

Session Number : 5
Session Title : Education - experience in measuring output growth
Session Chair : Paul KONIJN

**Paper prepared for the joint OECD/ONS/Government of Norway workshop
"Measurement of non-market output in education and health"**

London, Brunei Gallery, October 3 – 5, 2006

**Measuring Government Education Output Quality in England: an
overview of the issues and approaches developed by the
Department for Education and Skills**

Deborah Garniss

For additional information, please contact:

Author name: Deborah Garniss
Author address : Area 6S, Sanctuary Buildings, Great Smith Street, London SW1P
3BT, UK
Author E-mail: Deborah.garniss@dfes.gsi.gov.uk
Author fax: 020 7925 5097
Author telephone: 020 7925 5016

This paper is posted on the following website :
http://www.oecd.org/document/34/0,2340,en_2649_33715_36450978_1_1_1_1,00.html

EDUCATION OUTPUT

What exactly do we mean by education output?

1.1 Education is a complex process which aims to help achieve a range of ultimate outcomes. In broad terms, these are the maximisation of social and economic welfare: promotion of economic growth; reduction of social inequality; enhancement of happiness or well-being. So when we think about measuring what the education system ‘produces’, we want ideally to measure the contribution that education makes to these outcomes.

1.2 In our Green Paper, *Every Child Matters (ECM): Change for Children*, we identified five specific outcomes that underpin the Department’s focus on services to children and young people:

- Being Healthy
- Staying Safe
- Enjoying and Achieving
- Making a Positive Contribution
- Achieving Economic Wellbeing.

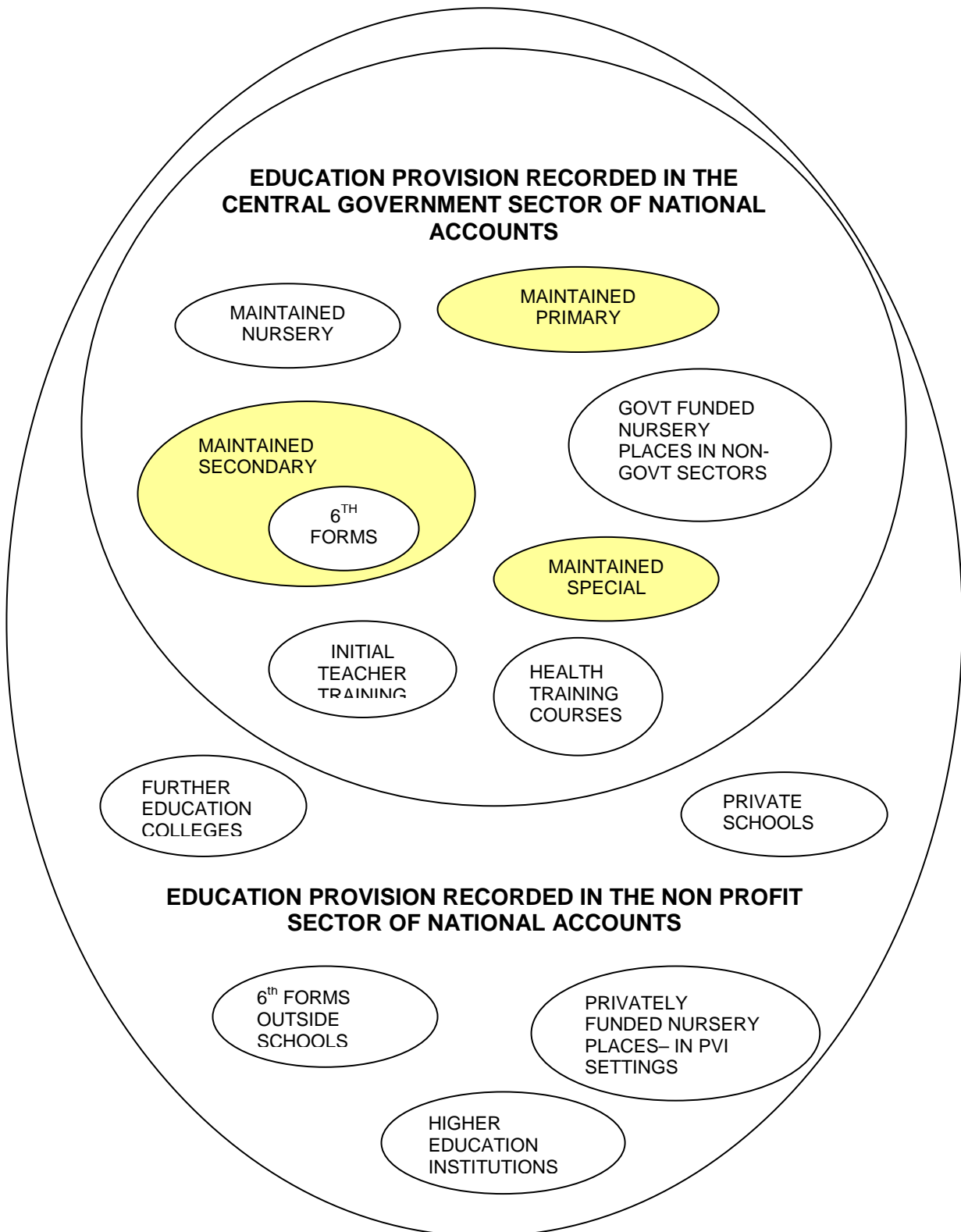
1.3 All five outcomes are important and inter-dependent. Importantly, they can also be affected by government intervention outside of education, and by factors outside of government activities. Different outputs from government provided education contribute to these five ultimate outcomes. For instance, time spent on sporting activities at school contributes towards ‘being healthy’ and ‘enjoying and achieving’. However, measuring the contribution that education makes to each of these outcomes is complex. Furthermore, not all of the outputs of education can be measured in the robust and consistent way required for National Accounting purposes – though this does not mean they contribute in any lesser way to the final outcomes.

1.4 Our aim is to identify measures of direct, quality-adjusted output which have statistical integrity, and which capture the essence of what education delivers. Given the complexity of measuring and attributing outcomes, the next best alternative is to identify and measure direct outputs that affect these ultimate outcomes.

Which education activities are we trying to measure?

We have talked about the National Accounts measuring the outputs of publicly provided education, but what does this mean? The figure overleaf shows all the education activities provided by the public and private sectors, and where they are reflected in the National Accounts. Activities in the large circle are publicly funded and covered in the central government part of National Accounts. These activities were within the scope of the Atkinson Review of public sector output measurement. The shaded activities are those we have been concerned with to date, and covered in this paper. Future work will look at the other (non-shaded)

central government education activities. The activities recorded in the non-profit sector include private education, but are mostly made up of provision by institutions which are not controlled by the government, some of which do receive funding from the State. Hence they are classified in the National Accounts as non-profit institutions serving households. These activities were beyond the scope of the Atkinson Review.

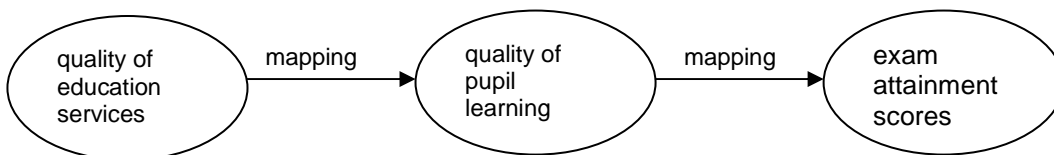


How can we measure the quality of education output?

1.5 Work to date has focused on how we can use existing data to measure the quality of education delivered. It has thrown up numerous conceptual and technical issues, which are summarised in this paper. Future work will start to focus on how we can measure the wider outputs identified earlier.

1.6 First, we have looked at using pupil attainment data to measure the quality of education services delivered: this is the most obvious and transparent measure of education output. The rationale for using attainment to measure the quality of education services delivered is that attainment depends on the quality of learning, which itself depends on the quality of education provided. We can thus use changes in attainment to proxy changes in the quality of education, although this is not without problems. Second, we looked at using measures of school effectiveness to capture more directly the quality of schooling provided.

Figure 1



CURRENT MEASURE OF EDUCATION OUTPUT

Overview of current measure

2.1 Output in the publicly-funded compulsory education sector currently comprises two components, one reflecting the volume of output and the other reflecting its quality.

2.2 In the National Accounts, and in accordance with international convention, education is viewed as an activity, so the volume measure captures activity levels using pupil numbers. On recommendation of the Atkinson Review, pupil numbers are now adjusted by school absence rates, to reflect the volume of education services actually delivered (or the amount of learning provided) by nursery, primary, secondary and specialist schools.

2.3 The quality of education output is currently based on the average improvement in exam results at the end of compulsory education; specifically, changes in GCSE average point scores between 1994 and 1997, in England. Since GCSE attainment reflects 11 years of schooling, the eleventh root of the annual percentage improvement in performance is taken. This reflects the ONS' current assumption that learning accumulates linearly over 11 years of schooling. The resulting quality adjustment factor is fixed at 0.25% per year.

Issues arising with current quality-adjustment measure

- being fixed it doesn't allow for actual year-on-year changes in quality;

- it is based on outputs at the end of compulsory schooling and doesn't reflect output or attainment at earlier stages in the education process;
- it ignores changes in the value of education output, which it is argued rises in an economy with rising GDP;
- it is possible to see a fall in education productivity (outputs over inputs) simply because education expenditure has risen. For example, a reduction in class sizes leading to increases in the quality of education would result in an apparent fall in productivity, whilst a doubling of class sizes would suggest virtually a doubling in measured productivity;
- finally, it raises the issue of the nature of the education process and how we should apply quality adjustments. The essence is whether learning is accumulated in an additive or multiplicative way and can significantly affect the resulting quality adjustment factor.

Atkinson Review recommendations

2.4 The Atkinson Review made three main recommendations for improving the current method, specifically:

- The **volume** measure should be based on actual school attendance rather than registered pupil numbers (already implemented);
- The **quality** adjustment should be: revised in terms of methodology and/or updated when new data becomes available; be based on a wider range of information; should take account of success rates for education provided to pupils aged 16 and over; and
- The measure should account for the changing value of education output in an economy with rising GDP, and we should consider ways of directly **valuing** public sector education output in monetary terms.

ALTERNATIVE ATTAINMENT-BASED MEASURES OF QUALITY

Measuring output at the end of compulsory schooling

Threshold attainment method

3.1 This method is very simple, and is based on the trend in the percentage of pupils who achieve a threshold level of attainment at the end of compulsory schooling. The threshold can in principle be set at any level. In the UK, we use the annual percentage of 15-year-olds in each year who achieve 5 or more A* to C grades in their GCSE exams (or equivalents).

Average Points Score Method

3.2 This method is similar to the threshold measure but is based on the average number of points that students get in each of their subjects at GCSE

level, rather than the percentage of pupils passing a threshold level of attainment. Pupils are awarded points for each grade attained in each subject, and the points are summed across all the subjects that they were enrolled for at the beginning of their GCSE exam year. The measure is based on taking the average of these points scores across all pupils in a cohort.

Issues arising from the GCSE attainment methods

3.3 Both of these methods improve on the current method because they reflect annual changes, rather than averaging over a number of years to obtain a fixed rate quality improvement. They also raise a number of issues, relating to how well even these improved measures capture the quality of outputs from the education system in any given year. These issues include:

- a threshold approach does not reflect changes over the full range of ability levels. Improvements in the attainment of low-achieving pupils will not be reflected if they move towards but do not cross the threshold, and vice versa at the high-achieving end;
- any attainment-based measures imply some strong assumptions about the link between attainment and quality:
 - Proportionality – assumption that a grade G (counted as 1 point under old system) represents 8 times more learning than an A (counted as 8 points). It implies that a change in attainment from a G to an A represents a quality change of 700%, this is an unlikely conclusion;
 - Arbitrary zero – assumption that zero GCSE points or 0% pupils achieving 5 good GCSEs equates to zero learning; it affects the extent to which attainment changes accurately reflect quality changes.
- scales for measuring attainment can change over time, as can the scope of what is counted in the attainment measures, thus leading to potential inconsistencies in the time series;
- they still only capture output for only one cohort of pupils in education in any given year (i.e. those sitting GCSEs) as a proxy for quality changes across the whole school system in that year;
- attainment at the end of compulsory education reflects a pupil's education over 11 years of schooling. Such measures are insensitive to concurrent changes in quality in the rest of the education system, and will take up to 11 years to feed through into the quality measure.

3.4 Despite the limitations, these approaches can be measured annually and are relatively simple. Furthermore, employers value qualifications, and on this basis the GCSE-based measures are appropriate output measures. Since employers value these qualifications through the wages they pay for them, in principle, there is potential scope for valuing these outputs directly.

Measuring output throughout the education system

3.5 Compulsory education in England is divided into four Key Stages, and pupils sit national examinations at the end of each phase. The Key Stages vary in length from 2 years to 4 years, with pupils sitting exams at ages 7, 11, 14 and 16. Pupils score National Curriculum points in these tests and are expected to progress by three points per year. Data on these examinations allows us to derive a more refined measure of the quality of education outputs based on more than one phase in the education process. Moreover, such rich data allows us to develop a measure which is closer to the concept of value added in the private sector, by measuring the progress made between the Key Stages.

Measuring cohort progress

3.6 We can measure average progress over an entire Key Stage by the difference between (i) the average points score for a given cohort of pupils at one Key Stage and (ii) their average points score at the next Key Stage.

Issues arising with a simple cohort progress measure (CPM)

3.7 The main advantage of this simple cohort progress measure is that it is based on the outcomes of 4 cohorts of pupils, not just one as in the GCSE methods. So it more closely measures quality of the whole education system.

3.8 There are, however, some disadvantages:

- Progress made by the other seven year groups is not covered;
- In the absence of data, we assume that pupils enter compulsory education with a fixed level of prior attainment (equal to nine points¹);
- Using cohort means ignores losses to or gains from the private education sector (or from immigration and death), as pupils usually switch to private education at the end of primary school. We could use matched individual pupil level data to explore how sensitive the measure is to using means or individual level data;
- Measured progress between the last two years of secondary education is dependent on the mapping we use to transform GCSE scores onto the Key Stage scoring scale. More exploratory work is needed to test how sensitive the results are to the transformation model used.

Adjusted cohort progress

3.9 One way of moving further towards a measure reflecting all 11 pupil cohorts in any given year, is to interpolate the progress made by the cohorts not

¹ The Curriculum System assumes that an average pupil will progress at a rate of one point per term. As the expected level of attainment for KS1 is 15 points, and KS1 is two years in length, or worth 6 points, we obtain a level of 9 points on entry by subtraction.

sitting exams from the results of those cohorts who did sit exams. To do this we have to make assumptions about how much progress is expected to be made in each year of a Key Stage. The simplest assumption is that pupils make even progress throughout each stage.

3.10 One drawback of this extension to the cohort progress method is that the adjustment process becomes less complete for the most recent years of a time series. For example, we won't know until 2007 the actual Key Stage 2 attainment of those in their first year of KS2 in 2004 – so we have no a priori basis for distributing their annual attainment backwards.

INSPECTION-BASED METHODS OF MEASURING EDUCATION QUALITY

Overall School Effectiveness Measure

3.11 Since 2000, inspectors at the Office for Standards in Education (Ofsted) have judged the 'overall effectiveness' of a school. This takes into account the school's strengths and weaknesses, the context of the school, and how it performs compared with similar schools. It is based on a careful review and assessment of all the evidence gathered about a school. Judgements are graded on a scale of 1 (highest score) to 7 (lowest).

3.12 We can use these inspection results to derive weighted average scores for primary and secondary school effectiveness; the weights being a reverse grading scale, so the highest scores are weighted by 7 and the lowest by 1.

Issues arising from the overall school effectiveness measure

3.13 A measure based on inspection results encompasses all judgements that inspectors make about a school. It has two implications. Firstly, the coverage of school outputs is wide – it will include not simply attainment, but teaching and behaviour and other things being assessed. Secondly, as the inspector's judgement combines all elements of a school in one metric, it is not necessary for us to weight together each dimension of school performance.

3.14 However, it also has a number of disadvantages:

- A weighting has to be applied to the grading scale. Although the scale is numerical it is not necessarily linear: a school graded a 2 is not necessarily twice as 'effective' as a school given a 4;
- The framework of assessment has changed a number of times. The overall effectiveness judgement has only been used since 2000, was changed in 2003 and again in 2005, to a 4 point scale. This highlights the problems associated with deriving time-series data;
- It is sensitive to the weighting used by Ofsted to gross up individual school results to a national level, as inspections are not based on a random sample of schools. When Ofsted present national figures they take into account the types of schools inspected that year, and aim to adjust to present a nationally

representative picture;

- The time series is not very long;
- In addition, the Atkinson Review Interim Report notes that;
 - Schools prepare thoroughly for Ofsted inspections and there may be an artificial climate, and no certainty that the criteria for teaching and school quality would be consistent across schools and time. (Although the reduction in the period of notice since September 2005 has to a large extent addressed this issue); and Ofsted's remit is not the same as needed for National Accounts, i.e. a measure of value-added by the school for individual pupils.

Combined Ofsted Measure

3.15 Ofsted also make judgements about individual elements of a school, so in principle, a composite measure of the quality of a school could be derived from combining results in the following categories:

- Quality of teaching;
- Pupil's achievement;
- Leadership and management;
- Personal development (including attitudes and behaviour).

3.16 However, there are limitations and due to the generic weaknesses outlined above we have not developed this measure further:

- it would require an arbitrary weighting between the different components. For instance, should teaching be judged to be more important, and thus have a greater weighting than leadership?
- disparity between a composite measure and the 'overall effectiveness' judgement. Although both are collected by the same inspection team they can produce inconsistent results. Inspectors make a judgement about the schools 'overall effectiveness' based on evidence which may not be wholly represented in the individual assessment categories.

VALUING EDUCATION OUTPUT IN MONETARY TERMS

Current valuation of education output in the National Accounts

4.1 We have been discussing education output in terms of a quality-adjusted volume measure, which is translated into an index with base year set to 100. Education output is also expressed in money terms in the National Accounts. In the base year, the value of output is set as equal to the value of education inputs (deflated education spending); i.e. output equals input. The quality-adjusted volume index is then applied to this base year value of outputs to derive a series for the value of education output. For example, in the 2004 National Accounts, general government final consumption expenditure on education in 2001 was

£33,480 million. The education output index grew by 0.319% between 2001 and 2002, so the value of output for 2002 was recorded as £33,587 million.

Alternative ways of reflecting the value of education output

Adjusting the volume of output in line with real earnings growth

4.2 Education output could be valued more accurately to reflect the value that society and the economy places on it - as is done for private sector output. Education – according to one theory – increases the productivity of the workforce, which is reflected in increased earnings and leads to rising living standards. The Atkinson Review concluded that the value of public sector output should be seen as growing in line with marketed sector output.

4.3 In the case of education, a qualified person produced in 10 years time has more productive potential than one produced now because of future technological progress. Since, historically, UK earnings have grown on average by around 1.5% per year the Review suggested that education output should be adjusted upwards by 1-1.5% per year, to reflect its rising value. Although it also suggested that this option should be given careful consideration, especially as no other country makes such an adjustment.

Direct valuation using estimated lifetime earnings

4.4 An alternative way of valuing education output more directly is through the labour market value of qualifications over the lifetime of individuals possessing them. In theory, this can be measured as the estimated lifetime earnings of individuals with different levels and combinations of qualifications. Such labour market valuation is useful in that it is an absolute and direct measure of the value of education output. It implicitly incorporates the quality of those outputs, because if this changed, so would its labour market value.

4.5 However, National Accounts need to reflect changes over time, so any approach would need to accurately reflect annual changes in the value of education – which is in practice likely to be difficult. A key limitation is the time lag needed to properly detect changes in the *lifetime* value of education outputs, as we have to estimate this over a 40 year working life.

CONCLUDING REMARKS

5.1 We have made significant steps in developing better approaches to measuring the quality (and value) of education output for National Accounting purposes. Our focus to date has been largely on the most obvious output – attainment – but even this is less than straightforward in this particular context and raises many issues which need further scrutiny. Moreover, there are many other important outcomes from education which go beyond attainment, such as better health, lower crime, better parenting, greater civic engagement and others. Measuring the changing quality of these wider outcomes – and the contribution of education - is the subject of an ongoing work programme to improve the measurement of education output.