'TOUGH, INTELLIGENT ACCOUNTABILTY' IN SCOTTISH SECONDARY SCHOOLS AND THE ROLE OF STANDARD TABLES AND CHARTS (STACS): A CRITICAL APPRAISAL

MICHAEL COWIE, DAVE TAYLOR AND LINDA CROXFORD

ABSTRACT

'Systems of tough, intelligent accountability' are heralded by the recent policy document Ambitious Excellent Schools (Scottish Executive, 2004a). This paper considers what 'tough, intelligent accountability' might mean, and how this might differ from systems that are currently in place in Scottish secondary schools. It focuses on the role of the Standard Tables and Charts (STACS) in current accountability systems. STACS are derived from National Qualifications data for each Scottish secondary school, to compare the performance of each subject in the school and to analyse performance in the attainment measures within National Priorities. They are used in performance reviews and inspections to hold schools and teachers to account for students' performance. The paper considers whether STACS can be considered a form of 'intelligent accountability' or whether perverse results are intrinsic to such statistical systems. It further considers how the accountability system could be made more intelligent.

INTRODUCTION

'Systems of tough, intelligent accountability' are heralded by the policy document Ambitious Excellent Schools (SEED, 2004a). But, what constitutes 'intelligent accountability' and how do the adjectives 'tough' and 'intelligent' relate to current accountability systems in Scottish education?

In this paper we discuss the emergence of the concept of 'intelligent accountability' and consider its implications for the systems of accountability currently in place in Scottish schools. We argue that the appropriation of 'tough, intelligent accountability' situates the concept within a continuing attempt to establish a culture of performance and performativity in Scotland. This is associated with a set of assumptions about how performance data can or should be used to improve attainment, and we argue that this has changed or is changing the governance of schooling in Scotland. This approach may be tough but it is 'unintelligent' because it subjugates teachers and limits their professional autonomy.

Performance data invite comparisons, and while some may see benchmarking as a stimulus to improvement, performativity can also be seen in terms of a technology of control. Over 20 years ago, for example, Lyotard (1984:46) argued that:

> Performativity is a culture or a system of 'terror' that employs judgments, comparisons and displays as means of control, attrition and change. The performances (of individual subjects or organizations) serve as measures of productivity or output, or displays of 'quality' or 'moments' of promotion or inspection. They stand for, encapsulate or represent the worth, quality or value of an individual or organization within a field of judgment. An equation between wealth, efficiency and truth is then established.

What we have seen in Scotland over the past 20 years is the incremental development of fundamental change in the way that secondary school performance data have been constructed and used with associated change in the way in which the educational system is organised (Cowie and Croxford, 2006). What is not clear, however, is the extent to which change has occurred within schools. If the Scottish experience

is similar to that in England, where performance management does not appear to be improving the quality of learning and teaching (Gleeson and Husbands, 2003), perhaps there has been little fundamental change[1].

In Scotland, the STACS provided by the Scottish Executive Education Department (SEED) are a major part of the mechanism for holding teachers and schools to account in the secondary sector, and the key source of data for school and authority self-evaluation of attainment in public examinations. In this paper we take a closer look at the STACS, and focus particularly on the influence of the Relative Values (RV). The RV compare the performance in each subject department in a school against that of other subject departments in the school. This, therefore, is a measure where some evidence of improvement may be expected. If the provision of RV influences management policies in secondary schools and contributes to change in departmental and classroom practices and improved attainment in lower attaining subjects, then evidence of improvement in relation to the other subjects should be reflected in the RV over time.

First, however, we discuss the nature and meaning of 'intelligent accountability' and outline the respective roles and responsibilities at three levels in the Scottish educational system with regard to accountability. Then we outline, discuss and critique the use of the STACS before concluding with a discussion of the relationship between the STACS and the concept of 'tough, intelligent accountability'. The paper is part of a collaborative research study that is being carried out as part of the Applied Educational Research Scheme (AERS) School Management and Governance Network.

THE CONCEPT OF INTELLIGENT ACCOUNTABILITY

Accountability is a significant concept in education because it is central to an understanding of the relationship between schools, education authorities and central government. At least two parties are involved in an accountability relationship, with an expectation that one will behave in ways that are consistent with the legitimate requirements of the other (Simkins, 1997). One has therefore the potential to modify the actions of the other. In education, however, there are multiple accountability relationships and these are not clear cut. To whom are schools accountable, for what and for what reasons?

Schools and education authorities are staffed by people with certain expectations with regard to autonomy, consultation and participation, and they are accountable in a number of different ways to different groups. Our underlying concern in this paper is that the articulation of the need for systems of 'tough, intelligent accountability' is likely to establish, or at least help to confirm, the nature of a particular kind of relationship between schools, education authorities and the Scottish Executive. We say this because accountability is usually defined in terms of control and giving an account to those in authority. In its pure form accountability is tied up with power and control because if the account given by the accountable body is unsatisfactory then some form of constraint or sanction may be expected.

But there are other interpretations of the meaning of accountability and several models and accountability frameworks have been developed in educational systems (Adelman and Alexander, 1982; Kogan, 1986; Simkins, 1997; Halstead, 1994). What is clear in these models and frameworks is that there are substantial difference between a definition which involves a right to exercise sanctions and one focused on rights to information and duties to report. People understand different things by accountability, and one of the reasons why accountability is such a complex concept is that different value systems are involved with tension between them.

The idea of 'intelligent accountability' came to the attention of educators in 2002, when Onora O'Neill's Reith Lectures on 'A Question of Trust' explored the negative effects of the accountability culture. O'Neill (2002) argued that:

> The new accountability is widely experienced not just as changing... but distorting the proper aims of professional practice and indeed as damaging professional pride and integrity.

She went on to suggest that 'if we want greater accountability without damaging professional performance we need intelligent accountability' and that this requires 'more attention to good governance and fewer fantasies about total control'. She also argued that much that has to be accounted for is not easily measured and cannot be boiled down to a set of stock performance indicators.

Building on these ideas, the need for intelligent accountability in school education was explored further by the Secondary Heads Association (SHA) in England. SHA defined the concept as:

> a framework to ensure that schools work effectively and efficiently towards both the common good and the fullest development of their pupils. It uses a rich set of data that gives full expression to the strengths and weaknesses of the school in fulfilling the potential of pupils. It combines internal school processes with levels of external monitoring appropriate to the state of development of each individual school (SHA, 2003:8).

SHA's suggestions for moving towards intelligent accountability include removing performance indicators that give perverse incentives, removing top-down target-setting, ending publication of performance tables, a less prescriptive curriculum, reducing high-stakes testing, and a move to greater school self-evaluation with consequent reduction in external inspection. To some extent, the concept of intelligent accountability was endorsed for English education by the minister of state for school standards who agreed with the increasing focus on school self evaluation, to be supported by shorter sharper external inspection. However, he reaffirmed commitment to national tests and exams as a basis for the accountability mechanism:

> Test and exam results are a vital indicator of educational success, and when the rigour they bring is lost, the evidence from the history of the English education system is that it is the pupils in the poorest communities who suffer most (Milliband, 2004).

The discovery of 'intelligent accountability' in Scotland seems to have coincided with that in England, but the policy document Ambitious Excellent Schools did not refer to the English policy process[2]. Nor did it acknowledge the negative effects of an accountability culture featuring 'toughness' (Cowie and Croxford, 2006). Although an emphasis on accountability has been a recognized feature of educational policy in Scotland for some time, the introduction of the terms 'tough' and 'intelligent' is relatively new. The document which attaches these adjectives to 'accountability' does not explain what is meant by 'systems of tough, intelligent accountability':

> Delivering excellence in education requires both professional freedom and public accountability. Scotland already has a world renowned system of inspection and evaluation: we will build on this to ensure further, sustained improvement in our schools. We need systems that are proportionate, that focus on outcomes, that promote self evaluation, and that provide targeted support to those who are struggling… We will expect local authorities to drive improvement at the local level… We will expect schools to meet the needs of their community and each and every one of their pupils. We will act to build, at each level, systems of tough, intelligent accountability that foster ambition and allow proper, informed public scrutiny' (SEED, 2004a).

From this statement, it is not easy to determine what features we might expect in a system of 'tough, intelligent accountability' in Scottish education. One difficulty

is that the term itself may be an oxymoron with 'tough' suggesting a hard and uncompromising approach to accountability with associated punitive measures, and 'intelligent' denoting a more considered, thoughtful approach based on a range of different accountability models. If we focus on what characterizes 'intelligent' accountability, the argument presented by O'Neill (2002) and the SHA (2003) suggests that the following features should be included:

- trust and professionalism;

- focus on self-evaluation, with a minimum of external inspection targeted to where it may be most needed;

- measures of performance that cover appropriate outcomes of schooling – not just academic attainment - and do not distort the purposes of education;

- measures that encourage the fullest development of every pupil.

Within Ambitious Excellent Schools (SEED, 2004a), however, the 'tough, intelligent accountabilities' outlined are limited to:

- testing Scotland's performance internationally;

- integrated, proportionate inspection to support improvement and target action;

- support and challenge for local authorities to deliver continuous improvement.

No mention is made of the other aspects of 'intelligent accountability' that might have been anticipated from O'Neill's initial meaning, or the SHA's practical suggestions, or the 'softer' measures included in the National Priorities. If the term is to have more substance than a political 'sound bite' it is reasonable to expect consideration to be given to the links between accountability and improvement, with some practical demonstration that there have been significant changes from the top-down target-setting model introduced in 1997.

ACCOUNTABILTY IN SCOTTISH EDUCATION

In Scottish education, as in other public services in Britain, there has been a growing emphasis on performance management, quality assurance and accountability over the past two decades (Cowie and Croxford, 2006). The driving force behind this accountability movement has been Her Majesty's Inspectorate of Education (HMIe). The structure was formalised as the Quality Initiative in Scottish Schools (SOEID, 1997), with aims (1) to evaluate and systematically account for quality of educational provision, and (2) support quality assurance developments in order to raise standards and improve quality.

The Standards in Scotland's Schools Act 2000 (Scottish Parliament, 2000) made this accountability system a statutory obligation. The obligations placed on schools, education authorities and HMIe are outlined below.

SCHOOLS

Schools are required to perform the following tasks each year:

1 Evaluate their own performance using performance indicators. A key document is How Good is Our School? Self-evaluation using quality indicators (SEED, 2002). This requires schools to look at each aspect of provision and ask: How are we doing? How do we know? What are we going to do now? A set of quality indicators are defined and suggestions are given as to the evidence needed. For secondary schools a major source of evidence on attainment is

provided by STACS, which are described in detail in a later section.

2 Write a Standards and Quality report and a School Development Plan (sometimes referred to as a School Improvement Plan, thereby reflecting the improvement agenda). Schools are expected to: give details of self-evaluation; recognize key strengths; identify levels of service to be maintained; identify development needs; and set targets. The Standards and Quality Report must therefore include developments that have taken place over the previous session and outline proposed developments for the forthcoming session. It must also include pupils' attainment data in the context of the school's attainment targets (which must be agreed with the education authority). The School Improvement Plan is expected to draw from the National Priorities and from national initiatives (such as A Curriculum for Excellence) as reflected in the education authority's priorities and Improvement Plan for education.

3 Publish a school handbook for parents. The handbook sets out key statistics about the school, as well as key policies. For secondary schools these include indicators of attainment in National Qualifications (NQ).

EDUCATION AUTHORITIES (EAS)

EAs are required to identify and take action to continuously improve performance in schools. Quality Improvement Officers (QIO) within the EA 'challenge and support' schools for which they have responsibility. They scrutinise statistics on school performance, ensure a robust self-evaluation structure within schools, and identify areas that need to be addressed. They carry out a regular cycle of visits to schools to: assess the school's progress with its school development plan; discuss improvement issues with management and staff; and support the school's management in making improvement. At the start of the school session, the focus in secondary schools is on results in external examinations – the NQ3. Visits in the latter half of each session focus on issues for the school development plans for the following session and the QIO is required to scrutinise and approve development plans to ensure delivery of local and national priorities. Additionally, most authorities conduct Performance Review meetings annually in both primary and secondary schools. These meetings examine the school's progress in terms of attainment data, self-evaluation procedures, and the key points for action in the national and local improvement agendas.

EAs are required to prepare an annual statement of improvement objectives in the areas of national priority identified by Scottish Ministers. The statement of objectives is intended to give local effect to national priorities and show how the authority will implement each priority taking account of local circumstances and the views of parents, pupils and others with an interest.

EAs themselves are inspected by HMIe, with especial regard to their own self-evaluation and the extent to which they support and challenge their schools. National Government has effectively maintained control of the measures of 'success' in improving schools, but devolved the responsibility to Local Government (Cowie and Croxford, 2006).

HMIE

HMIe is an Executive Agency of the Scottish Executive which provides independent external inspection and evaluation of the quality of education. In undertaking a school inspection the inspection team uses the quality indicators from How good is our school? (HGIOS) as the basis of its evaluations. The HMIe assessment scale of 1 to 4 (with 4 representing Very Good) has been superseded by a 1 to 6 scale. 'Very Good' is no longer regarded as the pinnacle to which schools should aspire,

having been replaced by 'Excellent'. The Executive's expressed aim is for schools, in their on-going striving for improvement, to 'journey towards excellence', and this is reflected in the title of the document currently being published in installments and which ultimately will replace HGIOS.

The school's self-evaluation is key to the tone of HMIe reports. Previously, attainment on HMIe defined measures was likely to provide the critical determinant of HMIe judgment, while recent reports have sometimes glossed over poor attainment, if good self-evaluation procedures demonstrate that the school has the 'capacity for improvement'. This may suggest a more 'intelligent' accountability but the HGIOS quality indicators are themselves defined by HMIe and we discuss the implications of this later.

A major source of information for inspection of secondary schools is data on attainment in NQ. During the 1990s, accountability was defined by Government in terms of attainment through the publication of examination results, with a school's effectiveness being seen in relation to the performance of other schools. The introduction of target-setting in 1998, taking account of factors that influence attainment but are outwith the control of schools, acknowledged that some schools face greater challenges than others when attempting to raise the attainment of pupils. Free Meal Entitlement (FME) was taken as a proxy for deprivation, and the proposed attainment target for each school was the performance of the best in a group of 'similar' but better performing schools. The crudity of this measure has since been recognized and a more sophisticated measure based on a principal components analysis of a number of measures[4] is now used to compare school performance in NQ in groups of comparator schools.

However, this approach to evaluating school performance is fundamentally flawed because it is based on proxy measures of school context, does not use pupil-level data, and lacks statistical validity (Croxford, 1999). Research on school effectiveness shows that evaluation of pupils' progress using a value-added approach which takes account of pupils' attainment on entry to school is a more appropriate way for comparing the performance of schools. (Goldstein, 1984; Woodhouse and Goldstein, 1988; Paterson 1991, Fitz- Gibbon, 1996).

STANDARD TABLES AND CHARTS (STACS)

Background

Standard Tables and Charts are a set of performance statistics provided to secondary schools each year to provide evidence for self-evaluation of NQ attainment (see Appendix for content details). Standard Tables were introduced in 1991 as a paper-based resource that allowed each school to compare its own performance against the national figures using 'National Comparison Factors', and to compare the performance in each subject department against that of other subject departments in the school using 'Relative Ratings'. Instructions were provided on how to use these for self-evaluation (SOED, 1991). The system has been further refined over the years, including the introduction of value-added measures between S4 and S5/6. In 1997, a National Management Information Systems Project Team (MIS) produced a prototype 'National MIS Graphing Tool' which created graphs to accompany the paper-based Standard Tables. EAs could provide the data for comparator schools, so that more realistic comparisons could be made. HMIe initiated a review of the Standard Tables prior to the introduction of the new NQ in 1999. The review was designed to take account of new courses being introduced, of developments in information and communications technology (ICT) and of experience in using the Standard Tables and the Attainment Summary. In 2000, STACS amalgamated the Standard Tables and the Charts, and was provided on CD-ROM to authorities and schools along with printed guidance on its use. Since then, further improvements

have been made for each version of STACS including the replacement of Relative Ratings by Relative Values and Value Added by Progression Values. STACS has also moved from a CD-ROM to a web-based service. Historical data from 1999-2003 were re-analysed by the new measures, thus a significant body of data exists for longitudinal study.

Whole-school performance measures

The definitions of performance measures used in STACS have a long history, and are very similar to those defined in the early 1990s, by HMIe's Audit Unit, for the Conservative Government's policy of publishing raw data to encourage parental choice. When New Labour came to power in 1997, the Scottish Office Minister for Education and Industry (Brian Wilson) asked for measures appropriate for their policy of target-setting for improving performance in schools. Under the terms of SOEID Circular 5/98 (SOEID, 1998), EAs were now required to publish annually school-by-school reports of progress towards Standard Grade and Higher Grade targets in all of their secondary schools. The target measures (% of S4 roll achieving 5+ Standard Grade awards at 1–2, 1–4 and 1–6, and % of S4 roll achieving 3+ and 5+ Higher Grade awards at A–C) were a subset of those previously published by the Audit Unit. Additional target measures were the % of S4 roll achieving Standard Grade English and mathematics at 1–6 (SOEID, 1999).

Following the 2000 Education Act (Scottish Parliament, 2000), virtually the same measures were selected as National Priority performance measures (although phrased in terms of the Scottish Credit and Qualifications Framework (SCQF), defined as the rolling average of three consecutive S4 cohorts by the end of S6, (and the reintroduced '1 Higher' measure).

STACS and target-setting

Performance targets for schools were introduced in 1998. According to the Senior Chief HMI, these were not intended to be ends in themselves but designed to:

> promote a specific focus and greater rigour within the process of self-evaluation and planning for improvement which is embodied by The Quality Initiative in Scottish Schools. The outcome will be higher standards of attainment, bringing better prospects for individual children in later life and for Scotland overall. (SOED, 1999)

Initial targets were set for 1999–2001, and were followed by a further target-setting cycle for 2002–5. Despite the ending of national collection of 5–14 data, authorities were required to continue to set targets, including 5-14 targets. Circular No. 02 June 2005 states: 'For 2005–7 [sic: misprint for 2008], schools and local authorities should continue to set targets for raising attainment and achievement.' Again, the implication is that both 5–14 and Standard Grade and Higher Grade targets are required.

Target-setting continues to be a key issue at education authority level and is still to the fore in conversations between QIOs and school management teams. In secondary schools, STACS data are used as the basis for setting targets at school level in Performance Reviews. In most authorities they are also used at micro level - within each school - and there is an expectation on the part of school managers, as well as Quality Improvement staff at authority level, that Faculty Heads/Principal Teachers analyse the detail in the data for their courses and set targets for groups or categories of pupils to help raise attainment.

Target-setting, however, represents a naive understanding of how schools operate and fails to acknowledge the complexity of change. It also can be seen as an attempt by central government to continue to direct the policy agenda using the concepts of quality assurance, accountability and continuous improvement. We believe that

target-setting, with its roots in systems analysis in commercial settings (Binney and Williams, 1995), is problematic and essentially 'unintelligent' in terms of accountability because it reflects underlying managerialist assumptions about the nature of change and does not lend itself to participative and interactive ways of working. What it does is to legitimate and reinforce bureaucratic relationships.

In setting targets there is an assumption that this will act as a spur to improvement. There is indeed a considerable body of evidence and theory within psychology and organization theory that 'goal-setting' can improve performance in some circumstances (Locke and Latham, 1990). However, there are a number of conditions which optimise and limit the effects (Wood, et al., 1987): the more challenging the goal the greater the effect, but partial success must be rewarded rather than seen as failure; people must have feedback about their performance in order to monitor their progress; there must be 'goal commitment' (people accept the goals, whether or not they set them themselves); and the task should be straightforward because the effect decreases with complexity.

However, while target-setting may be successful in some settings, the conditions described are not characteristic of schools. Firstly, there is no evidence of teachers and schools being committed to target-setting. When first introduced, the Headteachers' Association of Scotland questioned the statistical and educational validity of the targets and argued that they represented a 'morale sapping rod with which to beat headteachers' (1998). Secondly, no attention is paid to the 'cause-effect link' (Howe, et al., 1993:6). The prescription of educational outcomes without any notion of how to achieve them exemplifies the rational perspective of policy makers, which leads some to engage in wishful thinking.

Raising attainment is problematic and, without a clear strategy for achieving improved performance, setting targets is likely to be unproductive. An 'intelligent' approach to accountability would need to recognise a much more complex reality before establishing expectations of schools.

Perverse Incentives

The definition and utilisation of performance measures and targets sometimes provide the perverse incentives that 'intelligent accountability' should seek to avoid. This is illustrated by testimony to a committee of the Scottish Parliament:

> The target-setting regime is very narrow. It focuses on very limited areas of the curriculum and leads to distortions, because attention will be focused on areas on which people will be judged, with the risk that other parts of the curriculum will be neglected. There must be very strong pressures on teachers and head teachers… to adopt strategies that will increase the number of pupils passing at such and such a level. In other words, teachers will focus on pupils who are on or near the threshold, which is tough luck for pupils who are well off the mark, because they will not be brought to the point where they will make a difference to a target. That means that the most disadvantaged pupils might suffer further exclusion.

Further testimony from a headteacher demonstrates the inequitable application of resources on those pupils who will help the school to meet its targets:

> We concentrate on pushing youngsters from Standard Grade 3 to Standard Grade 2, or from Standard Grade 5 to Standard Grade 4[5].

'Pushing youngsters' from Standard Grade 2 to Standard Grade 1 was not a priority for this school, although there is significant evidence that progress to Higher in S5 from Standard Grade 2 in S4 is unlikely. Similarly, while there may be a significant benefit to an individual to improve her/his performance from 3 to 4 awards at Level 5, this is meaningless for schools trying to boost their attainment to reach their

targets. Only pupils who can improve their performance from 4 to 5 awards at Level 5 will matter. Thus there are incentives to discriminate in favour of pupils near the borderline to an arbitrary level.

Further issues arise because the system is geared towards high attaining academic pupils who stay on in school to achieve entrance qualifications for higher education. For example, performance indicators for S5 and S6 are based on the S4 school roll, and this means that schools which encourage pupils to follow more appropriate courses in Further Education after S4 or S5 are likely to show lower attainment than those which persuade them to stay within the school for a further year.

In addition, schools serving areas without Special School provision take all pupils, and have such pupils effectively labelled as failures – Level 1 and 2 courses, as well as alternative syllabi such as those certificated by the Award Scheme Development and Accreditation Network (ASDAN) were disregarded in the performance measures identified in the National Priorities, the statistics collated in STACS[6] and the attainment measures valued by HMIe.

What, then, is the value and utility of STACS? The received view is that the STACS support schools trying to raise pupil attainment, but a closer look suggests that this is not necessarily the case. For example, because the Whole School measures have arbitrary cut-offs, a school which sees a need to increase the proportion of its pupils gaining 4 Higher Grade examination passes will have no support from STACS. The 'top-down' selection of whole school measures, determined at national level suggests 'unintelligent' accountability. To allow intelligent use of data, STACS could have been constructed to allow schools to measure their progress towards targets they set for themselves based on pupil-level data. It would be comparatively simple to make STACS 'intelligent' on whole-school measures by providing data on a much wider range, and not simply National Priority measures.

An important challenge for secondary schools is the need to improve the pupils' performance in individual classrooms. Performance in individual subjects is neither part of national targets nor publicly available. This is the area of STACS focused most closely on school self-evaluation and for this reason it merits closer attention.

Subject comparisons: the impact of STACS subject analysis over time

Subject comparisons are a unique feature of STACS, and we may speculate that their introduction in Scotland, but not in England, was made possible by the existence of a single national examination authority in Scotland. Subject comparisons provide the opportunity to investigate within-school differences, while the whole-school measures demonstrate between-school differences.

Performance indicators for subject departments are derived by comparing each pupil's performance in a subject of interest with their average performance in all the other subjects they attempted. By definition, since these comparisons are internal to the school, some subjects will have negative RV, and some will have positive RV. (See appendix for details of the statistical process). A key concept underlying the indicators of subject performance, such as RV (and their predecessor Relative Ratings) is that, other factors being equal, a pupil's performance in any one subject ought to be broadly in line with her/his performance in other subjects. The STACS analysis highlights where this is not the case, allowing schools to consider why these differences might arise.

However, we should note that methodologically RV are a poor substitute for Value Added, since the calculation of RV takes account of each pupil's concurrent attainment in other subjects, whereas Value Added is a measure of each pupil's progress and takes account of her/his prior attainment. RV is used in STACS because there is no national system of baseline assessment in Scotland, and therefore it is not possible to estimate Value Added until S5. (There are further statistical problems beyond the scope of this article.)

STACS represent a major investment in data analysis by government, aimed at improving performance by providing evidence for schools to use in self-evaluation. It is therefore important to test the assumption that the provision of STACS data leads or has led to improvement in the performance of subject departments within secondary schools. If this system works there should be evidence from STACS of a reduction in the number of subject departments that were performing below average and evidence of improvement in performance by subject departments that previously were 'failing'.

We set out to examine the STACS data to investigate whether there was any improvement over time. Our analysis of STACS data focused on S4 Standard Grade attainment and RV. Table 1 shows the overall distribution of RV that were significantly above or below average during the period for which data were available. If schools have been addressing under-performance in their weaker departments, then one would expect to see an increase in the number of school/subjects with no significant RV and a consequent decrease in those with positive or negative RV. Table 1 suggests that there has been some movement in this direction (although not 'continuous' as required by the 2000 Act), and that the weakest departments may be becoming a little more 'exposed'.

Standard Grade Relative Value Significance	1999	2000	2001	2002	2003	2004	2005	Change from 1999 to 2005
*** Positive	15.1%	15.3%	15.5%	15.4%	14.0%	14.1%	14.1%	-1.0%
** Positive	7.1%	6.7%	6.6%	6.3%	6.5%	5.7%	6.5%	-0.6%
* Positive	4.2%	3.9%	3.9%	3.4%	3.6%	3.8%	3.1%	-1.1%
No Significance	50.5%	51.0%	51.0%	52.5%	53.5%	54.5%	53.7%	3.2%
* Negative	3.9%	3.9%	3.5%	3.6%	3.3%	3.6%	3.6%	-0.3%
** Negative	6.7%	6.5%	6.4%	6.1%	6.4%	5.6%	5.9%	-0.8%
*** Negative	12.4%	12.8%	13.1%	12.8%	12.6%	12.7%	13.0%	0.6%

However, RV are only relative within each school, and these Scotland-wide figures may disguise radically different patterns within schools. It is, therefore, necessary to pursue the analysis through subjects that display a long term pattern of negative RV.

Overall patterns

We hypothesized that if the use of STACS within the improvement agenda was working as intended, then action would have been taken since 1998 to identify the cause of negative RV in any department and apply remediation. To investigate this hypothesis we used the data on RV to examine the incidence of subject departments with negative patterns of performance between 2001 and 2005. STACS use a formula to identify 'patterns' of attainment over 3, 5 and 7 year periods. A negative pattern in a subject is one where all RVs are negative and the negative RV in the current year, as well as a majority of years, are statistically significant. (In Table 2 below, N represents a negative pattern over three years, NN a negative pattern over five years, and NNN a negative pattern over seven years[7].)

If schools are using the evidence of negative RV to identify and address problems in performance in a subject department, we expect that the % of presenting school/

subjects[8] which had an identified negative pattern would have reduced. However, Table 2 shows the percentage of school/subjects with negative patterns of RV at Standard Grade, and indicates no significant change in the 3 and 5 year patterns. (Whether 2.1% of presenting subjects with a 7 year negative pattern is 'low', is a matter of judgment.)

Table 2: Percentage of subjects with negative RV patterns over three, five and seven years					
	2001	2002	2003	2004	2005
Number of subjects presenting at SG	8343	8219	8068	7772	7499
% N	6.6%	6.5%	6.7%	6.7%	6.8%
% NN			4.2%	4.5%	4.4%
% NNN					2.1%

The STACS-defined '5 year pattern' is available for 2003, and 340 S4 Standard Grade subjects in Local Authority schools were categorised as very significantly negative on this basis. This allows a more detailed examination of subsequent progress in these subject departments; by 2005:

- 53 (16%) showed potentially secure improvement in the 2004 and 2005 results;

- 71 (21%) showed improvement in 2005 only;

- 36 (11%) remained negative, but only at the 95% confidence level, so were not identified as having a 'pattern';

- 156 (46%) continued to show negative patterns. Of these 3 had re-established a 3-year negative pattern, 13 still had a 5-year negative pattern, and 140 had 'progressed' to a 7-year negative pattern;

- 24 (7%) of these had been cut from the curriculum (a number of which were replaced by Intermediate/Access courses).

This analysis, however, takes no account of the numbers of pupils being presented – small subjects are less likely to meet the confidence levels required to be highly significant. English and Mathematics, however, are studied by all S4 pupils[9]. Since these subjects also measure the school's literacy and numeracy they are also whole-school measures and might reasonably be expected to be the focus of special attention by school management and authority Quality Improvement staff.

Table 3 overleaf shows that a substantial proportion of 407 local authority schools exhibit negative patterns in English and mathematics. Since these patterns can be identified, it must be of concern that attainment raising procedures appear not to have been successful.

Although these percentages seem worryingly high, not all the 3000+ pupils in these departments can be assumed to be doing less well than expected. Pupil-level analysis of performance in a sample of 'NNN' subjects via their RV scatter plots

and tables by gender, indicates that the suggested under attainment may affect only a subset of pupils by one or more of gender, overall attainment and/or teacher.

Table 3: Percentage of schools with negative RV patterns in English and Mathematics					
English	2001	2002	2003	2004	2005
% N	11.3%	8.6%	10.6%	10.1%	11.1%
% NN			7.9%	7.6%	8.1%
% NNN					4.7%
Mathematics	2001	2002	2003	2004	2005
% N	13.5%	15.5%	13.8%	13.5%	12.5%
% NN			10.6%	9.6%	8.4%
% NNN					5.7%

Evidence of Improvement

Thus far the focus has been on school/subjects that have not improved. Is there evidence of improvement? There is no standard definition of whether improvement has occurred or not. For processing large quantities of data, the following arbitrary standard of 'improvement' has been adopted:

> where a school/subject had a highly significant negative RV in one or more years, it is considered to be securely improved, once it has shown no significant negative RVs for 3 successive years.

Table 4 below displays the percentage of the 8,831 EA Standard Grade school/subjects which had presentations in 1999-2002 and were rated as highly significantly negative in one or more years. The percentages of these negative subjects that had been cut from the curriculum by 2005, had improved, and which had not demonstrated 'secure improvement' are also shown. Each row represents one authority and the table is ranked by the 'improvement rate' in descending order. The final row shows the mean performance of all authorities.

Table 4 also shows that nationally 31% (2699) had a highly significant negative RV in one or more years. Of these, 15% were no longer presenting pupils, 24% had met the 'improvement test' above and 60% had not improved by this measure. Some Standard Grade subjects no longer appear as they have been replaced by Intermediate courses. This, however, may be evidence of an improvement agenda in operation. Others have simply been 'cut' from the curriculum.

The influence of EAs

EAs are responsible for operating quality assurance procedures in their schools, and it was hypothesized that some would be more effective than others in reducing the percentage of 'unimproved subjects'. Differences do appear. At one end of the scale are two EAs where 75% of their school/subjects have not improved, while at the opposite end is one authority where only 47% of its SG school/subjects have not improved. The range of 13% to 35% of negative subjects securely improved suggests that some EAs may have created a more effective improvement strategy than others.

To shed light on these differences, three geographically disparate authorities agreed to an external analysis of their Standard Grade attainment over time and to respond to further interview. In all three EAs, there are explicit and formal Quality

Table 4: Improvement of Negative RV subjects by 2005

% of SG school/subjects rated highly significantly negative 99-02	Cut	Improved	Not Improved
%	%	%	%
30%	15%	35%	49%
27%	17%	33%	50%
27%	14%	30%	56%
16%	11%	30%	59%
35%	13%	29%	58%
24%	7%	29%	64%
29%	10%	29%	62%
35%	8%	27%	65%
33%	8%	27%	65%
37%	4%	27%	69%
33%	13%	26%	61%
30%	15%	26%	59%
34%	14%	26%	60%
29%	27%	25%	48%
33%	11%	25%	64%
35%	25%	25%	51%
27%	9%	25%	66%
29%	18%	25%	58%
30%	25%	25%	50%
25%	19%	25%	57%
29%	16%	24%	61%
28%	30%	23%	47%
24%	17%	23%	60%
30%	16%	23%	61%
36%	16%	23%	61%
38%	18%	21%	61%
29%	5%	20%	75%
36%	20%	19%	61%
18%	19%	19%	63%
32%	14%	15%	71%
31%	11%	14%	75%
18%	13%	13%	73%
30%	**15%**	**24%**	**60%**

Figures placed to the left of the box are 1 SD lower than the
Figures placed to the right of the box are 1 SD higher than t

41

Assurance (QA) meetings with headteachers (and other staff) on attainment. There is significant variation between them as to the structure of such meetings, and all have recently changed or are about to change the structure of QA processes in order to focus attention more closely on indicative data and attainment raising. One EA is planning to extend the QA visit from 1 to 2.5 days, and to include peer review as part of the process.

In two EAs, the study of Inspection of Education Authority (INEA) reports by HMIe was specifically mentioned as a factor in current or previous procedural changes. It would seem that there has been significant movement since the 2000 Act. As one QIO noted:

> there had been no attention paid at an EA level to this [attainment data use] before that time… I therefore hesitate to draw a direct (or indeed any) link between improvement and the use of this data. I don't think we have developed sufficiently in our approach yet to make that kind of claim.

All three EAs valued STACS derived data and indicated that most staff thought they were excellent as a self-evaluation and planning tool. Additionally, they were well aware of the negatively rated subjects within their schools, and of the staffing and other issues which were obstacles to raising attainment in these areas.

Use of STACS in schools

The next phase of the research study involves case studies of the use of performance data in schools across Scotland. At the time of writing case-study work is in process, and it is too early to report firm conclusions. However, there are early indications that teachers accept the need to be held accountable by management for the performance of pupils. The majority of school staff interviewed so far appears to find STACS a useful tool, in so far as it helps them to identify strengths and weaknesses, but they see this as simply one tool among the many they employ.

There were some suggestions, however, that use of STACS by management is occasionally poorly handled and lacking intelligence. For example, references were made by several staff to the ease with which the statistics for Intermediate and Higher courses can be improved by the simple expedient of not presenting potential failures, or presenting them at a lower level than they might achieve. At Standard Grade, inflation of grades for internally assessed elements is also seen as a potential response to management pressure. None admitted to adopting this course of action, but some 'knew of others who had done so'. In other words, ill-informed discussions between management and teachers on STACS data could well result in a 'dumbing down' of standards, instead of challenging pupils to do better.

More positively, STACS subject analysis appears to motivate many staff to aim for pupils' improvement within the General and Credit Levels – in contrast to the Headteacher (whose testimony to the Scottish Parliament is reported above) who claimed to focus resources only on those pupils who might move from one National Priority Target category to a higher one.

DISCUSSION

Having examined STACS in some detail, we turn now to consideration of the extent to which STACS fits with the concept of Intelligent Accountability.

Performance indicators that measure appropriate outcomes of schooling

The whole focus of STACS is attainment in NQs. These are important outcomes of schooling, but they are narrowly focused on academic attainment. STACS measures that which is easy to measure – not the wide range of educational outcomes that should be the aim of education, and are embodied in the National Priorities. This

problem of easy measurement was criticised by O'Neill, who argued that while in theory:

> the new culture of accountability and audit makes professionals and institutions more accountable for good performance...the real focus is on performance indicators chosen for ease of measurement and control rather than because they measure accurately what the quality of performance is.

It might be argued that the degree to which an over emphasis continues to be placed on easily measured outcomes is a reflection on the extent to which accountability remains 'unintelligent'.

Measures that encourage the fullest development of every pupil

We have argued that attempts to contextualise examination outcomes, using school characteristics based on aggregated data, have limited value and that the focus should be on pupil-level data. However, it is quite difficult to use STACS to focus on individual pupils, since the focus of STACS is on comparison – of the school versus other schools, and of subject departments versus other subject departments. The facility to see limited pupil-level data exists in the scatter plots provided to schools (see Appendix) and we would suggest that a more productive use of data would be to extend the provision of pupil-level data to enable teachers to track the progress of individual pupils at all stages. This is not a current function of STACS, yet it would sit comfortably and support the emphasis on personalization contained within A Curriculum for Excellence (SEED, 2004b). It has been left to schools and authorities to develop their own systems of tracking. It is not clear why the retrospective analysis of performance, that STACS embodies, should have attracted such significant national funding, while tracking the progress of individual pupils has enjoyed no such benefit.

Focus on self-evaluation

STACS are ostensibly designed to provide evidence for schools to evaluate their own performance, and so we might at first sight conclude that they are conducive to 'intelligent' accountability. However, while the focus on school self-evaluation is generally portrayed as fairly benign, some commentators give a more negative interpretation. It has been argued, for example, that:

> The supposed 'self-empowerment' of this system rests upon simultaneous imposition of external control from above and internalization of new norms so that individuals can continuously improve themselves (Shore and Wright, 2000:61).

This suggests a tension between different models and purposes of accountability, which is insufficiently recognised. It may be difficult to reconcile this tension, particularly at a national level, but in the interests of improved schooling and 'intelligent' accountability, the consequences of continued reliance on underlying managerialist assumptions about the rational nature of change need to be reconsidered. Furthermore, there are problems with the arbitrary definition of performance measures, with STACS data representing an essentially 'top-down' set of performance indicators. Our analysis suggests that they are not necessarily useful to those who are intended to use them.

Trust and professionalism

We have argued that accountability relationships in the school system have changed considerably and become confused in the past 15 years. Professional accountability has been reduced in significance and replaced by an ambiguous mix of bureaucratic

central control models and consumerist public and market models of accountability, which is curiously defined as 'tough' and 'intelligent'. Several aspects of these changed relationships suggest subordination to managerial control of professional autonomy and judgment, as well as the school's accountability to parents and the community. Our analysis suggests that this subordination is accomplished either through application of the principles of performance management, or through softer approaches which emphasise a strong achievement orientated culture.

To carry out self-evaluation using the off-the-shelf measures defined by STACS and the HGIOS Quality Indicators, schools must adopt these measures and targets as their own, and work towards achieving them, rather than consider what the most appropriate outcomes of their pupils' schooling should be. Ball (1994:54) describes this as a 'no hands' form of control, and refers to, 'steering at a distance' with 'prescription replaced by ex post accountability' with the appearance of autonomy.

One problem with accountability mechanisms is that they focus on performance and do not address issues relating to policy formulation (Ball, et al., 1997). Genuine self-evaluation would involve the school and its community having greater control over the measures by which it is assessed. Accountability could then be seen in terms of the quality and robustness of the school's self-evaluation structures and processes. This argument links accountability to the concept of culture and gives rise to consideration of power and organisational structures — the frameworks which govern relationships — and to how people interact within an organisation. Different conceptions of organisational relationships and structures and the dispersal of power have different accountability implications that we believe truly 'intelligent' needs to take account of.

Accountability and organisational culture

In her critique of accountability cultures, O'Neill suggested that while

> in theory the new culture of accountability and audit makes professionals and institutions accountable to the public… the real requirements are for accountability to regulators, to departments of government, to funders, to legal standards.

Analysis of the development of the Quality Initiative in Scottish Education suggests that these 'real requirements' continue to characterize the accountability culture in Scotland. The driving force in developing accountability systems since the early 1990s has been an Inspectorate committed to the quality improvement and quality management, which has justified the use of performance indicators through providing 'Information for Parents'. STACS and the self-evaluation process of HGIOS were initially developed by the HMIe Audit Unit, and the HGIOS quality indicators are an enduring feature of individual school inspections.

Couched in the language of partnership the 'Quality Initiative' might seem a productive way of improving schools but the emphasis on 'tough, intelligent' accountability suggests that the discourse of partnership and consultation masks a concern for control based on underlying managerialist assumptions about the rational nature of change. However, the assumption that private sector concepts such as quality and management performance will improve what goes on in schools is challenged by the literature which suggests that change is not controllable (Gleik, 1987; Waldrop, 1992), and by evidence that has been available for many years about how and why reform in education fails or succeeds (Cuban, 1988; Pink, 1989; Sarason, 1990).

One lesson that can be taken from this literature is that schools are part of wider organizations, which suggests that the relationship between SEED, EAs and schools may be crucial. In distinguishing between 'first-order' and 'second-order' change, for example, Cuban (1988) points to the structural, systemic and system-disturbing nature of change. First-order change is relatively straightforward and focused on

improving the efficiency and effectiveness of existing practices without changing organisational features, but second-order change attempts to change organisational roles and relationships.

Empirical evidence on school district relationships in North America supports our analysis (Murphy and Hallinger, 1993; Leithwood, 1996). Leithwood argues that one of the lessons to be drawn from his work is that policy implementation should be oriented 'more towards the development of commitment (and capacity) and less towards compliance' (p.392). He also argues that provincial and state government should change from a compliance or regulatory approach to policy development and implementation to an 'educative' one, which includes high involvement principles within a systemic approach. Indeed, the recent work of both Fullan, et al. (2004) and Leithwood, et al. (2006) is focused on reshaping school districts and developing both schools and districts as professional learning communities. Fullan (1991) argues that improvement is about capacity building and that this involves developing and using policies that increase the collective power of every level in the system to generate improvement. This involves focusing on the culture of the EA and developing a set of relationships across the system based on trust and shared commitment.

Although Fullan and Leithwood and their colleagues mount powerful arguments for intelligent accountability based on the development of professional learning communities, there is no empirical evidence to suggest that such communities would necessarily flourish if current accountability systems were removed. What does seem clear, however, is that the context within which performance measures such as the STACS are used is likely to affect whether they are used intelligently. Models of accountability based on compliance represent the antithesis of a professional learning community because there is little scope for collaborative working or collegiate decision-making, and limited opportunity to learn from collective experience. An 'educative' approach may therefore offer a means of developing more direct, appropriate and intelligent accountability because evaluation criteria could then be negotiated and set against negotiated school purposes.

The work of Fullan, et al. (2004) and Leithwood, et al. (2006) suggests that if there is respect and trust between EAs and schools, and within schools, performance data can be used in a positive way as a monitoring tool, as a source of evidence to inform decision-making and teaching in schools and as a tool to look for problems and weaknesses so that they can be addressed. If, however, there is a context of distrust, cynicism, or even fear, problems may be concealed and performance data may be ignored (Cowie, 2001) or used in tactical ways to deliver short term 'improvement' that is not sustainable and ultimately self defeating (Gray and Wilcox, 1995). Deming (1982) argued that, where there is fear, people produce the wrong figures and there is some evidence to suggest that within an accountability culture teachers may focus on impression management, and generate performances of their work in order to satisfy the accountability demands (Webb, 2006).

CONCLUDING COMMENTS

At the national level the temptation to focus on accountability has proved irresistible. We do not find this surprising because accountability provides what appears to be a quick and easy means of generating improvement. Our analysis, however, suggests that the presumed improvement in standards offered by 'tough, intelligent accountability' may be illusory. Where there is emphasis on control, managers may abdicate responsibility for improvement, and 'tough, intelligent accountability' provides a means of scape-goating or apportioning blame for failings in the system as a whole.

By definition, learning communities are constantly developing. Meaningful change is complex, open ended and not well defined and because it involves changes in conception and role behaviour it is difficult to achieve (Fullan, 1991). Such change

means change in work patterns, perceptions and relationships and we have argued that the managerial approach, which the concept of 'tough, intelligent accountability' is grounded on, may discourage the kind of constructive partnerships which might be capable of dealing with the complexity of change. We are currently exploring these issues further in case studies of schools.

'Intelligent accountability' should be concerned about the implications that judgments of performance and mechanisms for ensuring accountability might have for what goes on in schools. This would involve questioning the extent to which quality assurance activities and accountability mechanisms are likely to encourage a sense of coherence and purpose, which would support improvement, or undermine improvement processes through the creation of a climate of fear.

Deming (1982) argued against exhortations and target setting because they help create adversarial relations and demoralise staff, and focused instead on the importance of system-wide responsibility, team work and self-improvement. Improvement is about progress and change and so underlying managerialist assumptions about the nature of change, as reflected in the rhetoric of 'tough, intelligent accountability,' may need to be set against an alternative conception in support of an open, intelligent and professional, process-based approach to accountability. This would place the emphasis on changed relationships and on continuous debate, within and among schools, between schools, officers of education authorities and HMIe, between schools and the communities they serve, and between teachers and pupils.

NOTES

1. Gleeson and Husbands cite Elliot (2001) and Merson (2001) who suggest that this is because performance management does not connect with the contextual realities of learning and teaching.
2. Similarly, the English policy documents do not refer to the Scottish system of school self-evaluation that has been in place for many years and is embodied in How Good is Our School? In the same way, proportionate inspection was simultaneously introduced on both sides of the border, with neither acknowledging the other. Despite devolution, there seems to be a common source of policy development.
3. NQ include Standard Grade, Access, Intermediate, Higher and Advanced Higher courses.
4. Including local area measures of the percentage of women with degrees and households where the main householder has never worked, together with the percentage of pupils entitled to free meals, living in the 15% most deprived data zones, those who have additional support needs (ASN) or live in urban areas.
5. Testimony to the Scottish Parliament Education, Culture and Sport Committee Wednesday 9 February 2000. www.scottish.parliament.uk/business/committees/historic/education/or-00/ed00-0502.htm
6. However, a number of non SQA courses will be incorporated into the Unified Points Scale used by STACS, in 2007.
7. N could be first measured in 2001, NN in 2003 and NNN in 2005. Subjects which had an identified negative pattern over 7(NNN) years would be relatively few in number.
8. One 'school/subject' indicates one subject presentation in one school.
9. The number of S4 presentations in these subjects at other levels (Access/Intermediate) was extremely small, and these were sometimes dual presentations with SG.

REFERENCES

Adelman, C. and Alexander, R.J. (1982) The Self-Evaluating Institution: Practices and Principles, The Management of Educational Change, London: Methuen.
Ball, S.J. (1994) Education Reform: A Critical and Post-Structural Approach, Buckingham: Open University Press.
Ball, S.J., Vincent, C. and Radnor, H. (1997) Into Confusion: LEAs, Accountability and Democracy, Journal of Education Policy, 12(3), 147–163.

Binney, G. and Williams, C. (1995) Learning into the Future - Changing the way People Change in Organisations, London: New York.

Cowie, M. (2001) Talking Heads: A Critical Analysis of the Quality Assurance Relationship between Headteachers and an Education Authority, Research Paper 5, Centre for Educational Research, University of Aberdeen, March 2001.

Cowie, M. and Croxford, L. (2006) Measuring and Monitoring School Performance in Scotland, Education in the North, 13, 23–36.

Croxford, L. (1999) The statistical background to government target-setting policy, Edinburgh: Educational Institute of Scotland.

Cuban, L. (1988) A fundamental puzzle of school reform, Phi Delta Kappan, 70(5), 341–44.

Deming, W.E. (1982) Out of the Crisis: Quality, Productivity and Competitive Position, Cambridge, Mass: Cambridge University Press.

Elliot, J. (2001) Characteristics of performative cultures: their central characteristics and limitations as resources for educational reform, in D. Gleeson and C. Husbands (eds), The Performing School: managing, teaching and learning in a performance culture, London: Routledge Falmer.

Fitz-Gibbon, C.T. (1996) Monitoring Education: Indicators, Quality and Effectiveness, London: Cassell.

Fullan, M. (1991) The New Meaning of Educational Change, London: Cassell.

Fullan, M., Bertani, A. and Quinn, J. (2004) Lessons from district wide reform, Educational Leadership, 61 (6), 42–46.

Gleeson, D. and Husbands, C. (2003) Modernizing schooling through performance appraisal, Journal of Education Policy, 18 (5) 499–511.

Gleick, J. (1987) Chaos, London: Abacus.

Goldstein, H. (1984) The Methodology of School Comparisons, Oxford Educational Review, 10.

Gray, J. and Wilcox, B. (1995) Good School, Bad School: Evaluating Performance and Encouraging Improvement, Buckingham: Open University Press.

Halstead, M. (1994) Accountability and Values, in D. Scott (ed) Accountability and Control in Educational Settings, London: Cassell.

Headteachers' Association of Scotland (HAS), (1998) Target Showdown, Times Educational Supplement (Scotland), 1 May 1998.

Howe, R.J., Gaeddert, D. and Maynard, H.M. (1993) Quality on Trial, St.Paul:West Publishing Company.

Kogan, M. (1986) Educational Accountability: An Analytical Overview, London: Hutchinson.

Leithwood, K. (1996) Commitment - Building Approaches to School Restructuring: Lessons from a Longitudinal Study, Journal of Educational Policy, 11(3) 377–398.

Leithwood, K., Aitken, R. and Jantzi, D. (2006) Making Schools Smarter: Leading with Evidence, Sage.

Lyotard, J.F. (1984) The postmodern condition: a report on knowledge, Manchester: Manchester University Press.

Locke, E.A. and Latham, G.P. (1990) A Theory of Goal Setting and Task Performance, Englwood Cliffs, NJ: Prentice Hall.

Merson, M. (2001) Teachers and the myth of modernization, in D. Gleeson and C. Husbands (eds), The Performing School: managing, teaching and learning in a performance culture. London: Routledge Falmer.

Milliband, D. (2004) Personalised learning: Building a new relationship with schools, speech by Minister of State for School Standards to the North of England Education Conference, Belfast, 8th January 2004, www.standards.dfes.gov.uk/personalisedlearning/about/

Murphy, J.M. and Hallinger, P. (1993) Restructured Schooling: Learning from Ongoing Efforts, Murphy, P. and Hallinger, P. (eds), (1993) Restructured Schooling: Learning from Ongoing Efforts, Newbury Park, CA: Corwin Press, Inc.

O'Neill, O. (2002) Lecture 3: Called to Account, Reith Lectures 2002: A question of Trust, www.bbc.co.uk/radio4/reith2002

Paterson, L. (1991) An Introduction to Multilevel Modelling, in S.W. Raudenbusch. and J.D. Willms, (eds) Schools, Classrooms and Pupils: International Studies of Schooling from a Multilevel Perspective, London: Academic Press.

Pink, W.T. (1989) Effective Staff Development for Urban School Improvement, Paper presented at American Educational Research Association annual meeting, (cited in Ross, 1993).

Sarason, S. (1990) The Predictable Failure of Educational Reform, San Francisco: Jossey-Bass.

Scottish Office Education Department (1991) Using Examination Results in School Self-Evaluation: Relative Ratings and National Comparison Factors, HM Inspectors of Schools, Edinburgh: HMSO.

Scottish Office Education and Industry Department (1997) The Quality Initiative in Scottish Schools

(QISS) - final draft of a QISS discussion paper presented at a seminar organised by HM Inspectors of Schools Audit Unit, 10 June 1997, Edinburgh.

Scottish Office Education and Industry Department (1998) Reporting on School Performance Against Standards For Improvement In Standard Grade And Higher Grade Attainment, Circular No. 5/98, Edinburgh: SOEID.

Scottish Office Education and Industry Department (1999) Raising Standards – Setting Targets: Setting targets in Scottish Schools, National and Education Authority Information 1999, HM Inspectors of Schools Audit Unit, Edinburgh: HMSO.

Scottish Executive Education Department (2002) How good is our school? Self-evaluation using performance indicators, HM Inspectorate of Education, Edinburgh: HMSO.

Scottish Executive Education Department (2004a) Ambitious Excellent Schools: our agenda for action, Edinburgh: Scottish Executive.

Scottish Executive Education Department (2004b) A Curriculum for Excellence, Edinburgh: Scottish Executive.

Scottish Parliament (2000) Standards in Scotland's Schools etc. Act, www.opsi.gov.uk/legislation/scotland/acts2000/20000006.htm

Secondary Heads Association (2003) Towards Intelligent Accountability for Schools, Policy Paper 5, www.ascl.org.uk

Shore, C. and Wright, S. (2000) Coercive accountability: the rise of the audit culture in higher education, in Strathearn, M. (ed) Audit Cultures: Anthropological studies in accountability, ethics and the academy, London: Routledge.

Simkins, T. (1997) Autonomy and Accountability, in B. Fidler, S. Russell, and T. Simkins (eds) Choices for Self-Managing Schools: Autonomy and Accountability, London: Paul Chapman.

Waldrop, M.M. (1992) Complexity: The Emerging Science at the end of Order and Chaos, Harmondsworth: Penguin.

Webb, P.T (2006) The choreography of accountability, Journal of Education Policy, 21 (2) 201–214.

Wood, R.E., Mento, A.J., and Locke, E.A. (1987) Task Complexity as a Moderator of Goal Effects: A Meta-analysis, Journal of Applied Psychology, 72, 416–25.

Woodhouse, G. and Goldstein, H. (1988) Educational Performance Indicators and LEA League Tables, Oxford Review of Education, 14(3), 301–320.

APPENDIX - STACS OUTPUT AND CALCULATIONS

As the name suggests, STACS are a set of standard tables and charts that are provided to all state-funded secondary schools. They are based on NQ results. The set of tables and charts includes:

- National Priority whole school measures by school, authority or national

  – % of the S4 cohort who attain 5 or more awards at SCQF Level 3 (equivalent to Standard Grade Foundation) or better
  – % of the S4 cohort who attain 5 or more awards at SCQF Level 4 (equivalent to Standard Grade General) or better
  – % of the S4 cohort who attain 5 or more awards at SCQF Level 5 (equivalent to Standard Grade Credit) or better
  – % of the S4 cohort who attain 1 or more awards at SCQF Level 6 (equivalent to Higher) or better
  – % of the S4 cohort who attain 3 or more awards at SCQF Level 6 (equivalent to Higher) or better
  – % of the S4 cohort who attain 5 or more awards at SCQF Level 6 (equivalent to Higher) or better

- Subject course reports for S4, S5, and S5/6, which include:

  – Relative Value (RV) table for All, Male and Female
  – RV scatter plot diagram for current year
  – Progression Value (PV) table for All, Male and Female - S5 and S5/6 only

- PV scatter plot diagram for current year - S5 and S5/6 only
- Table of actual numbers, percentages and National Comparison Deciles (NCD)

Relative Values (RV) address the question: 'How well did the students do in my course compared with how well the same students did in their other courses in the school, taking account of how students of similar general attainment nationally perform in the examination?'. They are calculated using a regression analysis. The RV is the average of the residuals for the school. A positive value for RV shows that on average pupils did relatively well in the subject/course, while a negative RV shows that pupils did relatively badly in the subject/course.

Calculation of RV

For each course with entries nationally, for the entire national cohort, and separately for each stage and males, females and all candidates, fit a generalised model with cubic, quadratic, linear, and intercept terms between the actual grade in that course and the pupil's UMEG.

MODEL $Y = ß1X + ß2X2 + ß3X3 + a$

Store these parameters
Store also the R-square value (RSQ)

2. Apply the following 2 tests to each model.

- N (National entries with at least one other subject) < 60 – if TRUE reject this target subject.

  Or

- $N < 400(1-RSQ)$ – if TRUE reject target subject.

3. Use the parameter estimates to derive the predicted value for each pupil in the cohort under consideration:

   Predict $= ß1X + ß2X2 + ß3X3 + a$

4. The pupil residual is the predicted value - the actual value

5. For a particular regression model take the mean of all the pupil residuals at a Centre/LA to give the new 'Relative Value' for each centre/LA.

6. The Standard Error of the mean residuals for this centre or LA is stored to allow calculation of asterisks and confidence limits.

The scatter plot below shows a typical example of the regression analysis used to calculate RV. In this case the target course is Biology. Each point represents a student. The Y axis is the actual grade the pupil achieved in the target course. The x axis is the pupil's average grade using the Unified Points Score Scale (UPS) in all his/her other courses in the same diet of examinations. The curved line is the

High School Standard Grade Biology Relative Value 2003
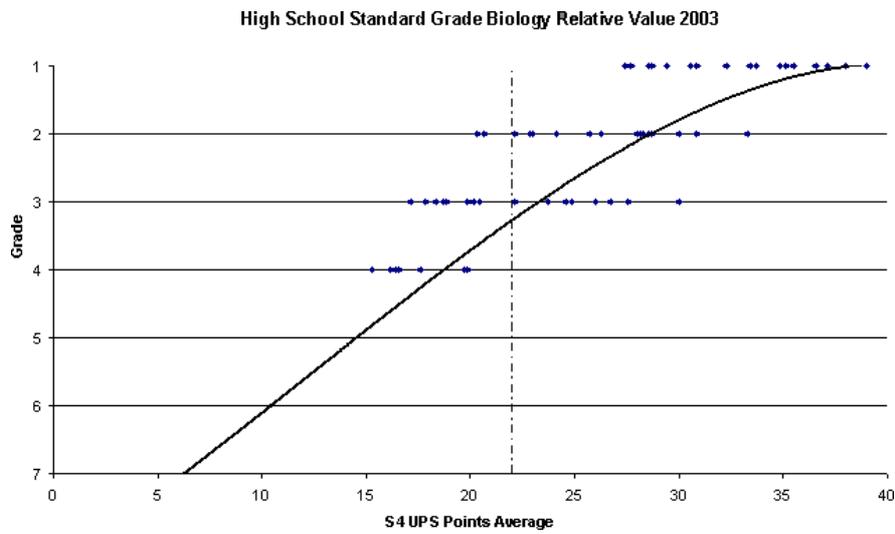
national regression line and represents the grade predicted in the target course for any UPS average grade. In this case the balance of points lie above the regression line and so the mean residual i.e. the Relative Value is positive.

Progression Value (PV) uses the same statistical technique as Relative Values but compares the performance of the students in the target course, with the performance of the same students in the courses they took in S4. It therefore provides a measure of the progress made by students in a course over an identified period of schooling, compared with the national norm for students with the same prior attainment undertaking the same course.

National Comparison Deciles (NCD) indicate where the school's value for a measure comes in a National Ranking. NCDs normally take values between 1 and 10. An NCD of 1 indicates that the score places the school in the highest 10% of EA schools nationally and an NCD of 10 places the school in the lowest 10% on any particular