OTHER TITLES IN THE SCHOOL DEVELOPMENT SERIES

R. Bollington, D. Hopkins and M. West:
An Introduction to Teacher Appraisal

J. Chapman, Vern Wilkinson and D. Aspin:
Quality Schooling

B. Creemers:
The Effective Classroom

P. Dalin:
Changing the School Culture

P. Dalin:
How Schools Improve

M. Fullan:
The New Meaning of Educational Change

D. Hargreaves and D. Hopkins:
The Empowered School

D. Hopkins, M. Ainscow and M. West:
School Improvement in an Era of Change

D. Hopkins and D. Hargreaves:
Development Planning for School Improvement

K. S. Louis and M. B. Miles:
Improving the Urban High School

J. Murphy:
Restructuring Schools

D. Reynolds and P. Cuttance:
School Effectiveness

P. Ribbins and E. Burridge:
Improving Education

J. Scheerens:
Effective Schooling

H. Silver:
Good Schools, Effective Schools

C. Taylor Fitz-Gibbon:
Monitoring Education

M. Wallace and A. McMahon:
Planning for Change in Turbulent Times

MERGING TRADITIONS

The Future of Research on School Effectiveness and School Improvement

Edited by

John Gray,
David Reynolds,
Carol Fitz-Gibbon
and David Jesson

CASSELL

Cassell
Wellington House
125 Strand
London WC2R 0BB

127 West 24th Street
New York
NY 10011

First published 1996

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

Library of Congress Cataloging-in-Publication Data
Merging traditions : the future of research on school effectiveness and school improvement / edited by John Gray . . . [et al.].
p. cm. — (School development series)
'The chapters presented in this book were first given at a seminar series funded by the Economic and Social Research Council' — Pref.
Includes bibliographical references and index.
ISBN 0–304–33653–X. — ISBN 0–304–33647–5 (pbk.)
1. School improvement programs—Research—Great Britain.
2. Educational evaluation—Research—Great Britain. I. Gray, John (John Michael), 1948– . II. Series.
LB2822.84.G7M47 1996
371.2'007—dc20 96–6375
CIP

ISBN 0–304–33653–X (hardback)
0–304–33647–5 (paperback)

Typeset by Action Typesetting Limited, Gloucester
Printed and bound in Great Britain by
Redwood Books, Trowbridge, Wiltshire

Contents

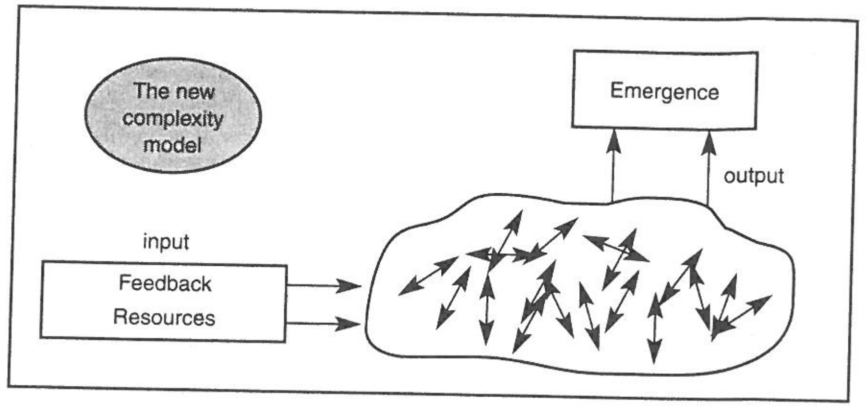

Figure 4.1c *Feedback focus*

Some Key Events

To select a few major currents from the rapid and turbulent flow of human affairs is a difficult if not distressing activity. As historians must be painfully aware, what is represented in a few pen sketches, however long, is but a fraction of what has gone before us. However, just as science looks for broad and simple outlines, which are, to some extent, only an approximation, so must anyone attempting to give a view of the years which have preceded our own.

Although 'complexity' as an area of study is a new development, there were already hints around in physics and mathematics. As long ago as 1908 Henri Poincaré wrote, 'It may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former, will produce an enormous error in the latter. Prediction becomes impossible, and we have the fortuitous phenomenon' (cited in Davies, 1987, p. 53). Decades later, notions of unpredictable systems, non-linear dynamics, cellular automata, emergence, sensitivity to initial conditions and all the delights that you can enjoy by reading Waldrop's 1992 book on *Complexity: the Emerging Science at the Edge of Order and Chaos* are providing a new set of mental models for how we can approach science, the universe, and everything. How did this transformation occur?

We could divide science into two phases: the first phase could be called the discovery of simplicity, and the second the understanding of complexity. The first phase of science is well illustrated by Fourier analysis. French mathematician and physicist Jean G. Fourier, showed that highly complex signals such as are carried on the radio or emerge from outer space, although they seem formless and random, can, if they are periodic, be broken down into a series of sine waves (simple up and down waves) and all the original complexity can be reconstructed by simply combining the pure sine waves. This possibility of adding up simple parts to get complex wholes has made enormous advances in science possible. The whole as a simple sum of parts applies not only to periodic waves as proved mathematically by Fourier, but also to electric and magnetic fields, stresses and strains in many materials, heat flow, diffusion of gases and liquids, and many other

phenomena (Davies, 1987, p. 94). Social scientists will recognize this 'additivity' as underpinning statistical models such as analysis of variance. Systems which can be decomposed and analysed by linear models, multilevel models or 'hierarchical linear' models are called linear; in these systems the whole is simply the sum of the parts.

The second phase of science arises with the study of non-linear systems:

> In a non-linear system a whole is much more than the sum of its parts, and it cannot be reduced or analysed in terms of simple sub-units acting together. The resulting properties can often be unexpected, complicated and mathematically intractable (Davies, 1987, p. 25).

Mathematically intractable! This seems to sound the death knell for science as normally conducted. If systems are chaotic, full of discontinuities, catastrophes, unevenness and unpredictability, science cannot construct mathematical laws that will predict the future. Mathematical modelling meets its limits. Are we then left with nothing but a complexity that we are unable to analyse? Not at all, but the tools of analysis become a kind of modelling using computers, a method which is different from the highly successful previous work in science resting on mathematical equations. Computer modelling can attempt to simulate events in a direct fashion, by rules and procedures, incorporating feedback, rather than by equations (see Tymms in this volume, chapter 6). For example, the universe (scientists are nothing if not willing to consider the big questions) is modelled by a clock ticking away, which is time, and a lattice of cells which can interact according to certain rules. The lattice defines space. Instead of the compelling logic of numbers, the compelling logic of a computer simulation is sought. If a few simple rules are implemented for the interactions of the cells in the lattice and certain patterns emerge, then there is a compelling logic that those few rules are sufficient to produce the emergent patterns. This does not mean they are necessarily the only way to produce the emergent patterns but, since they have done so once, the sufficiency is demonstrated.

These kinds of interactions are called 'cellular automata'. Chris Langton, who had become interested in computing when undertaking alternative military service in a hospital in preference to participation in the Vietnam war, did a literature search on the key words 'self-reproduction'. Von Neumann's *Theory of Self-Reproducing Automata* and a piece of work called *Cellular Automata* by Ted Codd (who had invented relational databases) caught his attention. Langton knew that, in dynamical systems, whether the outcomes were simple or chaotic was often a function of a single parameter; as the parameter increases, the outcomes pass through a chaotic phase. (This interesting behaviour of non-linear systems can be very simply demonstrated on a spreadsheet using the equation:

$$X_n = p^* x_{(n-1)}[1 - X_{(n-1)}]$$

and putting in the values X_0=0.2 and n=1, 2, ... 17 and for successive computation series using these kinds of values for the parameter: p=1, 2, 2.5, 3, 3.5, 3.57, 3.7, 4 and 5.) Langton set out to look for such a parameter in the systems of cellular automata. Eventually, he hit on a simple parameter which could be embodied

Table 4.1 *An Artefact: part of the list discussed by a Senior Management Team* (Figures in the margin show that more than one person suggested this factor)

'Student' Factors

Part-time jobs 11
'Unrealistic' students admitted to courses
Unrealistic parental demands/expectations
(Parents and pupils) lack of real understanding of what A level entails 3
Too many extra-curricular activities 2
Insufficient involvement in life of school/community 2
Students need to know what they are aiming for
Casual approach after easy ride to 16+
Marital/family break-ups
Student absence 2
Attitudes of students to school
Drugs and alcohol
Student responsibility
Career/job prospects
Instrumental approach to being a sixth-form student
Student–teacher relationship
Teachers' reaction to attendance & punctuality
Social activities in sixth form
Total demand made of students

'Teaching/Learning' Factors

Teaching and learning strategies/range of styles (some outmoded) 3
No structured teaching of study skills (e.g. note-taking) 5
Quality of teaching
Resources for teaching and learning (see separate list)
Time for 1:1 discussions (teacher/student) 2
Lack of 'bridging course' induction programme 3
Inadequate concentration on *process* of learning
Inconsistent expectations re marked written work
Lack of exam practice/revision techniques
Importance of homework in students' eyes
Teaching approaches not responsive enough to needs of *all* students
Teacher-pairs not communicating sufficiently
Students not encouraged to read widely enough
Student self-assessment
Progress not reviewed often enough
Number and use of study periods
Lack of common deadlines within and between departments
Teacher expectations of students

GCSE and the Transition to A Level

Cumulative effects of GCSE on students' study habits
GCSE as preparation for A level 3
Transition to A level is difficult
Transition to A level needs planning
Good GCSE grades give false confidence to students
Relevance of GCSE teaching strategies
Students used to coursework demands dislike courses with final exam only
Subjects seen as 'easy' or 'difficult'
Choosing subjects which are inappropriate

Policy Factors

Open-entry policy
Negotiated time out 6
Paired sharing of classes (= diminished teacher responsibility) 2
Inadequate *range* of courses
Poor/old-fashioned syllabus – choice which ignores changes at GCSE
Too many A levels on offer – resources too thinly stretched
Too many A levels on offer – poor *selection* of courses
No credible alternative for the many students who can manage only two (or even one) A level
Inadequate resourcing of vocational courses (e.g. low staffing priority) leads to too many choosing A level inappropriately
Homework policy? 3
Quality of information in Y11
Quality of guidance in Y11, 12, 13
Approach of subject choice at A level (several aspects of this noted, e.g. 'wrong' choices, inappropriate combinations, influence of teachers on subject choice) 5
Lack of academic emphasis in school development
Perceived management indifference to exam results

School Organization

Structure of school day since 1990
Use of registration and tutorial time
Assemblies
Leave of absence: timing and *use* of
Inadequacy of previous pastoral system: inadequate monitoring
Inadequacy of arrangements when staff miss lessons 4

Amount or use of 'free' time
Visits/fieldwork, etc.
Ethos of school

Problems

Too many interruptions to a 5-term course
Too much time wasted between June and September
We have given too much football during lesson time
Staff turnover 2
Retirements, resignations, illnesses, maternity leave
Total sustained demands on staff time and energy in non-academic directions
Instability: many changes since 1988
Complacency: our results seemed OK
General teacher demoralization
Diversion of teachers'/departments'/school's energy towards 2
- GCSE
- National Curriculum
- LMS
- Pastoral reorganization
- Appraisal
- Parent's Charter
- etc., etc.

Lack of debate on A-level teaching
Diversion of SMG attention from essentially curricular matters

Questions

Is our 'performance' as implied by national statistics distorted by the proportion of those statistics which emanate from:
- independent sector (massively better resourced)?
- sixth-form college sector: not distracted by KS3 and KS4 planning?

Are we losing more quality sixth-form students:
- from the state education altogether?
- to neighbouring schools/colleges?

Resources and Facilities

Teaching environment 2
Availability (or not) of study areas especially silent areas and supervised areas 6
General decline in funding for teaching materials – inadequate resourcing 3
Books get (and look) older
Poor, largely book-based, study facilities (library stock; IT resources and facilities) 2

The second illustration of this assertion of the complexity of the system with which we are dealing is due to Ibrahim (1992) and based on data from the A-Level Information System (ALIS), supported by many hours of classroom observation and interviews with teachers and pupils. In Figure 4.2, the Y-axis represents residuals (what is left over after we have taken account of the best single predictor of achievement at A level – the average achievement score two years earlier). Positive residuals indicate performance above that which would be expected on the basis of prior achievement, negative residuals represent achievement below what would be expected on the basis of the prior-achievement measure. Ibrahim divided the sample into high and low ability according to their scores on the International Test of Developed Abilities (Ottobre and Turnbull, 1987). As can be seen from the graph, the amount of individual help given by the teacher was associated with quite different outcomes, according to the ability category of the pupils. More able pupils appeared to do better when not receiving individual help, and less able pupils seemed to need an extensive amount of individual help to achieve at a level consistent with their prior achievement. Interestingly, in discussion, a number of able pupils stated explicitly that, in mathematics, if they could work it out, they liked to work it out for themselves rather than seek help from the teacher because that way they understood it better and remembered it better (Ibrahim, 1992, p. 298). This is correlational data so the direction of cause (if cause there is) cannot be assumed from the data. It serves to illustrate, however, that teaching is directed at individuals and different individuals may need different styles of teaching. Furthermore each action will have consequences and reactions. Giving help

Prediction and Feedback

> An adaptive agent is constantly playing a game with its environment. ... What actually has to happen to game-playing agents to survive and prosper? Two things ... prediction and feedback (Holland, quoted by Waldrop, 1992, p. 282).

In order to make effective decisions for school improvement we need values (to know in which directions we wish to move) and information. To thrive on the edge of chaos, the system needs the capacity to act in multiple self-organizing, adequately informed, local units, i.e. the capacity to act on the basis of feedback. (Indeed the Hawthorne Effect has been reanalysed to show it to have been largely due to feedback (Parsons, 1974).)

Monitoring systems enable schools to make accurate predictions where these are possible and provide a constant flow of feedback. Thus monitoring systems fit in with the model of a school as an evolving complex system.

SIMPLICITY – THE NEED TO KNOW THE MAJOR PATTERNS IN DATA

Although it has been argued above that the educational enterprise is exceedingly complex it can also be shown that stable features can be derived from monitoring data. Just as the *weather* is variable and unpredictable – a frequently cited example of a non-linear, open system, determined by the laws of physics but essentially unpredictable – and yet the *climate* is a fairly reliable guide to the range of weather to be expected, so there are broad features in the system which adequate monitoring can clarify and, by so doing, can inform the interpretation of data and the feedback into the system.

The major point which needs emphasis, and which it is the intention in this section to illustrate, is the extent to which *mistakes can be made if the broad features which pertain throughout a system are not known and/or are not taken into account*. In other words there will be variations in outcomes which are seen throughout the system ('noise') and there will be some stable broad features. Four examples will be drawn from the ALIS monitoring system.

Are some teachers to blame for inaccurate predictions? David Elsom, a former college vice-principal, draws on his experience to portray a scenario which he has frequently encountered: complaints against the English department for being unable to predict accurately their pupils' A-level grades. Anyone thinking to blame individual English department staff needs to recognize that the prediction of grades in English is a problem throughout the system and may have more to do with the nature of the A-level English examination, and its lack of discrimination, than with the competence or otherwise of individual A-level English teachers. Appeals against grades are more frequent in English than in other subjects.

This is an important illustration of two principles: that the broad features of the system must be known before individuals are judged and that the competencies of teachers can only be considered in comparison with similar teachers *teaching the same subject* to similar pupils.

Subject differences are substantial and should not be neglected. Indeed *school effectiveness may need to be reconceptualized, if not for every subject*

then certainly for broad bands of subjects such as the foreign languages, the sciences, mathematics, the humanities, practical or vocational subjects. What applies in one curriculum area may not apply in another.

Can examination results be fairly contextualized by SES data? Rarely is socio-economic status a good predictor of achievement in the UK – which is perhaps a tribute to the UK system of education. SES is not the most important covariate so that comparisons which rely on SES comparisons will be less fair than comparisons of pupils' achievements based on cognitive measures. In general SES may explain about 9 per cent of the variance whereas any prior achievement measure would generally explain about four times as much. Furthermore, within schools the composition of classes varies from year to year and this variation will itself vary from school to school depending perhaps on the 'pulling power' of various departments (Fitz-Gibbon, 1984). If one inner-city French department had recruited all the highly able students it would be an error to adjust its results for the average SES of the school, or even the average achievement level of the school.

Are A-level subjects 'level' in difficulty? Rhodes Robson is an accountancy firm. It has written to schools offering to work out residuals for A-level subjects. How does it compute the predicted/expected grade for a student? By assuming that all subjects are equally difficult at A level and therefore if your student has 24 UCAS points the 'predicted' grade in each of three subjects can be taken as 24/3 or 8. The fact is that A-level subjects differ in difficulty on any reasonable definition (Fitz-Gibbon, 1988). Consequently the approach suggested will lead to unfair chastisement of teachers teaching difficult A levels: maths, sciences and foreign languages. Do trading standards apply in social science?

Table 4.2 shows subjects rank ordered according to their intercept as calculated using ordinary least squares (OLS) regression. We see physics as having the lowest intercept. It appears to have been, that year, the most difficult subject and history was at the other end of the rank order appearing to have been the easiest subject in the 1989 sample. A second method of calculating difficulty (Kelly–Lawley adjustments) gave similar results (Kelly, 1976; Fitz-Gibbon, 1991).

Table 4.2 *The relative difficulty of the 1989 A-level examinations*

Subject	Intercept from OLS regression[1]	Difficulty rank OLS	Difficulty rank Lawley[2]	Lawley correction factors	N
Physics	–9.8	1	1	0.58	867
French	–8.5	2	5	0.12	371
Chemistry	–8.0	3	4	0.28	853
Maths	–7.4	4	3	0.31	1,357
General Studies	–7.4	5	2	0.36	1,087
Biology	–6.7	6	9	–0.54	667
Geography	–6.0	7	8	–0.66	630
Economics	–5.7	8	7	–0.18	606
English	–4.6	9	10	–0.75	831
History	–4.1	10	6	–0.15	674

[1] Based on 1989 ALIS data.
[2] Based on the subset of candidates taking 2 or more subjects.

OTHER TITLES IN THE SCHOOL DEVELOPMENT SERIES

R. Bollington, D. Hopkins and M. West:
An Introduction to Teacher Appraisal

J. Chapman, Vern Wilkinson and D. Aspin:
Quality Schooling

B. Creemers:
The Effective Classroom

P. Dalin:
Changing the School Culture

P. Dalin:
How Schools Improve

M. Fullan:
The New Meaning of Educational Change

D. Hargreaves and D. Hopkins:
The Empowered School

D. Hopkins, M. Ainscow and M. West:
School Improvement in an Era of Change

D. Hopkins and D. Hargreaves:
Development Planning for School Improvement

K. S. Louis and M. B. Miles:
Improving the Urban High School

J. Murphy:
Restructuring Schools

D. Reynolds and P. Cuttance:
School Effectiveness

P. Ribbins and E. Burridge:
Improving Education

J. Scheerens:
Effective Schooling

H. Silver:
Good Schools, Effective Schools

C. Taylor Fitz-Gibbon:
Monitoring Education

M. Wallace and A. McMahon:
Planning for Change in Turbulent Times

MERGING TRADITIONS

The Future of Research on School Effectiveness and School Improvement

Edited by

John Gray,
David Reynolds,
Carol Fitz-Gibbon
and David Jesson

CASSELL

Cassell
Wellington House
125 Strand
London WC2R 0BB

127 West 24th Street
New York
NY 10011

First published 1996

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

Library of Congress Cataloging-in-Publication Data
Merging traditions : the future of research on school effectiveness
and school improvement / edited by John Gray . . . [et al.].
p. cm. — (School development series)
'The chapters presented in this book were first given at a seminar series funded by the Economic and Social Research Council' — Pref.
Includes bibliographical references and index.
ISBN 0-304-33653-X. — ISBN 0-304-33647-5 (pbk.)
1. School improvement programs—Research—Great Britain.
2. Educational evaluation—Research—Great Britain. I. Gray, John (John Michael), 1948– . II. Series.
LB2822.84.G7M47 1996
371.2'007—dc20 96-6375
CIP

ISBN 0-304-33653-X (hardback)
0-304-33647-5 (paperback)

Typeset by Action Typesetting Limited, Gloucester
Printed and bound in Great Britain by
Redwood Books, Trowbridge, Wiltshire

Contents

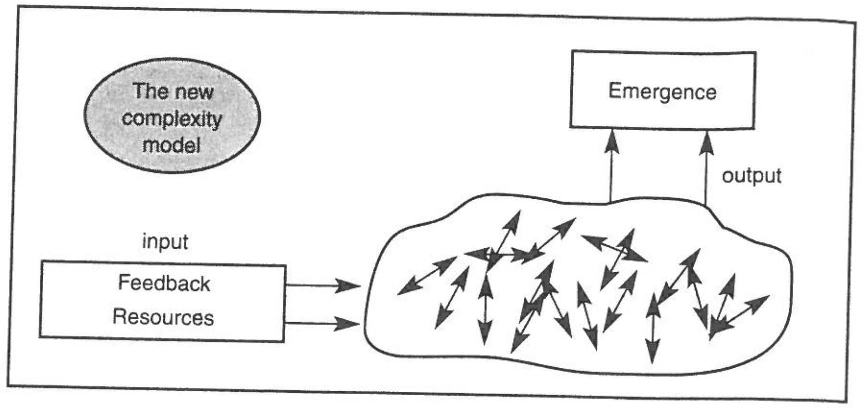

Figure 4.1c *Feedback focus*

Some Key Events

To select a few major currents from the rapid and turbulent flow of human affairs is a difficult if not distressing activity. As historians must be painfully aware, what is represented in a few pen sketches, however long, is but a fraction of what has gone before us. However, just as science looks for broad and simple outlines, which are, to some extent, only an approximation, so must anyone attempting to give a view of the years which have preceded our own.

Although 'complexity' as an area of study is a new development, there were already hints around in physics and mathematics. As long ago as 1908 Henri Poincaré wrote, 'It may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former, will produce an enormous error in the latter. Prediction becomes impossible, and we have the fortuitous phenomenon' (cited in Davies, 1987, p. 53). Decades later, notions of unpredictable systems, non-linear dynamics, cellular automata, emergence, sensitivity to initial conditions and all the delights that you can enjoy by reading Waldrop's 1992 book on *Complexity: the Emerging Science at the Edge of Order and Chaos* are providing a new set of mental models for how we can approach science, the universe, and everything. How did this transformation occur?

We could divide science into two phases: the first phase could be called the discovery of simplicity, and the second the understanding of complexity. The first phase of science is well illustrated by Fourier analysis. French mathematician and physicist Jean G. Fourier, showed that highly complex signals such as are carried on the radio or emerge from outer space, although they seem formless and random, can, if they are periodic, be broken down into a series of sine waves (simple up and down waves) and all the original complexity can be reconstructed by simply combining the pure sine waves. This possibility of adding up simple parts to get complex wholes has made enormous advances in science possible. The whole as a simple sum of parts applies not only to periodic waves as proved mathematically by Fourier, but also to electric and magnetic fields, stresses and strains in many materials, heat flow, diffusion of gases and liquids, and many other

phenomena (Davies, 1987, p. 94). Social scientists will recognize this 'additivity' as underpinning statistical models such as analysis of variance. Systems which can be decomposed and analysed by linear models, multilevel models or 'hierarchical linear' models are called linear; in these systems the whole is simply the sum of the parts.

The second phase of science arises with the study of non-linear systems:

> In a non-linear system a whole is much more than the sum of its parts, and it cannot be reduced or analysed in terms of simple sub-units acting together. The resulting properties can often be unexpected, complicated and mathematically intractable (Davies, 1987, p. 25).

Mathematically intractable! This seems to sound the death knell for science as normally conducted. If systems are chaotic, full of discontinuities, catastrophes, unevenness and unpredictability, science cannot construct mathematical laws that will predict the future. Mathematical modelling meets its limits. Are we then left with nothing but a complexity that we are unable to analyse? Not at all, but the tools of analysis become a kind of modelling using computers, a method which is different from the highly successful previous work in science resting on mathematical equations. Computer modelling can attempt to simulate events in a direct fashion, by rules and procedures, incorporating feedback, rather than by equations (see Tymms in this volume, chapter 6). For example, the universe (scientists are nothing if not willing to consider the big questions) is modelled by a clock ticking away, which is time, and a lattice of cells which can interact according to certain rules. The lattice defines space. Instead of the compelling logic of numbers, the compelling logic of a computer simulation is sought. If a few simple rules are implemented for the interactions of the cells in the lattice and certain patterns emerge, then there is a compelling logic that those few rules are sufficient to produce the emergent patterns. This does not mean they are necessarily the only way to produce the emergent patterns but, since they have done so once, the sufficiency is demonstrated.

These kinds of interactions are called 'cellular automata'. Chris Langton, who had become interested in computing when undertaking alternative military service in a hospital in preference to participation in the Vietnam war, did a literature search on the key words 'self-reproduction'. Von Neumann's *Theory of Self-Reproducing Automata* and a piece of work called *Cellular Automata* by Ted Codd (who had invented relational databases) caught his attention. Langton knew that, in dynamical systems, whether the outcomes were simple or chaotic was often a function of a single parameter; as the parameter increases, the outcomes pass through a chaotic phase. (This interesting behaviour of non-linear systems can be very simply demonstrated on a spreadsheet using the equation:

$$X_n = p^* x_{(n-1)}[1 - X_{(n-1)}]$$

and putting in the values X_0=0.2 and n=1, 2, ... 17 and for successive computation series using these kinds of values for the parameter: p=1, 2, 2.5, 3, 3.5, 3.57, 3.7, 4 and 5.) Langton set out to look for such a parameter in the systems of cellular automata. Eventually, he hit on a simple parameter which could be embodied in the rules for the computer program, the probability that any given cell would be alive in the next generation. Exploring values of this parameter, he found, for low values, simple frozen patterns or periodic repeating patterns. With slightly larger values there was a transition phase and then, with even larger values, there was chaos. Langton was aware of the work of a British mathematician, John Conway, who had created, on computers, a program called the *Game of Life*. It too was a 'cellular automaton' program simulating evolution. With a few simple rules the program would run and create successful species, predator species, extinctions, patterns which were pictured on the screen as coloured cells and which mimicked biological evolution. Langton recognized the *Game of Life* as being in the transition part between the frozen, periodic, ordered phase and the completely chaotic phase. It seems that life emerges, self-reproduction occurs, patterns evolve, compete, cooperate and behave in a lifelike manner *on the edge of chaos*, in a region of 'complexity'.

The ideas of the evolution of effectiveness in open interacting systems are now being applied in many fields: economics, biology, archaeology and speculations on the stock market. Perhaps we are seeing the development of newly discovered laws of nature, laws about how complex systems manage to develop effectiveness – survival, success. One message is that the systems are often *locally organized*. They are not told what to do but they do get regular feedback. Can these ideas apply to schools?

In 1979, in an article entitled 'Policy for the unpredictable', Gene Glass (1979) suggested that it might not be possible to evaluate schools or create widely applicable research findings. Education was so complex that we might simply have to be content with 'fire-fighting': having monitoring systems in place which could alert us when untoward events were happening. This was a prescient article and fits well with modern theories of complexity.

SCHOOLS AS COMPLEX SYSTEMS

Schools are nothing if not complex systems subject to feedback such as examination results and enrolment patterns. They are very likely, then, to show the kind of behaviours to be expected in complexity. Furthermore, they contain within themselves further self-organizing units such as departments and classrooms.

The *flow of information* plays a key role in any complex system. However, it is important to note that in all the writings on complex systems scientists do not seem to be considering the possibility of *disinformation as opposed to valid information*. This is quite understandable in that in the evolution of cells in response to the availability of food sources, light, other cells etc., the feedback to the cell is veridical, uncorrupted, actual. The light is there where the light intensity is greatest; the organism that is eaten provides nourishment, or if it poisons, that poison feeds back into the survival mechanism. There is little in natural systems as clearly corrupted as in human information systems; a point to which we shall return later in asking how schools as complex systems can be effective.

There is a reason why the information flow back to humans or to groups of humans in complex systems is not as accurate as that in the natural systems which have been subject to the greatest amounts of modelling and investigation. The feedback which is needed from a complex activity may be very difficult to acquire.

Take teaching as an example, the teacher gets immediate feedback from pupils about their levels of enjoyment and cooperation, and teachers generally learn class control with practice. The immediacy of the feedback and its lack of ambiguity make it useful for the teachers' learning. *But some feedback, such as how effectively they have enabled pupils to learn, is very difficult to acquire*. The pupils' final learning level can be measured, but a vital piece of information is needed if this final level is to be interpreted: what would it have been if the pupils had been taught some other way? If teachers were always assigned randomly selected, representative groups of pupils then the simple end achievement levels would be interpretable by simple comparisons between teachers. But schools are not like that. Teachers face classes which are exceedingly different from each other, even in the same subject, even, possibly, in the same school. (Vague notions of general correction factors as suggested by the politicians and the Office of Standards in Education make no sense. For example, in one inner-city school, every pupil taking A-level mathematics was highly able and from a professional home background. Schools are not homogeneous and classes within schools can vary substantially in their intake.) Only pupil-by-pupil information interpreted against a large database can sort out what kind of results should be expected from a class. In other words, without access to data on the performance of similar pupils in other schools, it is exceedingly difficult for teachers to evaluate how effective they have been. This lack of clear, unambiguous, accurate information is typical of many of our complex social systems. Until recently this has been an inevitable situation but now the growth of computing has finally made the management of very large and complex data sets economically feasible.

We are moving into an era in which information can be made available in great detail at every level of complex systems. The impact of such information needs investigating, but first the quality of the information and its availability needs to be developed. In other words, we need monitoring systems. Whether we are looking at sentencing policies for juveniles, welfare policies, reward structures for directors of companies, productivity, truancy, satisfaction levels, health or school effectiveness, extensive monitoring is needed and is indeed developing in many areas of activity.

Few would set their face against the provision of information although we get some researchers more concerned with the second decimal place than with the method of collection of the data. What I am arguing here is that the growing understanding in the hard sciences of how complex systems function emphasizes the need for the flow of feedback into the system. With new insights provided by what is coming to be called 'complexity theory', our mental models of how the world functions need some adjustment. *In short, education is a complex, adaptive system which can probably become self-improving if given better data.*

Two Illustrations of Complexity

Since the complexity of education is not strongly questioned, just two illustrations will be given. One is an artefact: a list, drawn up by just one school for one lunchtime meeting, of factors which might be affecting their examination results. More than 100 factors were listed, most of which are shown in Table 4.1. None of these factors could be ruled out as unimportant for some students in some classes.

Table 4.1 *An Artefact: part of the list discussed by a Senior Management Team*
(Figures in the margin show that more than one person suggested this factor)

'Student' Factors
- Part-time jobs 11
- 'Unrealistic' students admitted to courses
- Unrealistic parental demands/expectations
- (Parents and pupils) lack of real understanding of what A level entails 3
- Too many extra-curricular activities 2
- Insufficient involvement in life of school/ community 2
- Students need to know what they are aiming for
- Casual approach after easy ride to 16+
- Marital/family break-ups
- Student absence 2
- Attitudes of students to school
- Drugs and alcohol
- Student responsibility
- Career/job prospects
- Instrumental approach to being a sixth-form student
- Student–teacher relationship
- Teachers' reaction to attendance & punctuality
- Social activities in sixth form
- Total demand made of students

'Teaching/Learning' Factors
- Teaching and learning strategies/range of styles (some outmoded) 3
- No structured teaching of study skills (e.g. note-taking) 5
- Quality of teaching
- Resources for teaching and learning (see separate list)
- Time for 1:1 discussions (teacher/student) 2
- Lack of 'bridging course' induction programme 3
- Inadequate concentration on *process* of learning
- Inconsistent expectations re marked written work
- Lack of exam practice/revision techniques
- Importance of homework in students' eyes
- Teaching approaches not responsive enough to needs of *all* students
- Teacher-pairs not communicating sufficiently
- Students not encouraged to read widely enough
- Student self-assessment
- Progress not reviewed often enough
- Number and use of study periods
- Lack of common deadlines within and between departments
- Teacher expectations of students

GCSE and the Transition to A Level
- Cumulative effects of GCSE on students' study habits
- GCSE as preparation for A level 3
- Transition to A level is difficult
- Transition to A level needs planning
- Good GCSE grades give false confidence to students
- Relevance of GCSE teaching strategies
- Students used to coursework demands dislike courses with final exam only
- Subjects seen as 'easy' or 'difficult'
- Choosing subjects which are inappropriate

Policy Factors
- Open-entry policy
- Negotiated time out 6
- Paired sharing of classes (= diminished teacher responsibility) 2
- Inadequate *range* of courses
- Poor/old-fashioned syllabus – choice which ignores changes at GCSE
- Too many A levels on offer – resources too thinly stretched
- Too many A levels on offer – poor *selection* of courses
- No credible alternative for the many students who can manage only two (or even one) A level
- Inadequate resourcing of vocational courses (e.g. low staffing priority) leads to too many choosing A level inappropriately
- Homework policy? 3
- Quality of information in Y11
- Quality of guidance in Y11, 12, 13
- Approach of subject choice at A level (several aspects of this noted, e.g. 'wrong' choices, inappropriate combinations, influence of teachers on subject choice) 5
- Lack of academic emphasis in school development
- Perceived management indifference to exam results

School Organization
- Structure of school day since 1990
- Use of registration and tutorial time
- Assemblies
- Leave of absence: timing and *use* of
- Inadequacy of previous pastoral system: inadequate monitoring
- Inadequacy of arrangements when staff miss lessons 4

- Amount or use of 'free' time
- Visits/fieldwork, etc.
- Ethos of school

Problems
- Too many interruptions to a 5-term course
- Too much time wasted between June and September
- We have given too much football during lesson time
- Staff turnover 2
- Retirements, resignations, illnesses, maternity leave
- Total sustained demands on staff time and energy in non-academic directions
- Instability: many changes since 1988
- Complacency: our results seemed OK
- General teacher demoralization
- Diversion of teachers'/departments'/school's energy towards 2
  - GCSE
  - National Curriculum
  - LMS
  - Pastoral reorganization
  - Appraisal
  - Parent's Charter
  - etc., etc.
- Lack of debate on A-level teaching
- Diversion of SMG attention from essentially curricular matters

Questions
- Is our 'performance' as implied by national statistics distorted by the proportion of those statistics which emanate from:
  - independent sector (massively better resourced)?
  - sixth-form college sector: not distracted by KS3 and KS4 planning?
- Are we losing more quality sixth-form students:
  - from the state education altogether?
  - to neighbouring schools/colleges?

Resources and Facilities
- Teaching environment 2
- Availability (or not) of study areas especially silent areas and supervised areas 6
- General decline in funding for teaching materials – inadequate resourcing 3
- Books get (and look) older
- Poor, largely book-based, study facilities (library stock; IT resources and facilities) 2

The second illustration of this assertion of the complexity of the system with which we are dealing is due to Ibrahim (1992) and based on data from the A-Level Information System (ALIS), supported by many hours of classroom observation and interviews with teachers and pupils. In Figure 4.2, the Y-axis represents residuals (what is left over after we have taken account of the best single predictor of achievement at A level – the average achievement score two years earlier). Positive residuals indicate performance above that which would be expected on the basis of prior achievement, negative residuals represent achievement below what would be expected on the basis of the prior-achievement measure. Ibrahim divided the sample into high and low ability according to their scores on the International Test of Developed Abilities (Ottobre and Turnbull, 1987). As can be seen from the graph, the amount of individual help given by the teacher was associated with quite different outcomes, according to the ability category of the pupils. More able pupils appeared to do better when not receiving individual help, and less able pupils seemed to need an extensive amount of individual help to achieve at a level consistent with their prior achievement. Interestingly, in discussion, a number of able pupils stated explicitly that, in mathematics, if they could work it out, they liked to work it out for themselves rather than seek help from the teacher because that way they understood it better and remembered it better (Ibrahim, 1992, p. 298). This is correlational data so the direction of cause (if cause there is) cannot be assumed from the data. It serves to illustrate, however, that teaching is directed at individuals and different individuals may need different styles of teaching. Furthermore each action will have consequences and reactions. Giving help

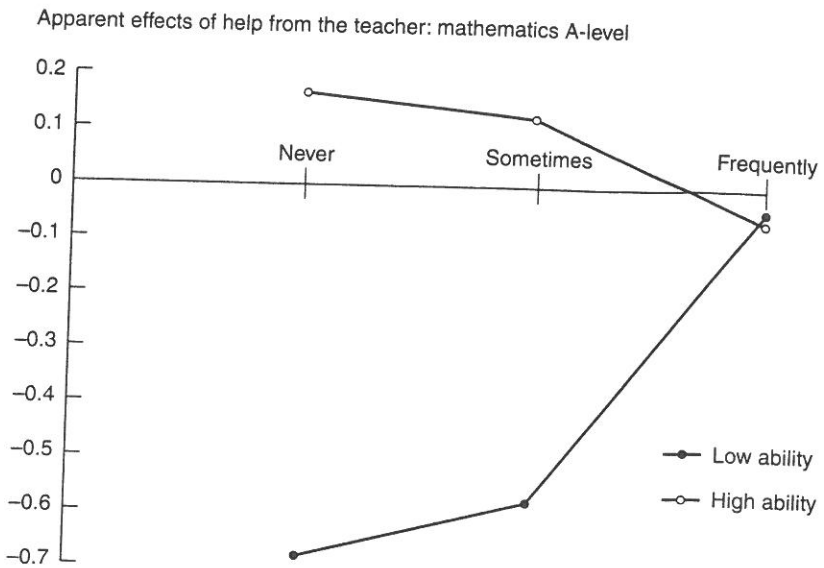

Figure 4.2 *Residuals associated with various amounts of help from the teacher (after Ibrahim, 1992)*

contains the message that help is needed – a message of incompetence. Help can be harmful. Whether or not it is may depend upon many subtle factors of interpretation, style of help, personalities. Here, in Ibrahim's work, we see the effect of help *apparently* dependent upon the ability of the student.

In summary, even if we confine our attention to just one simple, central outcome of education, namely pupils' cognitive achievement, the numerous factors affecting this outcome, and the contingencies in patterns of interactions, are possibly so locally determined as to be *un*amenable to much in the way of general rules which say 'this is good practice'. The specification of good practice may be as difficult to catch hold of as the end of the rainbow.

It is not just the presence of many factors which makes the discovery of rules for action unlikely. If the factors acted consistently, in a linear, additive fashion, and could all be measured, the mathematical modelling and the normal procedures for establishing 'research has shown'-type findings could yield results. We might manage to 'explain' more variance simply by measuring more factors. It is the feedback loops (the non-linearity) in the system which make for predictable unpredictability.

COMPLEXITY – HOW DO WE RESPOND?

How then, if we cannot pursue research as normal and locate the principles of good practice and pass on these rules to teachers, can we improve the complex system of education?

The question that we must ask in pursuit of school improvement is how do complex adaptive systems become most effective? Not an easy question, but it is the essential question at the heart of trying to influence a social system as complex as education. To answer this question requires that we return to the sea changes taking place in science and social science disciplines, driven forward by the fascination with computers as a tool for the simulation of complex information processing.

Perhaps the first thing to learn is to think in terms of systems rather than picking out pieces of a system. To think in terms of getting the whole system to work, not all at once, not in some distant future when we finally understand some fundamental laws. What can we do now, for the entire educational system, to improve its effectiveness?

Emergence

One relevant construct is that of emergence – the emergence of complexity and order. Many of the studies in complexity have been concerned with the nature of life, its characteristics and its emergence. The work is getting very close to explanations as to how disordered systems can become self-organizing in local areas, using local laws with feedback, selection, adaptation, conflict and cooperation. The emergence of organization when there are simple local laws is a highly important concept for it suggests that complex systems *are* locally organized and perhaps *have to be* locally organized, with their sub-units making independent decisions, not governed by some Great Plan from higher in the system. What do these ideas suggest about management? The ideas seem to fit well with LMS – the local management of schools which has devolved large proportions of LEA budgets to schools. Whatever the problems, there is a widely held impression that local management has been strongly welcomed, at least by secondary schools and colleges. Classrooms have always enjoyed a good deal of independence – a situation sometimes bemoaned by management but, it would seem from complexity theories, highly important.

The Flow of Information

What, then *is* the role of management? As Davies (1987) says, complex systems are 'describable by a web of informational interactions' (p. 159). If schools evolve like organisms they do so in response to information about their students, their potential students, external requirements and internal information about resources, human and material. Self-organizing units can only respond effectively if they have the necessary information. Management must provide the information infrastructure which enables local problem-solving to be effective.

What if disinformation is introduced? Presumably this can distract attention and effort from more valid information which could have led to more useful activities. If inspectors, for example, declare that teachers' expectations are too low, what are teachers to do about this? Feel bad? Does this help? After 25 years of developing this deeply flawed research (Elashoff and Snow, 1971) we are no wiser as to its implications for action and recent work continues to cast doubt on its validity (Goldenberg, 1992; Raudenbush, 1984).

If inspectors demand written development plans (in this unpredictable world) does this lead to endless hours of writing and discussion? Is there any evidence that written development plans are beneficial for any activity other than satisfying inspectors? These issues need evidence. Inspection itself, and the efforts schools put into responding to inspection, represent enormous amounts of time and effort diverted from teaching. Were this misdirected it would have a negative cost–benefit ratio. One task of management for improvement is to remove or discount poor information.

Chapter 4

Monitoring School Effectiveness: Simplicity and Complexity

Carol Fitz-Gibbon

INTRODUCTION

Educational research has been too ready to leap from correlation to causation, too ready to neglect, after the epidemiology, the necessary clinical trials. The hope has been that finding 'the correlates of effectiveness', i.e. features common to 'effective schools', would enable researchers to guide 'ineffective schools' towards improvement (Figure 4.1a).

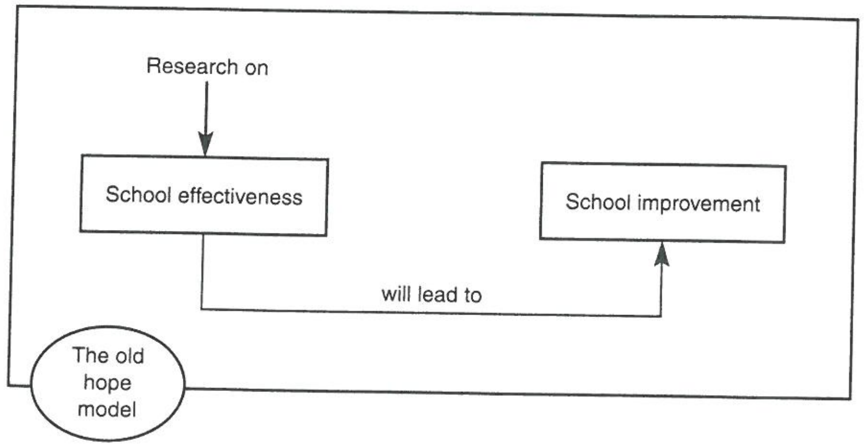

Figure 4.1a *Effectiveness focus*

But correlations only generate hypotheses which then need testing, ideally by well-designed multi-site field trials, the sort of replicated experiments which provided evidence for the effectiveness of pre-school education in highly deprived inner cities in the USA (Lazar and Darlington, 1982). Furthermore a correlate such as 'a safe and orderly environment' is of little use. There are few campaigners for unsafe and disorderly environments and the question which faces schools is how to create the safe and orderly environment. What actions are needed? This requires experimental efforts at school improvement. Indeed, it is likely that careful monitoring of school improvement efforts would be a faster route to a knowledge of effective actions than further surveys attempting to find correlates of effectiveness by comparing schools deemed effective or ineffective (Figure 4.1b). To know which actions work it is actions which must be studied.

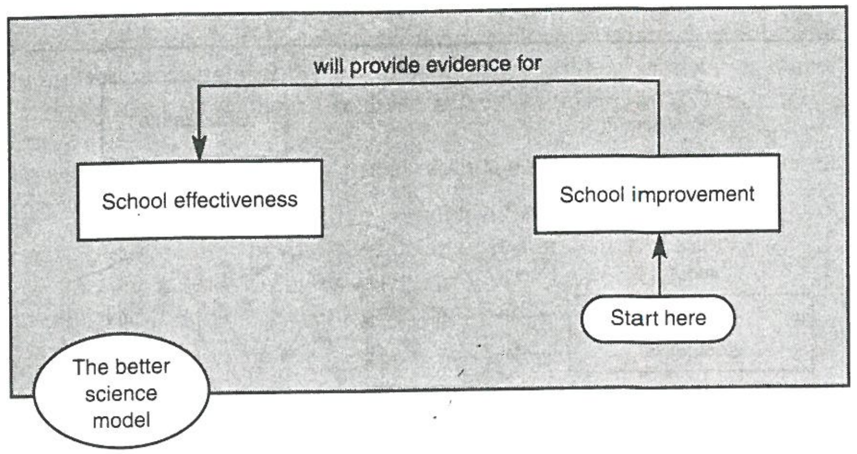

Figure 4.1b *Improvement focus*

Yet either approach – testing and validating hypotheses or collecting evidence of effective school improvement efforts – would take years of research to arrive at reliable, generalizable findings and would require a level of funding which is unlikely. Furthermore after adopting this approach of 'normal science', i.e. pursuing valid research strategies to find strong nomothetic rules, there would then follow the problems of dissemination and uptake, both of them fraught and lengthy processes. Set against this long time horizon we have to recognize a sense of urgency which is based on perceptions of serious problems in schools (particularly in the USA more so than in the UK). If children feel afraid, fail to learn, stay away from school, disrupt when they are in school, then teachers cannot wait for validated research findings to seek improvements; they cannot wait for evaluators to find the dream programme which applies to their particular patch of the educational pastiche.

Another major problem with the 'science as normal' RD&D (research, development and dissemination) approach to school effectiveness is more fundamental: there may be very few nomothetic rules or generalizable findings, very few 'laws' which are locally applicable. This is not to assert that phenomenology rules and we must all retreat into qualitative studies along with our one-legged friends, the illogical negativists. The point here is that we may be dealing with a system which is so complex as to be fundamentally unpredictable in a way which is only just beginning to be understood in science proper. If this is the case we have to rethink how to approach the vital task of school improvement.

THE EMERGING SCIENCE OF COMPLEXITY

Science itself is currently undergoing a sea change, possibly even a paradigm shift. P. C. W. Davies, the famous cosmologist and professor of physics has described the change as 'nothing less than a brand new start in the description of nature' (Davies, 1987, p. 23). In the next few paragraphs an attempt is made to trace the ideas which have led to the present mix of models and beliefs reflected in Figure 4.1c.

Prediction and Feedback

> An adaptive agent is constantly playing a game with its environment. ... What actually has to happen to game-playing agents to survive and prosper? Two things ... prediction and feedback (Holland, quoted by Waldrop, 1992, p. 282).

In order to make effective decisions for school improvement we need values (to know in which directions we wish to move) and information. To thrive on the edge of chaos, the system needs the capacity to act in multiple self-organizing, adequately informed, local units, i.e. the capacity to act on the basis of feedback. (Indeed the Hawthorne Effect has been reanalysed to show it to have been largely due to feedback (Parsons, 1974).)

Monitoring systems enable schools to make accurate predictions where these are possible and provide a constant flow of feedback. Thus monitoring systems fit in with the model of a school as an evolving complex system.

SIMPLICITY – THE NEED TO KNOW THE MAJOR PATTERNS IN DATA

Although it has been argued above that the educational enterprise is exceedingly complex it can also be shown that stable features can be derived from monitoring data. Just as the *weather* is variable and unpredictable – a frequently cited example of a non-linear, open system, determined by the laws of physics but essentially unpredictable – and yet the *climate* is a fairly reliable guide to the range of weather to be expected, so there are broad features in the system which adequate monitoring can clarify and, by so doing, can inform the interpretation of data and the feedback into the system.

The major point which needs emphasis, and which it is the intention in this section to illustrate, is the extent to which *mistakes can be made if the broad features which pertain throughout a system are not known and/or are not taken into account*. In other words there will be variations in outcomes which are seen throughout the system ('noise') and there will be some stable broad features. Four examples will be drawn from the ALIS monitoring system.

Are some teachers to blame for inaccurate predictions? David Elsom, a former college vice-principal, draws on his experience to portray a scenario which he has frequently encountered: complaints against the English department for being unable to predict accurately their pupils' A-level grades. Anyone thinking to blame individual English department staff needs to recognize that the prediction of grades in English is a problem throughout the system and may have more to do with the nature of the A-level English examination, and its lack of discrimination, than with the competence or otherwise of individual A-level English teachers. Appeals against grades are more frequent in English than in other subjects.

This is an important illustration of two principles: that the broad features of the system must be known before individuals are judged and that the competencies of teachers can only be considered in comparison with similar teachers *teaching the same subject* to similar pupils.

Subject differences are substantial and should not be neglected. Indeed *school effectiveness may need to be reconceptualized, if not for every subject*

then certainly for broad bands of subjects such as the foreign languages, the sciences, mathematics, the humanities, practical or vocational subjects. What applies in one curriculum area may not apply in another.

Can examination results be fairly contextualized by SES data? Rarely is socio-economic status a good predictor of achievement in the UK – which is perhaps a tribute to the UK system of education. SES is not the most important covariate so that comparisons which rely on SES comparisons will be less fair than comparisons of pupils' achievements based on cognitive measures. In general SES may explain about 9 per cent of the variance whereas any prior achievement measure would generally explain about four times as much. Furthermore, within schools the composition of classes varies from year to year and this variation will itself vary from school to school depending perhaps on the 'pulling power' of various departments (Fitz-Gibbon, 1984). If one inner-city French department had recruited all the highly able students it would be an error to adjust its results for the average SES of the school, or even the average achievement level of the school.

Are A-level subjects 'level' in difficulty? Rhodes Robson is an accountancy firm. It has written to schools offering to work out residuals for A-level subjects. How does it compute the predicted/expected grade for a student? By assuming that all subjects are equally difficult at A level and therefore if your student has 24 UCAS points the 'predicted' grade in each of three subjects can be taken as 24/3 or 8. The fact is that A-level subjects differ in difficulty on any reasonable definition (Fitz-Gibbon, 1988). Consequently the approach suggested will lead to unfair chastisement of teachers teaching difficult A levels: maths, sciences and foreign languages. Do trading standards apply in social science?

Table 4.2 shows subjects rank ordered according to their intercept as calculated using ordinary least squares (OLS) regression. We see physics as having the lowest intercept. It appears to have been, that year, the most difficult subject and history was at the other end of the rank order appearing to have been the easiest subject in the 1989 sample. A second method of calculating difficulty (Kelly–Lawley adjustments) gave similar results (Kelly, 1976; Fitz-Gibbon, 1991).

Table 4.2 *The relative difficulty of the 1989 A-level examinations*

Subject	*Intercept from OLS regression*[1]	*Difficulty rank OLS*	*Difficulty rank Lawley*[2]	*Lawley correction factors*	*N*
Physics	–9.8	1	1	0.58	867
French	–8.5	2	5	0.12	371
Chemistry	–8.0	3	4	0.28	853
Maths	–7.4	4	3	0.31	1,357
General Studies	–7.4	5	2	0.36	1,087
Biology	–6.7	6	9	–0.54	667
Geography	–6.0	7	8	–0.66	630
Economics	–5.7	8	7	–0.18	606
English	–4.6	9	10	–0.75	831
History	–4.1	10	6	–0.15	674

[1] Based on 1989 ALIS data.
[2] Based on the subset of candidates taking 2 or more subjects.

There is no excuse for the use of untenable assumptions demonstrated by the accountancy firm. The differences illustrated here have been a consistent feature of the data for years and apply to Scotland's Highers as well as to English/Welsh A levels. If these differences in difficulty are not (a) recognized and (b) monitored accurately from year to year then unfair comparisons are likely to be made in schools and colleges trying to assess the performance of departments by means of comparing the grades pupils achieve in one subject as compared with another. Such errors can demoralize teachers or even, in a harsher world, cost them their jobs.

Are departments differentially effective to a substantive degree? Note that the question posed above is not 'Are departments differentially effective to a *statistically significant* degree?' The sizes of differences which are important to schools cannot be determined *a priori* by reference to levels of statistical significance habitually employed in other studies, but will come to be understood over the years as people work with the data. Statisticians should not mindlessly adopt significance testing in novel situations without justifying their weighting of Type I and Type II errors. (A touching faith in the 0.05 level is not science but habit.)

We can use regression lines to look at differences in effectiveness and try to get some handle on the *substantive* difference that effective and ineffective departments make to A-level grades. Departments can be rank ordered on the average residual obtained by the pupils on the examination in a particular subject, for example, biology. If we select, from this rank-ordered league table of average residuals, the top quarter and the bottom quarter and plot the regression lines, we get the graph shown in Figure 4.3. There appears to be a difference of more than one and a half grades across the range of GCSE scores. Thus, for a given average GCSE score, those pupils who were in departments in the top quarter

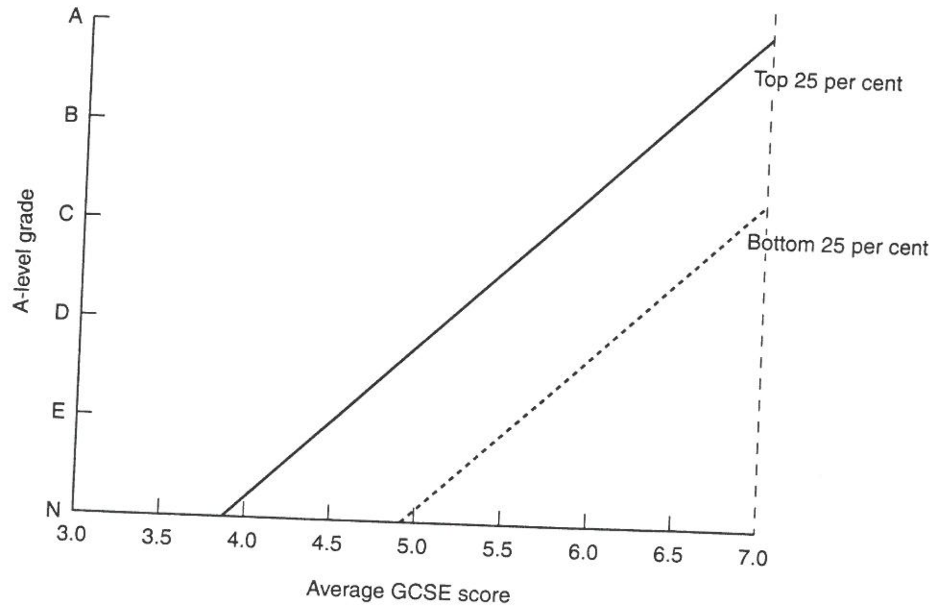

Figure 4.3 *Biology 1992: regression lines for top and bottom 25 per cent of institutions in ALIS*

generally obtained substantially better grades in biology than pupils of similar prior achievement in departments in the bottom quarter. What are we to make of these differences? Are they just noise or can they be altered? Only repeated efforts to alter the residuals will inform us as to whether they are alterable or just noise or features of particular student bodies, or schools, in ways that we have not measured. In other words, we can only establish what is or is not alterable by monitoring the outcomes of attempts at alteration, preferably in the framework of an experimental design (Campbell and Stanley, 1966). As Willms (1992) has pointed out, experimentation is greatly facilitated by being set up in the framework of a monitoring system.

SUMMARY AND PROPOSALS

We argue frequently by metaphor and with underlying mental models. I have tried to give some flavour of the sea change taking place in science as physicists, mathematicians and those who can follow in their footsteps tackle not the simple, predictable systems of stars and galaxies (not so simple and not always predictable), but complex systems of economies, living organisms and the complexities of organizations. One can certainly draw the conclusion from much recent work on complexity that the number of connections in a system (who talks to whom, where information flows, how many people receive the information, the amount and quality of information) have to be crucial variables in the functioning of any system, and are probably the variables which should be manipulated in order to improve the functioning of the system, although in what direction and by how much will be a matter of trial-and-error learning. These are not answers we can dream up; we have to run the programs and see what happens.

As already noted, current writing on complex systems seems to pay no attention to the effects of feedback which is false. I would like to suggest the hypothesis that an important way to improve education is to *increase the amount of valid feedback and to decrease the amount of misleading feedback; increase fair comparisons and decrease disinformation such as the overinterpreted generalizations and opinions offered by inspection and the disinformation of inadequate models, illustrated by the offer from an accountancy firm.*

The validity of the judgements made by inspectors has been called into question many times (Bennett, 1978; Gray and Hannon, 1986; Fitz-Gibbon, 1994) yet it appears that the system which has operated for more than 100 years has made no effort to have independent checks made on its reliability or validity, let alone value for money.

In this chapter an attempt has been made to show that monitoring in a complex system like post-16 education can draw attention to some broad, simplifying features in the data and that lack of knowledge of these features can leave people open to unjust criticisms and can mislead people into trying to take actions on the basis of inadequate information. Misinformation can cause pain and can distract the system from working towards the outcomes it values.

The setting up of systems of *monitoring with feedback* is the most vital task of the next decade – and we have only just begun. Whilst we must increase the amount and quality of feedback, all the while monitoring the effect of feedback – monitoring the monitoring – we also need to be concerned to eliminate from the system false feedback, misinformation, distraction from the critical task of focusing on outcomes.

The questions, then, are:

- What are the outcomes of concern? What do we value enough to measure? What kind of feedback is required? Which variables? How measured? And how frequently measured? And which covariates are needed to make fair comparisons? (i.e. what predicts the outcomes of concern?).
- How should feedback be provided? What level of aggregation should be used? What degree of confidentiality must be provided at every level of the system? (i.e. who gets to know what?) What kind of feedback is understood with or without additional training?
- Are there process variables which relate to the outcomes of concern? i.e. are there 'alterable variables', things people might choose to do which might make a difference and should therefore be monitored? Can these be measured without raising false hopes of easy answers or risking widespread misinterpretation of correlation as causation?

Thus may we improve the flow of valid information into the system and integrate the old-style search for simplicity with the newly developing mental models of complex systems. We can set up monitoring systems which have a role in exposing both the complexity of the system and its simplicity.

Note

1. The A-Level Information System, the ALIS project, has been based on collaboration between researchers (previously at Newcastle University and now at Durham) and schools and colleges, supported in many instances by local education authorities and training and enterprise councils. Data are collected relating to:

- pupils' achievement in examinations at age 16 (General Certificate of Secondary Education (GCSE)) and at age 18 (A levels);
- pupils' attitudes to each subject studied for the examinations at age 18;
- pupils' attitudes to their school or college and the facilities and resources provided;
- pupils' aspirations for further education;
- pupils' participation in extramural activities (as a 'quality of life' indicator);
- demographic factors;
- process data in the form of classroom teaching and learning activities, and school organizational factors.

The data are specially collected, by representatives of the university directly from pupils, schools and colleges. Reports are prepared for each A-level subject department and fed back promptly to each department.

REFERENCES

Bennett, N. (1978) 'Surveyed from a shaky base', *The Times Educational Supplement*.

Campbell, D. T. and Stanley, J. C. (1966) *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Davies, P. C. W. (1987) *The Cosmic Blueprint*. London: Heinemann.

Elashoff, J. D. and Snow, R. E. (1971) *Pygmalion Reconsidered*. New York: Wadsworth.

Fitz-Gibbon, C. T. (1984) *Report to Schools: Confidential Measurement-Based, Self-Evaluation*. Newcastle upon Tyne: School of Education.

Fitz-Gibbon, C. T. (1988) 'Recalculating the standard', *The Times Educational Supplement*, **15**, 26 August.

Fitz-Gibbon, C. T. (1991) *Evaluation of school perfomance in public examinations: a report for the Scottish Office Education Department*. Newcastle upon Tyne: Curriculum, Evaluation and Management Centre.

Fitz-Gibbon, C. T. (1994) *Ofsted, Schmofsted*. Newcastle upon Tyne: Curriculum, Evaluation and Management Centre.

Glass, G. V. (1979) 'Policy for the unpredictable (uncertainty research and policy)', *Educational Researcher*, **8** (9), 12–14.

Goldenberg, C. (1992) 'The limits of expectations: a case for case knowledge about teacher expectancy effects', *American Educational Research Journal*, **229** (3), 517–44.

Gray, J. and Hannon, V. (1986) 'HMI interpretation of schools' examination results', *Journal of Educational Policy*, **1** (1), 23–33.

Ibrahim, A. bin (1992) *The A-Level Examination: Qualitative and Quantitative Data in the context of a Performance Monitoring System*. Ph.D. Thesis, University of Newcastle upon Tyne.

Kelly, A. (1976) 'A study of the comparability of external examinations in different subjects', *Research in Education*, (16), 50–63.

Lazar, I. and Darlington, R. (1982) *Monographs of the Society for Research in Child Development: Lasting Effects of Early Education*. Chicago: University of Chicago Press.

Ottobre, F. M. and Turnbull, W. W. (1987) *The International Test for Developed Abilities: A report on the feasibility study*. Princeton, NJ: Report for the International Association for Educational Assessment.

Parsons, H. M. (1974) 'What happened at Hawthorne', *Science* (183), 922–32.

Raudenbush, S. (1984) 'Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of the expectancy induction: a synthesis of findings from 18 experiments', *Journal of Educational Psychology*, **76** (1), 85–97.

Tymms, P. B. (1994) *Theories, Models and Simulation: School Effectiveness at an Impasse*. Paper presented at the ESRC Seminar Series (see Chapter 6).

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

Willms, J. D. (1992) *Monitoring School Performance: A Guide for Educators*. Lewes: Falmer Press.

Part 2

Some Recent Evidence from the Field