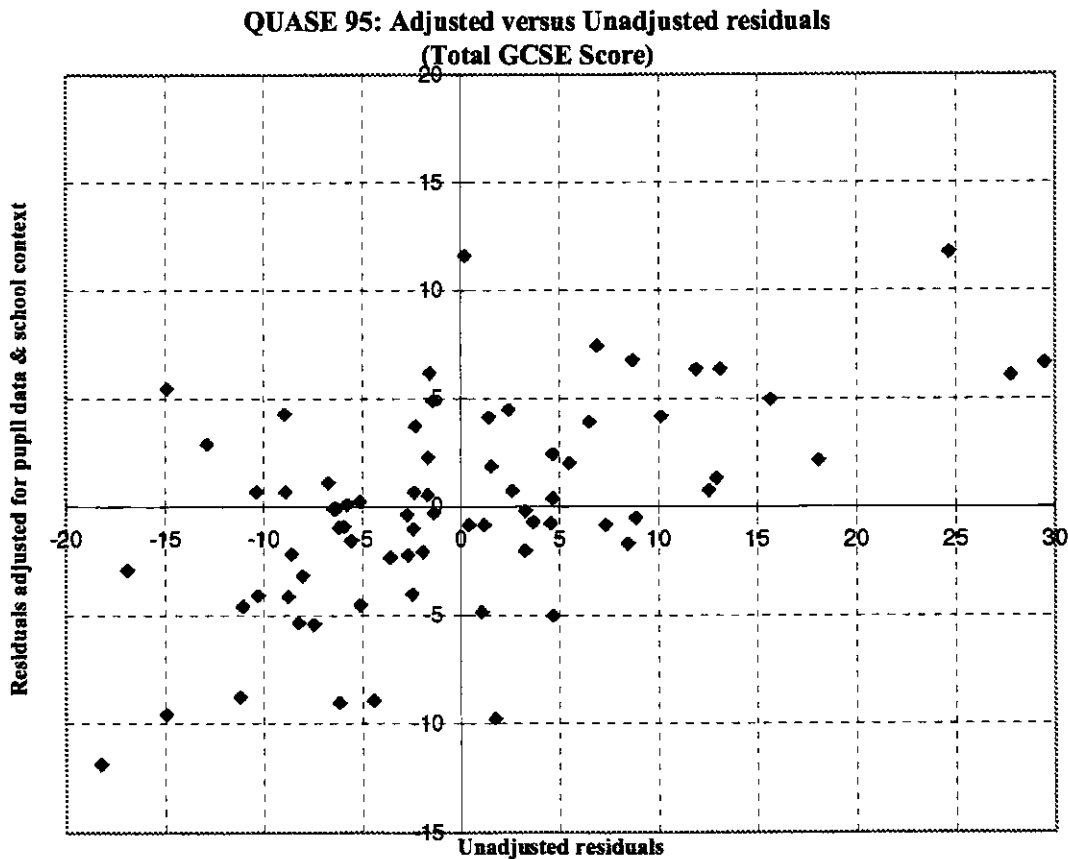
Factors to be taken into account in choosing case study schools for fieldwork would ideally include:

- ♦ the desirability of conducting case study work with low-, medium- and high-performing schools as reflected in their raw scores, on the supposition that their managerial and pedagogical priorities will be different;
- ♦ the further insights which might be gained by including schools with positive and with negative value added scores, irrespective of their raw scores. The quadrant diagram below (Figure 5.1), based on NFER's QUASE data on secondary schools, illustrates this principle. Schools in the upper right-hand quadrant are those which performed well on raw scores and showed better than expected performance ('value added') even when the analysis took into account a whole range of factors which might have affected their pupils' performance. Conversely, schools in the

lower left-hand quadrant are those which showed below average performance on both raw and adjusted scores, that is, even after all available adjustments had been made for pupils' prior attainment and socio-economic context;

- ♦ the potential relevance of the LEA as a promoter or otherwise of collective self-review using performance data;
- ♦ the desirability of considering different school 'cultures', in order to explore the explanatory potential of, for example, traditional versus collegial types of school and/or of instrumental/control versus expressive/cohesion emphasis: see Hargreaves (1995) for a discussion of how these notions might help to clarify hypotheses about the relationship between school effectiveness and improvement and school culture. See also Lauder *et al.* (1998) for the argument that effective schools are those which evolve structures and climates precisely in order to be different things to different pupils.

Figure 5.1 Quadrant diagram illustrating different ways of grouping schools in terms of their 'value added'



5.5 Using Value Added for School Improvement?

The current state of knowledge suggests that for value added analyses to serve the cause of school improvement, the following points need to be borne in mind:

- **Data rarely speaks for itself:** even the most rigorous quantitative analysis needs to be treated with caution, since it depends on a series of prior decisions – not necessarily obvious to the recipients – about *what* has been measured and *how* it has been measured. There are always important aspects of school performance that have not yet been measured, and often other ways, better or worse, of measuring things. One item that always needs to be made plain, for example, is whether the measured differences between schools (or departments) are significant, i.e. not just a function of the normal random distribution of results. So the full meaning of the data must surely emerge from the *interaction* between 'the numbers on the page' and what school staff bring from their professional judgement of their pupils, the school's context, etcetera.
- **'School improvement' is not self-evident:** despite the years of painstaking research into school effectiveness and improvement, there is no formula by which we can link identified inputs at classroom, year group or school level with desirable outcomes, like an increase in test scores. The notion that one can transfer practices in effective schools to ineffective schools has been shown to be particularly problematic. This is hardly surprising: schools are complex institutions and 'what works' is likely to be a combination of the transferable and the unique, arising from professionalism *and* personality; pedagogy *and* pastorality, mission *and* ethos; the rational *and* non-rational. A major hypothesis of any guidance should be that different kinds of schools need different kinds of strategy – schools are complex organisations with different backgrounds and starting points.
- **Schools are not homogenous:** within any organisation, there are likely to be both missionaries and blockers of a new idea or initiative; left and right hands; change agents, vested interests, defensive positions, hidden agendas, intrigues; more, or less, commitment to consultation at different levels within the hierarchy; delaying tactics, forced pace, wait-and-see management stances; sometimes major differences of style and substance between key people on crucial issues. And so on. It is a short-hand bordering on an untruth to talk of 'schools' – as we often do – as if they were monolithic agents, rather than micro-political institutions. Value added analyses are impartial in their intention, but rarely neutral in their effect in such a context.
- Therefore, **value added analyses are an innovation to be managed like any other.** All of the above factors mean that when it comes to working with performance data of a kind that attempts to identify strengths and expose weaknesses in performance, planned active management of its dissemination is necessary. But not sufficient: there also needs to be an understanding of the in-school processes by which the data can be 'translated' into an instrument for helping staff make improvements, rather than simply undermining them or entrenching some of them more deeply. What was discovered in the 1980s about the complexities of managing change needs to be re-activated.

In conclusion, even though value added measures are only part of the story of school effectiveness, it is hard to see how schools can take steps towards improving pupils' performance without using – and actively seeking to use – data of the highest possible quality. Value added analyses represent a major step towards this goal. But it is imperative that the analyses are based on valid and reliable data which is analysed and disseminated in such a way to be *useful and appropriate to the complex and often long-term task of raising achievement*. The emphasis in the management of valued added should be on continuing professional development and diagnosing the quality of teaching and learning.

6. CONCLUSIONS: THE FUTURE FOR VALUE ADDED IN EDUCATION

Value added is clearly here to stay, at least for the foreseeable future. As was said in Section 1, the introduction of the national value added system from autumn 1998 can be seen as the culmination of a decade of sustained and public argument about how to measure the performance of pupils in the nation's schools in a way which sheds light on progress as well as standards. It should be clear from the evidence in Sections 2, 3 and 4 how much work has been put into developing the concepts, methods and procedures for doing so.

Even so, as was claimed in the preceding section, one of the most difficult questions to answer with any certainty is: Is the effort involved, at every level, in producing all this information – especially highly sophisticated analyses of performance data – worthwhile? Does it help improve education, and young people's achievements, in practice? These must be the overarching questions for the value added agenda in future.

Whilst researchers have an ethical as well as political duty to investigate ways in which these questions can continue to be addressed, we need to be aware of the limits of research. First of all, all data is imperfect and historical. (This does not, of course, absolve researchers from the requirement to establish as far as possible the validity, reliability, generalisability and transferability of their findings.) Value added data (or any quantitative data based on averages and aggregates) is inherently probabilistic and has a degree of statistical uncertainty built into it. The search for a school effectiveness measure which can summarise how well a school is performing, and is also simple to calculate, understand and use, is bound to be ultimately fruitless. And of course, not all that is desirable in terms of educational outcomes is measurable. The requirement is rather that data should be fit for the purpose; and this means knowing what the purpose of data is. The key questions to which school managers, LEA advisers/inspectors and national government agencies need data-supported answers are likely to overlap but also to have irreducible differences. As was argued in Section 3, value added data which is relevant and useful for internal evaluation and review, for example, is unlikely to be sufficiently valid and reliable for making public comparisons of effectiveness and decisions about spending or school closures.

Secondly, the school improvement agenda is a particularly challenging one, for government, for schools, for teachers and for the community at large. It entails a

pluralist and dynamic view of success which takes into account the changing needs of society and the labour market on the one hand and the devolution of decision-making to local level on the other. Although managers and policy-makers should be tireless in their search for high quality evidence, research and evaluation in themselves are a necessary but not sufficient basis for decision-making about what to do, and what will work, in different institutional, local and national contexts.

However, three major areas of school improvement research which can support the agenda for value added in future are:

6.1 Identifying and Addressing Under-Achievement

Tackling the under-achievement of a substantial proportion of young people is a priority for most developed industrial countries, in order to reduce economic wastage, retain people in learning later in life, and provide greater satisfaction and career progression for individuals. Under-achievement in school may take different forms, with different aetiologies, such as:

- ♦ individual subject areas with problems of capacity, organisation and management;
- ♦ teaching and learning styles, and/or the curriculum diet, not appropriate for a (changing) school population;
- ♦ inadequate flows, and use, of information about the progress being made by pupils between public tests and examinations;
- ♦ a small number of pupils who show persistent under-achievement throughout their school career;
- ♦ pupils not having equal access to resources in practice, because of low self- or teacher-expectations;
- ♦ deep-rooted problems facing many of the pupil population, such as high local rates of unemployment, poor housing and poverty.

There are no easy answers to such problems, and interactive solutions need to be devised at national, local, institutional, classroom and individual levels. Diagnosis may be impeded if under-achievement is conflated with low achievement. It is crucial to know whether pupils and schools are attaining the levels of performance which they could and should be expected to do. Schools getting good results may nonetheless be under-performing relative to what their pupils could be expected to achieve given, for example, their level of ability on entry to the school; and, of course, *vice versa*. Some recent unpublished NFER work analysing aggregated QUASE data (Schagen, 1998) suggests that schools which are under-performing in this way display distinctly

different characteristics from those which are over-performing. The former are more likely to be located in inner cities, with lower levels of parental support, and higher percentages of pupils eligible for free school meals and/or diagnosed as having learning difficulties.

As Mortimore and Whitty (1997) persuasively argued, the influence of socio-economic factors on school performance is clear, and a strong case could be made for targeting financial and material resources and other forms of support on disadvantaged schools. At the very least, better methods for controlling for these factors in value added analyses are clearly needed; there is also more work to be done on devising the most sensitive measures for assessing prior attainment and socio-economic factors, and on checking that the right models are being applied; the NFER is contributing to this important area of policy-related research.

Moreover, schools are often differentially effective for different groups of pupils – as much value added work, including that of the NFER, also shows. Yet only if the 'school effect' is enhanced for all pupils, irrespective of sex, age, ethnicity, ability or socio-economic context, can under-achievement be said to have been addressed. At the school level, diagnosing precisely where a school's problems lie is a critical first step for which detailed value added data is particularly useful. (See Saunders, 1996.)

But the body of knowledge about the ways schools influence outcomes for different pupils is not well developed: more research is needed both into the generalisability of models of educational effectiveness and into the links between class/teacher and school effects. The quote from Goldstein (1997) is worth repeating: *[The conditions for valid inference] imply considerable expenditure of thought and resources as well as the long term follow-up of large numbers of students, schools and classrooms. They need to be replicated across educational systems, phases and types of schooling.* Like other leading institutions, the NFER is committed to working collaboratively with practitioners and with academics in the UK and overseas in order to help build such a body of combined empirical and theoretical knowledge.

6.2 Institutional Development

The decision to devolve funding and planning to school or 'site' level has made great demands on both government and school managers. On the one hand, a national educational system is composed in practice of highly diverse individual 'ecologies', that is, schools with a huge range of different starting-points, profiles and priorities that require different strategies for change. No single approach to improvement will

therefore work in every case, and national policy can only ever be a crude instrument. A great onus is on headteachers and senior management teams to demonstrate improvements in their pupils' achievements through their strategic planning and resource deployment. The studies carried out in the late '80s and early '90s identified key characteristics of effective schools in terms of a policy and management orientation, such as:

- clear school mission;
- effective site management;
- instructional leadership;
- climate of high expectations;
- shared and clear goals;
- safe and orderly environment;
- collaborative working relationships.

Much has since been written about the shortcomings of school effectiveness models, both in themselves and for contributing relevant information for school improvement. For example, important work has been done on 'failing' schools which suggests that such institutions are not simply deficient in one or more of the effectiveness correlates but have their own 'pathologies' (see Reynolds, 1996). Yet one valuable insight was that a school is more than just the sum of its parts. It cannot be seriously doubted that there is a consequent need for whole-school planning, development and review, seen in terms of a continuous cycle usually referred to as school development planning. As the DfEE guidance argues (see, for example, GB. DfEE, 1996; GB. DfEE. SEU, 1997; OFSTED, 1998), an integral part of this cycle must be systematic self-evaluation by schools. This should be tailored to reflect their own challenges and priorities rather than be confined to a single standard model. The process needs to be supported by LEA advisors, and to make use of current research findings in a way which is sensitive to school managers' and teachers' thinking.

But it is clear that the culture of school self-evaluation is not yet fully formed. As the Centre for Educational Research and Innovation (St John-Brooks 1995) put it:

Although evaluators in most countries would like to develop a 'climate of review' in the schools they are assessing, this is hard to achieve – except in schools which already have a self-confident staff and effective leadership – without a substantial input of professional training. But the act of collecting data for the indicators and more general criteria used in evaluations, and discussing their use, can help schools to focus

on and analyse their task; this important aspect is ripe for development in many countries.

This would seem to summarise the current position in England and Wales.

The NFER has been active in developing models and materials to support systematic institutional self-evaluation: see, for example, Saunders *et al.*, 1996; Saunders, 1998. Even so – as was argued in Section 5 – more attention must be paid by researchers and policy-makers to *institutional context and culture* if performance data analyses are to make a difference to school practice.

One small note of warning may need to be sounded, however, so that school managers keep firmly in mind that the ultimate purpose of school self-evaluation is to get better at meeting the needs of their students. Otherwise, this may be the outcome:

I used to think that this school cared about how well I was doing. Now I just think the only thing it cares about is how well *it's* doing. (Year 10 student, quoted in MacBeath *et al.*, 1996, p. 47)

6.3 Partnership between Schools and Local Education Authorities

In a devolved system such as now exists in England and Wales, different functions must be carried out at different levels. A key issue for educational partnerships is to identify at what level – institutional, local, regional and national – the different functions of regulation, advice and guidance, strategic planning, resource allocation and quality assurance should come into play, and how the different levels can effectively be linked with each other. In this context, LEAs need to be bridges between national policy-making bodies and individual institutions. The most important kinds of function which can be fulfilled by LEAs in relation to school improvement are:

- ♦ to assist with quality control, by inspecting and regulating schools according to a statutory set of criteria or performance indicators;
- ♦ to develop quality assurance, by challenging and supporting schools to interrogate and improve their own practice; this includes the production and dissemination of management information, especially comparative performance data analyses, and support for school self-review;
- ♦ to help change the culture, by mediating long-term partnerships between the educational 'stakeholders' (schools, parents, enterprises, higher education institutions, social services agencies, the voluntary sector and local communities). When it comes to raising the level of young people's achievements and aspirations, schools clearly cannot do it all by themselves.

Different LEAs are at very different stages of development in these key areas. Interestingly, current work at the NFER is revealing that school senior managers' expectations of the LEA and its role in raising achievement are not the same in the late '90s as they were earlier in the decade. Rather than autonomy and devolution, headteachers seem to be looking for leadership, strategic planning and intervention on the part of the LEA. This is an important message for LEAs to take on board in view of their enhanced role in school improvement.

6.4 Final Thoughts

School effectiveness and improvement – and their connections with 'value added' analyses – are exciting, but sometimes frustrating, areas to be working in. Whilst there is a great deal that we now know about both effectiveness and improvement, the challenges which remain for research into value added are substantial. They include:

- ♦ isolating the pupil- and school-level factors which are associated with better or worse performance but are still undiagnosed;
- ♦ deriving better models of what makes schools effective for different groups of pupils, especially those who are at greatest risk of under-achievement;
- ♦ having a better theoretical grasp of the role of institutional and local 'climate' or micro-politics in school change and improvement;
- ♦ knowing how to involve teachers more deeply in action research so that teaching and learning processes in the classroom become a central instead of a peripheral aspect of this area of research;
- ♦ getting across key messages from the huge body of research findings in ways that make sense and are coherent and accessible to education managers and practitioners.

And above all, perhaps, researchers need to hold fast to the idea that there are no quick fixes, easy answers or standard formulae whilst still believing in the possibilities of, indeed the necessity of, sustained educational improvement.

APPENDIX A

BIBLIOGRAPHY OF THE NFER'S WORK IN VALUE ADDED

FLETCHER-CAMPBELL, F. (Ed) (1996). *Value-added and Special Educational Needs: Proceedings of EMIE/NFER Seminars 28 November 1995 (Slough) 14 March 1996 (Stockport)*. Slough: NFER.

HUTCHISON, D. (1991). 'Assessing the effectiveness of schools', *NFER News*, 55, 1.

HUTCHISON, D. (1993). 'School effectiveness studies using administrative data', *Educational Research*, 35, 1, 27-47.

HUTCHISON, D. (1997). 'The effect of group-level influences on pupils' progress in reading.' Presentation to the International Congress on School Effectiveness and Improvement, Memphis, USA, 8 January.

HUTCHISON, D. (1997). *The Effect of Group-Level Influences on Pupils' Progress in Reading*. Unpublished report.

KENDALL, L. (1995). 'Contextualization of school examination results 1992', *Educational Research*, 37, 2, 123-39.

KENDALL, L. (1996). *Examination Results in Context. Report on the Analysis of 1994 Examination Results*. London: Association of Metropolitan Authorities.

KENDALL, L. (1997). *Examination Results in Context: Report on the Analysis of 1995 Examination Results*. London: Association of Metropolitan Authorities.

KENDALL, L. (1998). 'Research into factors influencing GCSE performance in London boroughs.' Paper presented at the International Congress on School Effectiveness and Improvement, University of Manchester, 4-7 January.

KENDALL, L. and AINSWORTH, L. (1997). *Examination Results in Context: Analysis of 1996 Examination Results*. London: Local Government Association.

RUDDOCK, G., HUTCHISON, D. and SAINSBURY, M. (1993). *An Investigation into the Relationship Between the Ten-level Scale and the Development of Value-added Indicators of Achievement by Individual Schools in 1996*. Unpublished report.

SAUNDERS, L. (1996). 'QUASE and the special education sector.' In: FLETCHER-CAMPBELL, F. (Ed) *Value-added and Special Educational Needs: Proceedings of EMIE/NFER Seminars 28 November 1995 (Slough) 14 March 1996 (Stockport)*. Slough: NFER.

SAUNDERS, L. (1996). 'Raising attainment in secondary schools: how self-evaluation can help.' Presentation to European Conference on Educational Research, Seville, Spain, 26 September.

SAUNDERS, L. (1996). 'Using value-added measurement of performance for school improvement: what, how and why?' Presentation to European Conference on Educational Research, Seville, Spain, 28 September.

SAUNDERS, L. (1997). 'Value-added principles, practice and ethical considerations.' In: HARRIS, A., BENNET, N. and PREEDY, M. (Eds) *Organisational Effectiveness and Improvement in Education*. Milton Keynes: Open University Press.

SAUNDERS, L. (forthcoming). 'A brief history of educational "value added": how did we get to where we are?' *School Effectiveness and School Improvement*.

SAUNDERS, L. (forthcoming). 'Value added analyses of school performance', *The Letts Network Maths Advisor*.

SAUNDERS, L. (forthcoming). 'Value added analyses of school performance', *The Letts Network Science Advisor*.

SAUNDERS, L. (forthcoming). "'Value added": a primer', *TOPIC*.

SAUNDERS, L. (forthcoming). 'Value added measurement of school performance: the right thing for the wrong reason?' *Education, Public Law and the Individual*.

SAUNDERS, L. and SCHAGEN, I. (1994). 'Developing a system for school self-evaluation based on quantitative data.' Paper presented at the 20th Annual Conference of the British Educational Research Association, St Anne's College, Oxford, 9 September.

SAUNDERS, L. and THOMAS, S. (1998). 'Into an uncertain world', *Times Educ. Suppl.*, **4280**, 10 July, 26.

SCHAGEN, I. (1991). 'Beyond league tables. How modern statistical methods can give a truer picture of the effects of schools', *Educational Research*, **33**, 3, 216-22.

SCHAGEN, I. (1994). 'Multilevel analysis of the key stage 1 National Curriculum assessment data in 1991 and 1992', *Oxford Review of Education*, **20**, 2, 163-71.

SCHAGEN, I. (1995). 'In the premier league', *Guardian Education*, 23 May, 3.

SCHAGEN, I. (1996). *QUASE: Quantitative Analysis for Self-Evaluation. Technical Report 1996: Analysis of GCSE Cohorts 1993 to 1995*. Slough: NFER.

SCHAGEN, I. (1996). "'Value-added" versus self-evaluation in England and Wales.' Paper presented to the International Congress on School Effectiveness and Improvement, Minsk, Belarus, 4 January.

SCHAGEN, I. (1997). 'Value added taxes the statisticians', *Times Educ. Suppl.*, **4210**, 7 March, 14.

SCHAGEN, I. (1998). Exploring School Effectiveness and 'Value-added Quadrants' Via QUASE Data. Unpublished working paper.

SCHAGEN, I. and MORRISON, J. (1998). *QUASE: Quantitative Analysis for Self-Evaluation. Overview Report 1997: Analysis of GCSE Cohorts 1994 to 1996*. Slough: NFER.

SCHAGEN, I. and MORRISON, J. (forthcoming). 'A methodology for judging departmental performance within schools', *Educational Research*.

SCHAGEN, I. and WESTON, P. (1998). 'Insights into school effectiveness from analysis of OFSTED's school inspection database', *Oxford Review of Education*, **24**, 3, 337-44

SCHAGEN, I. and SAINSBURY, M. (1996). 'Multilevel analysis of the key stage 1 National Curriculum assessment data in 1995,' *Oxford Review of Education*, **22**, 3, 265-72.

WHETTON, C. (1998). 'The new chimera: evaluating progress in pupils and effectiveness in schools.' Paper presented at 'Testing and Evaluation: Confronting the Challenges of Rapid Social Changes', International Association for Educational Assessment 24th Annual Conference, Barbados, 12 May.

APPENDIX B

REFERENCES

- BARNARD, P. and DIXON, S. (1998). 'Value added: beyond A-levels to vocational programmes?' *FEDA Bulletin*, 2, 4, (whole issue).
- BIRNBAUM, I. (1993). 'Value added variations' (Digest), *Education*, 182, 20, 12 November.
- BLUESTONE, B. (1994). 'The inequality express', *The American Prospect*, 20, 81-93.
- BROWN, S., DUFFIELD, J. and RIDDELL, S. (1995). 'School effectiveness research: the policy makers' tool for school improvement?' *EERA Bulletin*, 1, 1, 6-15.
- BRYK, A. and WEISBERG, H.I. (1976). 'Value-added analysis: a dynamic approach to the estimation of treatment effects', *Journal of Educational Statistics*, 1, 2, 127-55.
- BUCK, D. (1993). 'Value addition: a reappraisal of school exam league tables', *Educational Psychology in Practice*, 9, 2, 89-93.
- BYATT, A.S. (1997). *Babel Tower*. London: Vintage.
- COOPERS & LYBRAND (1988). *Local Management of Schools: a Report to the DES*. London: DES.
- CORNALL, P. and LOFTHOUSE, M. (1993). *Value Added: "Taking Account of the Social, Economic and Cultural Context of Schools"* (CSCS Discussion Paper 2). Northampton: Centre for the Study of Comprehensive Schools.
- DEARING, R. (1993). *The National Curriculum and Its Assessment: Interim Report*. London: SCAA.
- EDUCATION (1992). 'The value-added tasks' (Digest), *Education*, 180, 5, 31 July.
- ELKIN, S. (1995). *Adding Value and Improving Performance*. London: City Technology Colleges Trust.
- FALLON, M. (1991). 'Education Matters' (Extract), Radio 5, 17 November.
- FITZ-GIBBON, C.T. (1994). 'Three routes to the wrong answer', *Times Educ. Suppl.*, 4093, 9 December, 11.
- FITZ-GIBBON, C.T. (1995). 'Measuring value added in schools.' In: IMAC RESEARCH *Education and Training Statistics Conference Papers, Statistics Users Council Annual Conference, 22 November 1995*. Esher: IMAC Research Ltd.

FITZ-GIBBON, C.T. (1996). 'The Value Added National Project', *Head Teachers Review*, Spring, 18-20.

FITZ-GIBBON, C.T. (1997). *The Value Added National Project: Final Report. Feasibility Studies for a National System of Value Added Indicators*. London: SCAA.

FITZ-GIBBON, C.T. and TYMMS, P. (1993). *Value Added: a Perspective on the Contribution from Examination Results* (CSCS Discussion Paper 1). Northampton: Centre for the Study of Comprehensive Schools.

FITZ-GIBBON, C.T. and TYMMS, P. (1995). *The Value Added National Project: General Report. Issues to be Considered in the Design of a National Value Added System*. London: SCAA.

FREEMAN, R. and KATZ, L. (1993). 'Rising wage inequality: the United States vs. other advanced countries.' In: FREEMAN, B. *Working Under Different Rules*. New York: Russell Sage Foundation.

GALLAGHER, A. (1991). 'Comparative value added as a performance indicator', *Higher Education Review*, 23, 3, 19-29.

GOLDSTEIN, H. (1987). *Multilevel Models in Educational and Social Research*. London: Charles Griffin.

GOLDSTEIN, H. (1997). 'From raw to half-baked', *Times Educ. Suppl.*, 4229, 18 July, 15.

GOLDSTEIN, H. (1997). 'Methods in school effectiveness research', *School Effectiveness and School Improvement*, 8, 4, 369-95.

GOLDSTEIN, H. (1997). 'Value added tables: the less-than-holy grail', *Managing Schools Today*, 6, 6, 18-19.

GOLDSTEIN, H. and CUTTANCE, P. (1988). 'A note on national assessment and school comparisons', *Journal of Education Policy*, 3, 2, 197-202.

GOLDSTEIN, H. and THOMAS, S. (1995). 'School effectiveness and "value-added" analysis', *Forum*, 37, 2, 36-8.

GRAY, J. (1993). *Value-added Approaches in School Evaluation: the Experiences of Three LEAs in England. Lessons and Challenges*. Edinburgh: The Scottish Office Audit Unit.

GRAY, J. (1996). 'The use of assessment to compare institutions.' In: GOLDSTEIN, H. and LEWIS, T. (Eds) *Assessment: Problems, Developments and Statistical Issues*. Chichester: John Wiley and Sons Ltd.

GRAY, J., JESSON, D. and JONES, B. (1984). 'Predicting differences in examination results between Local Education Authorities: does school organisation matter?' *Oxford Review of Education*, **10**, 1, 45-68.

GRAY, J., JESSON, D. and JONES, B. (1986). 'The search for a fairer way of comparing schools' examination results', *Research Papers in Education*, **1**, 2, 91-122.

GRAY, J., JESSON, D. and SIME, N. (1990). 'Estimating differences in the examination performances of secondary schools in six LEAs: a multi-level approach to school effectiveness', *Oxford Review of Education*, **16**, 2, 137-58.

GREAT BRITAIN. DEPARTMENT FOR EDUCATION (1995). *GCSE to GCE A/AS Value Added: Briefing for Schools and Colleges*. London: DfE.

GREAT BRITAIN. DEPARTMENT FOR EDUCATION (1995). *Value Added in Education: a Briefing Paper from the Department for Education*. London: DfE.

GREAT BRITAIN. DEPARTMENT FOR EDUCATION AND EMPLOYMENT (1996). *Setting Targets to Raise Standards: a Survey of Good Practice* (Improving Schools Series). London: DfEE.

GREAT BRITAIN. DEPARTMENT FOR EDUCATION AND EMPLOYMENT (1998). *Target-setting in Schools* (Circular No.11/98). London: DfEE.

GREAT BRITAIN. DEPARTMENT FOR EDUCATION AND EMPLOYMENT. STANDARDS AND EFFECTIVENESS UNIT (1997). *From Targets to Action: Guidance to Support Effective Target-setting in Schools*. London: DfEE.

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

GREAT BRITAIN. PARLIAMENT. HOUSE OF COMMONS (1997). *Excellence in Schools* (Cm.3681). London: The Stationery Office.

HARGREAVES, D. (1995). 'School culture, school effectiveness and school improvement', *School Effectiveness and School Improvement*, **6**, 1, 23-46.

HILL, P.W. and ROWE, K.J. (1996). 'Multilevel modelling in school effectiveness research', *School Effectiveness and School Improvement*, **7**, 1, 1-34.

HUTCHISON, D. (1993). 'School effectiveness studies using administrative data', *Educational Research*, **35**, 1, 27-47.

HUTCHISON, D. (1997). 'The effect of group-level influences on pupils' progress in reading.' Presentation to the International Congress on School Effectiveness and Improvement, Memphis, USA, 8 January.

HUTTON, W. (1996). 'Educated guesses are wide of mark', *Guardian*, 8 January, 15.

- JESSON, D. (1992). 'Valuable addition' (Digest), *Education*, **180**, 4, 24 July.
- JESSON, D. (1996). *Value Added Measures of School GCSE Performance: an Investigation into the Role of Key Stage 3 Assessments in Schools. Interim Report* (DFEE Research Studies 14). London: HMSO.
- KELLNER, P. (1997). 'Hit-and-miss affair', *Times Educ. Suppl.*, **4238**, 19 September, 23.
- KENNEDY, J. (1991). 'Adding value to raw scores', *Managing Schools Today*, **1**, 4, 13-14.
- LAUDER, H., JAMIESON, I. and WIKELEY, F. (1998). 'Models of effective schools: limits and capabilities.' In: SLEE, R., WEINER, G. and TOMLINSON, S. (Eds) *School Effectiveness for Whom? Challenges to the School Effectiveness and School Improvement Movements*. London: Falmer Press.
- MacBEATH, J., BOYD, B., RAND, J. and BELL, S. (1996). *Schools Speak for Themselves: Towards a Framework for Self-evaluation*. Glasgow: University of Strathclyde, Quality in Education Centre.
- McGEEVOR, P., GILES, C., LITTLE, B., HEAD, P. and BRENNAN, J. (1990). *The Measurement of Value Added in Higher Education: a Joint PCFC/CNAA Project Report*. London: Council for National Academic Awards.
- McPHERSON, A. (1992). *Measuring Added Value in Schools* (NCE Briefing 1). London: National Commission on Education.
- MANDEL, E. (1962). *Marxist Economic Theory*. London: Merlin Press.
- MAUGHAN, B. and RUTTER, M. (1986). 'Black pupils' progress in secondary schools: II. Examination attainments', *British Journal of Developmental Psychology*, **4**, 1, 19-29.
- MAUGHAN, B. and RUTTER, M. (1987). 'Pupils' progress in selective and non-selective schools', *School Organization*, **7**, 1, 50-68.
- MORTIMORE, P. (1998). *The Road to Improvement: Reflections on School Effectiveness*. Lisse: Swets & Zeitlinger.
- MORTIMORE, P., SAMMONS, P. and THOMAS, S. (1994). 'School effectiveness and value added measures', *Assessment in Education*, **1**, 3, 315-32.
- MORTIMORE, P. and STONE, C. (1990). 'Measuring educational quality', *British Journal of Educational Studies*, **39**, 1, 69-82.
- MORTIMORE, P. and WHITTY, G. (1997). *Can School Improvement Overcome the Effects of Disadvantage?* London: University of London, Institute of Education.

NICKELL, S. and BELL, B. (1995). 'The collapse in demand for the unskilled and unemployment across the OECD', *Oxford Review of Economic Policy*, **11**, 1, 40-62.

NUTTALL, D. (1991). 'An instrument to be honed: tables do not reflect schools' true performance', *Times Educ. Suppl.*, **3924**, 13 September, 22.

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

NUTTALL, D., THOMAS, S. and GOLDSTEIN, H. (1992). *Report on Analysis of 1990 Examination Results* (AMA Project on Putting Examination Results in Context). London: Association of Metropolitan Authorities.

OFFICE FOR STANDARDS IN EDUCATION (1998). *School Evaluation Matters*. London: OFSTED.

PATERSON, L. (1991). 'Socio-economic status and educational attainment: a multi-dimensional and multi-level study', *Evaluation and Research in Education*, **5**, 3, 97-121.

PRATT, J., TRAVERS, T. and BURGESS, T. (1978). *Costs and Control in Further Education*. Windsor: NFER Publishing Co.

QUALIFICATIONS AND CURRICULUM AUTHORITY (1998). *Value Added Measures in School Performance Tables: Consultation Paper*. London: QCA.

RAUDENSBUSCH, S. and BRYK, A.S. (1986). 'A hierarchical model for studying school effects', *Sociology of Education*, **59**, 1, 1-17.

REYNOLDS, D. (1995). 'The effective school: an inaugural lecture', *Evaluation and Research in Education*, **9**, 2, 57-73.

REYNOLDS, D. (1996). 'Turning around ineffective schools: some evidence and some speculations.' In: GRAY, J., REYNOLDS, D., FITZ-GIBBON, C. and JESSON, D. *Merging Traditions: the Future of Research on School Effectiveness and School Improvement*. London: Cassell.

REYNOLDS, D., BOLLEN, R., CREEMERS, B., HOPKINS, D., STOLL, L. and LAGERWEIJ, N. (Eds) (1996). *Making Good Schools: Linking School Effectiveness and School Improvement*. London: Routledge.

REYNOLDS, D., HOPKINS, D. and STOLL, L. (1993). 'Linking school effectiveness knowledge and school improvement practice: towards a synergy', *School Effectiveness and School Improvement*, **4**, 1, 37-58.

RUTTER, M., MAUGHAN, B., MORTIMORE, P. and OUSTON, J. with SMITH, A. (1979). *Fifteen Thousand Hours: Secondary Schools and their Effects on Children*. London: Open Books.

- ST. JOHN-BROOKS, C. (1995). *Schools Under Scrutiny*. Paris: OECD.
- SAMMONS, P. (1996). 'Complexities in the judgement of school effectiveness', *Educational Research and Evaluation*, 2, 2, 113-49.
- SAMMONS, P., THOMAS, S., MORTIMORE, P., OWEN, C., PENNELL, H. and HILLMAN, J. (1994). *Assessing School Effectiveness: Developing Measures to put Performance in Context (Summary Report)*. London: OFSTED and University of London, Institute of Education.
- SAUNDERS, L. (1998). 'Who or what is school 'self'-evaluation for?' Paper presented at the European Conference on Educational Research, University of Ljubljana, Slovenia, 20 September.
- SAUNDERS, L. (forthcoming). 'A brief history of educational "value added": how did we get to where we are?' *School Effectiveness and School Improvement*.
- SAUNDERS, L. (forthcoming). 'Value added analyses of school performance', *The Letts Network Maths Advisor*.
- SAUNDERS, L. (forthcoming). 'Value added analyses of school performance', *The Letts Network Science Advisor*.
- SAUNDERS, L. (forthcoming). "'Value added": a primer', *TOPIC*.
- SAUNDERS, L. (forthcoming). 'Value added measurement of school performance: the right thing for the wrong reason?' *Education, Public Law and the Individual*.
- SAUNDERS, L. and SCHAGEN, I. (1994). 'Developing a system for school self-evaluation based on quantitative data.' Paper presented at the 20th Annual Conference of the British Educational Research Association, University of Oxford, 9 September.
- SAUNDERS, L. and STRADLING, B. with GALLACHER, S. (1996). *Raising Attainment in Secondary Schools: a Handbook for Self-evaluation*. Slough: NFER.
- SAUNDERS, L. and THOMAS, S. (1998). 'Into an uncertain world', *Times Educ. Suppl.*, 4280, 10 July, 26.
- SCHAGEN, I. (1991). 'Beyond league tables. How modern statistical methods can give a truer picture of the effects of schools', *Educational Research*, 33, 3, 216-22.
- SCHAGEN, I. (1994). 'Multilevel analysis of the key stage 1 National Curriculum assessment data in 1991 and 1992', *Oxford Review of Education*, 20, 2, 163-71.
- SCHAGEN, I. (1996). *QUASE: Quantitative Analysis for Self-Evaluation. Technical Report 1996: Analysis of GCSE Cohorts 1993 to 1995*. Slough: NFER.
- SCHAGEN, I. (1997). 'Value added taxes the statisticians', *Times Educ. Suppl.*, 4210, 7 March, 14.

- SCHAGEN, I. (1998). Exploring School Effectiveness and 'Value-added Quadrants' Via QUASE Data. Unpublished working paper.
- SCHOOL CURRICULUM AND ASSESSMENT AUTHORITY (1994). *Value-added Performance Indicators for Schools*. London: SCAA.
- SMITH, D. and TOMLINSON, S. (1989). *The School Effect: a Study of Multiracial Comprehensives*. London: Policy Studies Institute.
- SMITH, P. (1995). 'On the unintended consequences of publishing performance data in the public sector', *Journal of Public Administration*, **18**, 2/3, 277-310. Cited in: FITZ-GIBBON, C.T. (1997). *The Value Added National Project: Final Report. Feasibility Studies for a National System of Value Added Indicators*. London: SCAA.
- SOCIETY OF EDUCATION OFFICERS (1995). *Value Added and School Improvement*. Manchester: SEO.
- SPOURS, K. and HODGSON, A. (1996). *Value-added & Raising Attainment: a Formative Approach. A Resource Pack for Practitioners*. Poole: BP Education Service.
- STRONACH, I. and MACLURE, M. (1997). *Educational Research Undone: the Postmodern Embrace*. Buckingham: Open University Press.
- SWEETMAN, J. (1997). *Curriculum Confidential 7: the Complete Guide to the National Curriculum*. Westley: Courseware Publications.
- THOMAS, S. (1998). 'Value-added measures of school effectiveness in the United Kingdom', *Prospects*, **28**, 1, 91-108.
- THOMAS, S. and GOLDSTEIN, H. (1995). 'Questionable value', *Education*, **185**, 11, 17 March, 17.
- THOMAS, S. and MORTIMORE, P. (1996). 'Comparison of value-added models for secondary-school effectiveness', *Research Papers in Education*, **11**, 1, 5-33.
- THOMAS, S., SAMMONS, P., MORTIMORE, P. and SMEES, R. (1997). 'Stability and consistency in secondary schools' effects on students' GCSE outcomes over three years', *School Effectiveness and School Improvement*, **8**, 2, 169-97.
- WEBSTER, W.J., MENDRO, R.L. and ALMAGUER, T.O. (1994). 'Effectiveness indices: a "value added" approach to measuring school effect', *Studies in Educational Evaluation*, **20**, 1, 113-45.
- WEST, M. and MOORE, S. (1998). 'Using added-value data to improve pupil attainment: a case study of practice in one secondary school.' Paper presented at the International Congress on School Effectiveness and Improvement, University of Manchester, 4-7 January.

WHITE, J. and BARBER, M. (Eds) (1997). *Perspectives on School Effectiveness and School Improvement* (Bedford Way Papers). London: University of London, Institute of Education.

WILIAM, D. (1992). 'Value-added attacks: technical issues in reporting National Curriculum assessments', *British Educational Research Journal*, **18**, 4, 329-41.

WILLIAMSON, J. and FITZ-GIBBON, C.T. (1990). 'The lack of impact of information: performance indicators for A levels', *Educational Management & Administration*, **18**, 1, 37-45.

WOODHEAD, C. (1997). 'Inspecting schools: the key to raising educational standards.' Presentation of 'The Last Word Lecture' at the Royal Geographical Society, London, 21 January.

nfer

'value added' measurement of school effectiveness: a critical review

'value added' is one of those ideas which make intuitive sense but are harder to pin down in practice. A critical review of 'value added' in education is particularly relevant in the late 1990s. The introduction of the national value added system from autumn 1998 comes at the end of a decade of sustained and public argument about how to measure pupils' performance in a way that sheds light on schools' contribution to their progress. The debate on school effectiveness and how to measure it is now integrally linked with the national political agenda for educational quality. Moreover, the statutory requirement to set 'challenging but realistic targets' for schools and LEAs is bringing value added analyses closer to the everyday life of schools and teachers.

The range of parties currently interested in value added extends from politicians to school senior managers, and from academic researchers to lay governors. There is consequently a need for continuing discussion which takes into account not only the technical questions – how most sensitively and informatively to measure added value – but also the different agendas, expectations and priorities of different stakeholders.

This report (which is an in-depth companion volume to the recently published Overview) reviews over one hundred articles, reports and features; it concludes that whilst much has been learnt about the contribution of 'value added' to school effectiveness and improvement, the challenges which the idea continues to hold out for practitioners, managers and researchers are substantial.

ISBN: 0 7005 1530 5
£10.00