CES Special Briefing

The Youth Cohort Surveys - How Good is the Evidence?

Special CES Briefing No. 38

May 2006

CENTRE *for* EDUCATIONAL SOCIOLOGY

A Research Centre within the **DEPARTMENT *of* EDUCATION *and* SOCIETY**

The Youth Cohort Surveys – How good is the evidence?

by Linda Croxford, Centre for Educational Sociology, University of Edinburgh

Introduction

Evidence-based policy-making should be based on high quality evidence, but how reliable are the data that are used? Regular surveys of young people aged 16-19 have been commissioned by government departments in England and Scotland over the last twenty years in order to evaluate the effectiveness of education and training provision. The surveys are known as the England and Wales Youth Cohort Study and the Scottish School Leavers Survey. This Briefing outlines some problems encountered when using these surveys to analyse change in young peoples' education and training over the last two decades, as part of a ESRC-funded research project - *Education and Youth Transitions in England, Wales and Scotland 1984-2002*. The Briefing describes how a lack of continuity and comparability in commissioning and design has reduced the value of the youth cohort surveys as an evidence base.

Key findings

- The youth cohort surveys are potentially a rich source of information about the experiences and outcomes of young people as they make the transition from compulsory schooling to tertiary education and/or labour market destinations.
- Education and training are investments that have long-term impacts, both on individuals and on society as a whole. Surveys that seek to understand changes in young peoples experiences and outcomes need to take a long view, and not focus simply on the short-term impact of the latest government policy.
- Comparable data over time are needed to analyse changes in young people's experiences and transitions, and the effects of the enormous changes in society, the economy, and education and training systems over the period since 1984.
- It is difficult to use the youth cohort surveys for analysing change over time because of changes in design and survey contractors, inadequate and inconsistent questions and coding as well as poor quality of documentation.
- Within Britain there are important differences between the national education and training systems of England, Wales and Scotland, and these may diverge further following devolution in Scotland and Wales. Comparative data are needed to assess the impact of these changes on young people's experiences, and to support policy learning within Britain.
- "Home-international" comparisons of differences between England, Wales and Scotland based on the youth cohort surveys are currently very difficult because of the inadequate sample size for Wales, and lack of comparability of timing and questions between the two surveys.
- The surveys would be more valuable if they were planned with a longer time horizon, and gave greater attention to comparability – over time and across the UK.

A potentially valuable source of information

The England and Wales Youth Cohort Study (YCS) and the Scottish School Leavers Survey (SSLS) are postal-questionnaire surveys that cover nationally representative samples of young people. They were designed as youth cohort surveys in the early 1980s, at a time when youth unemployment was rising, youth training schemes were being introduced, and policy makers felt the need to track young people's pathways from school to the labour market. Each youth cohort comprises young people who were in their final year of compulsory schooling at a particular point in time. The youth cohort surveys collect longitudinal data about these young people from age 16 to 18/19 (and for recent cohorts to age 23). The youth cohort surveys are thus, potentially, a rich source of data about young people's experiences, which we have attempted to use for a research project entitled *Education and Youth Transitions (EYT) in England, Wales and Scotland 1984-2002*, funded by the Economic and Social Research Council (ESRC). The project aims to bring together comparable data from the youth cohort surveys so that we can analyse change over time. (Annex 1 summarises the cohorts included in the time-series datasets for the EYT project.)

The youth cohort surveys are an important source of information about young people because they combine data on family background, school context, attitudes, truancy and attainment with information about subsequent moves to further and higher education, labour market and other destinations. Whereas other statistical sources, such as DfES reports on educational participation, can provide information on particular outcomes, the youth cohort surveys provide a rare opportunity to examine inequalities in relation to gender, social class, ethnicity and other sources of differentiation and to consider these in a wider context. Because the data are longitudinal, they also provide an opportunity to investigate the longer-term impacts on later outcomes, including links between inequalities in attainment and participation at age 16 and young people's destinations and outcomes at 18/19 (and at 23).

Long-term change

Education and training are investments that have long-term impacts, both for individuals and for society as a whole. For example, the increased provision of free public education after the war, and subsequent raising of the school leaving age to 15 in 1947, and 16 in 1972, has meant that the parents of school students have themselves experienced increasing levels of education. Therefore, surveys that seek to understand changes in young people's experiences and outcomes need to take a long view, and not focus simply on the short-term impact of the latest government policy. Raffe and Spours (forthcoming) draw attention to problems linked to insufficient policy memory in both England and Scotland, so that mistakes are repeated and learning from experience is reduced. Comparable data over time are needed to help provide an evidence base that places policy in the context of long-term change over the period since 1984.

Problems in using youth cohort surveys for analysing long-term change

The youth cohort surveys were set up with the primary aim of tracking the transitions of young people from the completion of compulsory schooling into further education, training and the labour market and were thus not designed as a time-series (although it would be possible to do so). Given the current design it is difficult to use the data for analysing change over time. Serious problems arise from the following:

- changes in design;
- the impact of competitive tendering and consequent changes in survey contractors;
- inadequate and inconsistent questions and coding;
- inadequate documentation of the procedures used in constructing datasets
- sample attrition due to non-response.

Changes in design. In 1992 the sample design of YCS and SSLS changed. Up to that time the sample for each YCS cohort was drawn from a sample of schools, with relatively large numbers of pupils included in the sample[1] for each school. This two-stage sample design was unsatisfactory, and was changed in 1992 to a simple random sample from all schools, with relatively smaller numbers of pupils per school. The samples of the England and Wales cohorts are also affected by the increasing number of schools who refused to take part, but the extent and effect of school non-response is not documented.

In Scotland the target sample for SSLS included all schools throughout the period, with random samples of pupils selected by birthdate. However, redesign of the SSLS in 1992 created more problems because it was made into a survey of school leavers, including those who left school from the S4, S5 or S6 year stages[2]. It was intended to retrospectively construct the cohort of young people who were in the S4 year stage in 1992 by taking data on S6 leavers from the 1994 leavers' survey, S5 leavers from the 1993 leavers' survey and S4 leavers from the 1992 leavers' survey. This design was very flawed, and in 1996 the SSLS was redesigned again as a cohort survey.

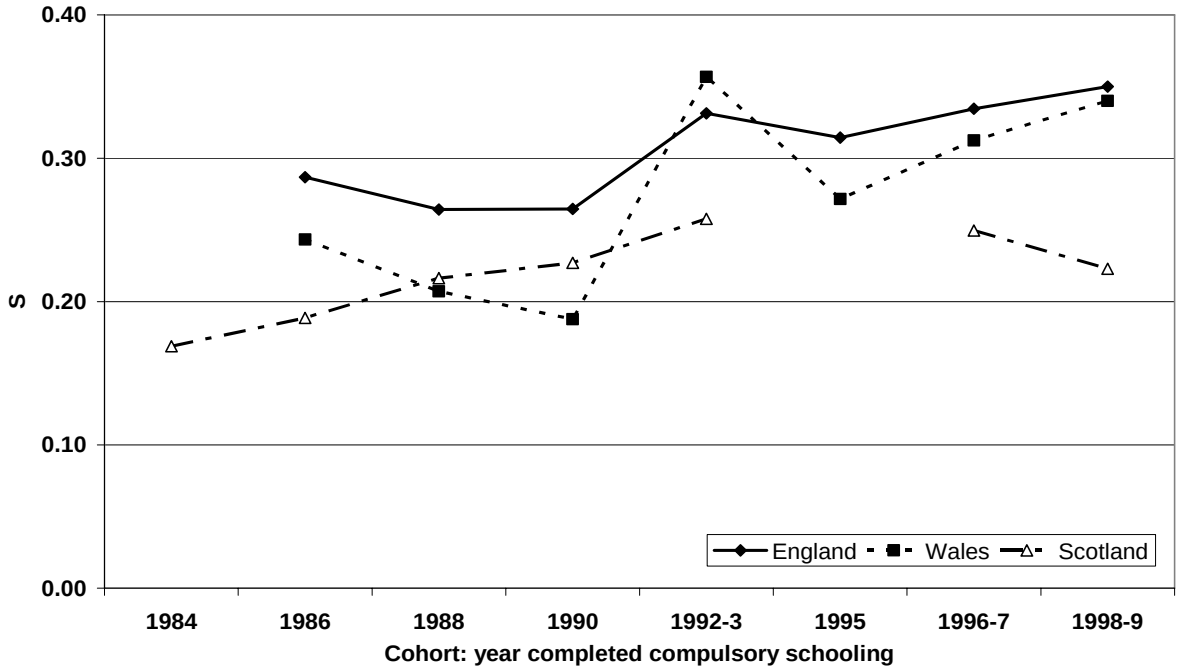
These changes in design of both YCS and SSLS have led to wide fluctuations in measures derived from answers to the surveys, and have created problems for interpreting trends over time, for example in respect of social segregation between schools (Croxford and Paterson forthcoming). Figure 1 shows an apparently sharp increase in segregation in England and Wales in 1992-3, which coincides with the change in sample design, while in Scotland the upward trend in segregation from 1984 to 1992-3 appears to have reversed from 1996 to1998, following changes in sample design: are these real trends, or merely artefacts of changes in samples?

---

[1] The sample of pupils was selected according to birth-dates, for example a 10% sample could be drawn by selecting pupils whose birthday fell on the 5th, 15th or 25th of any month. The aim was to create a "random" sample of pupils, that is a sample in which all members of the population have an equal chance of selection.

[2] The S4, S5 and S6 year stages in Scotland correspond to the Y11, Y12 and Y13 stages in England and Wales.

Figure 1: Segregation Index, Working Class SEC



The impact of competitive tendering. Between 1984 and 1991, both YCS and SSLS were planned as regular survey series and included substantial academic input to the design. Work on YCS was shared between Sheffield University and the survey organisation that is now called the National Centre for Social Research (NCSR). The SSLS was designed and conducted by the Centre for Educational Sociology (CES) at Edinburgh University. In each case the survey teams built up substantial expertise in issues relevant to surveys of young people. From 1992 competitive tendering was introduced, academic input to design greatly reduced, and the survey conduct and analysis functions separated.

The process of competitive tendering has eroded the quality of data. Currently, each survey sweep of the YCS is put out to competitive tender, so there are problems in continuity of questionnaire design and coding. For example, some survey contractors use manual coding to code the occupation details of young people and their parents, while other survey organisations use various computer packages, and these different approaches give slightly different results. These differences in methods of coding create inconsistent trends when the data are analysed - whether in a government department or academic institution – but the causes of the inconsistent trends are hidden. It is impossible to interpret trends accurately from inconsistent data sources. If the analyst finds change in an outcome, such as a reduction in social class inequality, how can s/he know whether it is real or merely an artefact of changing survey contractors? (For example Croxford and Raffe 2006.)

In Scotland, the competitive tendering process resulted in the design and conduct of the SSLS being taken over in 1992 by an English-based survey organisation. This led to serious flaws because their survey team did not understand crucial differences between the Scottish and English education systems. In particular, they assumed that school leaving arrangements in Scotland were the same as in

England, and designed questionnaires that failed to ask more than one third of the Scottish S4 cohort about their post-S4 destinations. This seriously reduced the value of the whole survey.

Inadequate and inconsistent questions and coding. Ideally, each survey in the time series should include the same core questions, coded in the same way, so that the data are consistent over time. Sometimes changes are necessary, for example the introduction of new types of qualifications or training schemes make it necessary to include additional questions or answer categories. However, there are many instances of changes in questions and categories between surveys for no apparent reason; the following example is from YCS and shows changes in question wording in different cohorts.

1984, 1986 *"Which of the following apply to your parents (or step parents)?"* (seven response categories including "in a full-time job", "in a part-time job", "unemployed", "retired" and "doing full-time housework").

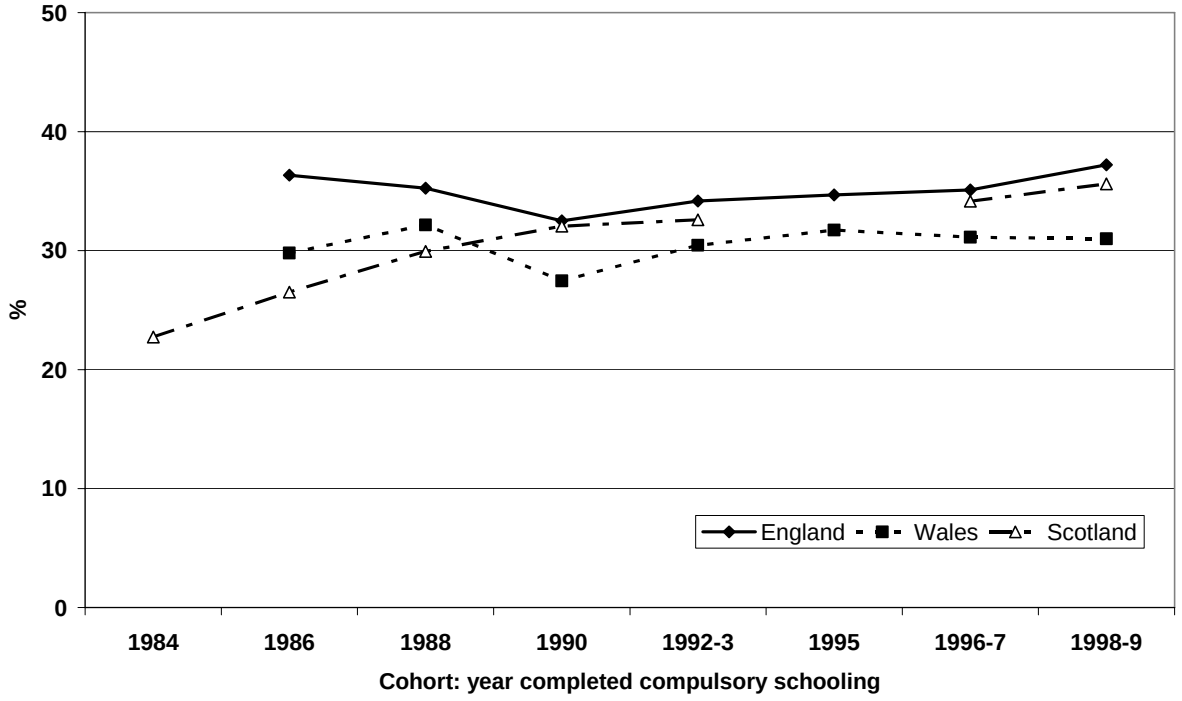
1988: *"Are your parents (or step-parents) employed (full-time or part-time) at the moment?"* (two response categories, "yes" or "no")

1990, 1993, 1995, 1999: *"Are your parents (or step-parents) employed full-time at the moment?* (two response categories, "yes" or "no")

1997, : *"Is your father/mother employed full-time at the moment?"* (two response categories, "yes" or "no")

Questions relating to further and higher education are inadequate and inconsistent over time, and are not coded in sufficient detail to allow analysis of the effects of institutions changing their designation – they do not distinguish "former-polytechnics" from "pre-1992 universities". Such differences in the status of institutions need to be analysed to accurately identify inequalities in access to higher education.

Figure 2: Parents in managerial & professional SEC (%)



Real world changes in occupational structures have necessitated changes in occupational coding and social-class classifications (see Annex 2). Such changes do not in themselves make it impossible to incorporate social class data in a time series, so long as the coding is carried out in appropriate detail and quality, and is well documented; for example, for the Scottish time series we have been able to accommodate the changes in occupational classifications between 1980, 1990 and 2000. The problems arise when the surveys include only the summarised categories of socio-economic group (as did the YCS up to 1988) or omit employment status (as in the case of YCS 1997-99). Obviously, if detailed codes are not provided it is difficult to create comparable variables over time. This is illustrated in Figure 2 which shows the proportion of each cohort with managerial/professional parents – the lack of detailed occupation coding in the 1986 and 1988 YCS cohorts results in inconsistent trends compared with those found in Scotland.

Inadequate documentation of the procedures used in creating youth cohort datasets. Comprehensive documentation of the procedures used in producing datasets from publicly-funded surveys is vital for subsequent users of the datasets – whether researchers are in government departments or elsewhere. Documentation should cover key issues such as method of calculating design weights and method of occupational coding, method of creating derived variables, as well as variable/value labels and links to questionnaires. Existing documentation of separate YCS datasets (and SSLS from 1993 onwards) is very scrappy and does not cover the key issues listed above. Inadequate documentation makes subsequent analysis more difficult and, in particular, means there is little basis from which to investigate inconsistent results. Adequate documentation is even more important when changes in personnel and survey contractors mean that it is not possible to ask the people responsible for the surveys. There is a need for the relevant government departments to ensure their survey contractors provide better documentation in future, and it would be helpful if they invested in retrospective documentation of historical datasets where possible.

Table 1: Respondents as % of initial target sample

	England & Wales		Scotland	
Cohort	survey at 16	survey at 18	survey at 16	survey at 18
1984	69	43	81	50
1986	77	44	81	52
1988	71	41	77	50
1990	72	42	69	43
1993	66	30		
1995	65	41		
1996-7	65	28	68	39
1998-9	55	34	63	42

Sample attrition due to non-response is an increasing problem in the youth cohort surveys. Not only does it introduce bias into the results, but this bias may change as attrition increases, reducing comparability across surveys. Table 1 shows that respondents to the YCS surveys at age 18 averaged 40% of the initial target sample between 1984 and 1990, but were less than a third of the initial target

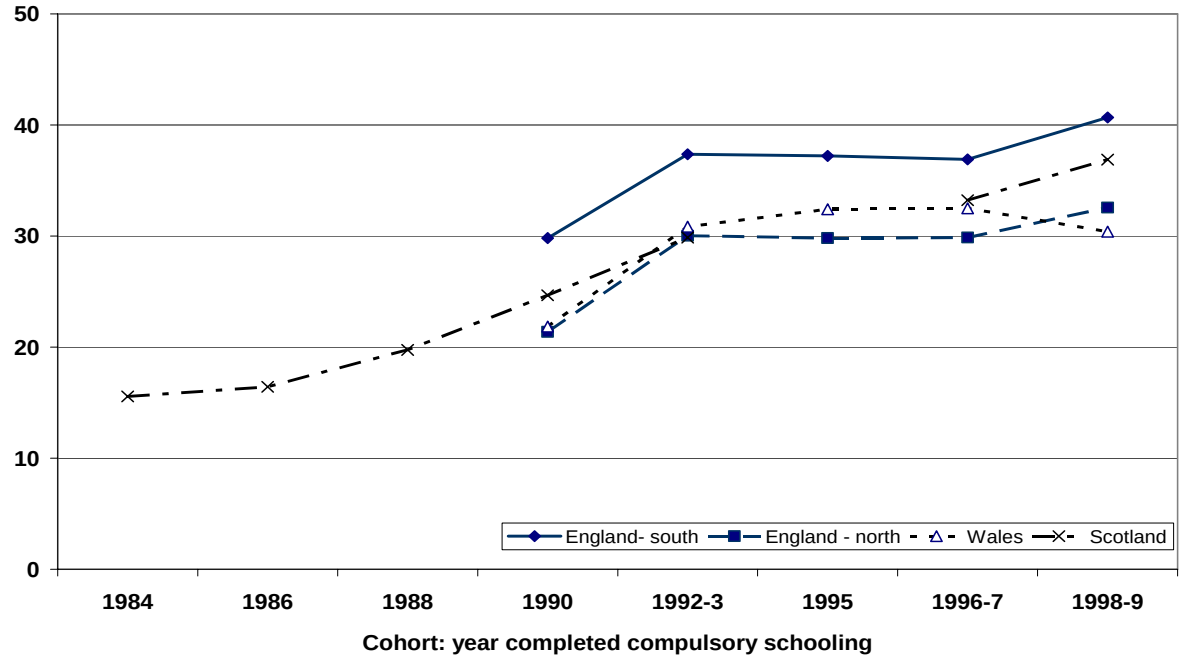
sample of the 1993 and 1997 cohorts. Young people from lower social classes, ethnic minorities and those with lower attainment are least likely to respond to surveys, and thus are under-represented in the data. In an attempt to compensate for these biases in the results, "weighting" variables are constructed from known characteristics of the sample and population, but are not wholly satisfactory, and, as stated earlier, documentation of the weighting procedures used in the surveys is inadequate.

Home-international comparison

Within Britain there are important differences between the national education and training systems of England, Wales and Scotland; education and training policy may diverge further following the creation of a Scottish Parliament and National Assembly for Wales (Raffe *et al* 1999). Comparative data are needed to assess the impact of these developments on young people's experiences, to support research on policy convergence/divergence, and to enable policy learning within Britain (Raffe and Byrne 2005, Raffe and Spours forthcoming).

Some home-international comparisons are intrinsically difficult because of actual differences in the real world. For example, the qualifications systems in Scotland are very different from those in England and Wales. In order to create comparable variables we have drawn on the frameworks developed by organisations that are working to establish equivalences between qualifications, such as those established for the Universities and Colleges Admissions Service (UCAS) and National Qualification Framework (NQF) as shown by Annex 3. Our approach is illustrated by Figure 3, which compares trends in achievement of NQF level 3 by age 18.

Figure 3: Achieved level 3 qualification by age 18 (%)



Note: Figures for 1993 cohort in England were depressed by survey error – some batches of questionnaires appear to have been printed without the questions on A-levels achieved.

Although the youth cohort surveys were not primarily designed for comparisons within Britain, when the YCS was first developed in the early 1980s the cohort design was coordinated with SSLS and common questions included to allow comparisons[3]. Over time the coordination of the surveys has lapsed and there are increasing difficulties in making comparisons between Scotland and the rest of Britain. Although the topics covered by the surveys remain very similar, the questionnaires gradually became more different, and since 1992 the timing of the surveys has not coincided (see Annex 1). For the sake of simplicity of presentation we have presented data for the SSLS 1992, 1996 and 1998 cohorts together with YCS 1993, 1997 and 1999 in charts and tables. More appropriate comparisons could be made if the relevant funding departments cooperated to arrange that cohorts surveyed in Scotland coincide with those in England and Wales.

There are further problems in comparison between Wales and the rest of Britain. The size of the sample for Wales is extremely small, in line with the relatively small size of the Welsh population, and consequently most comparative analyses find no significant differences between Wales and England because sample sizes are too small[4]. Comparative analysis within Britain will be even more difficult in future because the Welsh sample will be dropped from YCS from cohort 13 onwards.

Conclusion

Problems with the youth cohort surveys reduce their usefulness as a time series and for comparative analysis. These surveys are an important source of evidence for policy makers and researchers, providing nationally representative data on a wide range of issues relating to education and youth transitions. However, in recent years there has been a tendency to see them as 'market research', with one-off surveys of 'impact' rather than as a reliable research resource, constituted as a time series. Their primary purpose, at present, is to provide policy makers with information about whether the latest policy intervention has 'worked'. The problems discussed in this Briefing highlight the difficulty of reliance on evidence that is not based on long-term, consistent and comparable data. There is a risk that survey research is being used to 'authorize' policy interventions when the evidence is not secure. The youth cohort surveys are an important investment of public money, and with more attention to continuity and comparability, and investment in quality of coding and documentation, they would provide a reliable basis for evaluation of education and training and provision.

Acknowledgements

The support of the UK Economic and Social Research Council (R000239852) is gratefully acknowledged. However, the views expressed in this Briefing, and responsibility for any errors, are those of the author, and are not attributable to any funding agency or government department.

[3] The SSLS had been conducted as a survey of school leavers by the CES since the 1970s, and the CES shared experience of questions and coding with YCS designers at Sheffield University.

[4] The Scottish surveys have relatively large sample sizes because the target samples are 10% of the cohort (20% of 1998 cohort), and thus different trends can be identified (Croxford et al 2006).

Annex 1: Youth cohort datasets included in EYT time series

The datasets on which the Education and Youth Transitions (EYT) project is based come from nationally-representative cohort surveys:

- The England and Wales Youth Cohort Study (YCS), funded by the Department for Education and Skills (DfES)
- The Scottish Young Peoples Survey (SYPS, until 1991) and subsequently Scottish School Leavers Survey (SSLS), funded mainly by the Scottish Office/Scottish Executive.

Cohort (see note 1)	England & Wales (see notes 2 & 3)	Scotland (see notes 2, 3 & 6)
1984	YCS1 (1985, 86, 87) (see note 4)	SYPS85 (1985, 1986, 1987a)
1985		
1986	YCS3 (1987, 88, 89 + 94)	SYPS87 (1987,1989a)
1987		
1988	YCS4 (1989, 90, 91)	SYPS89 (1989,1991a)
1989		
1990	YCS5 (1991, 92, 93)	SYPS91 (1991, 1993a)
1991		
1992		SSLS Reconstructed Cohort (1993-5, 1999)
1993	YCS7 (1994,,96) (see note 5)	
1994		
1995	YCS8 (1996,,98, 2000a)	
1996		SSLS97 (1997, 1999)
1997	YCS9 (1998,99,2000s, 2000a +02)	
1998		SSLS99 (1999, 2001)
1999	YCS10 (2000s, 2000a, 2002)	

Notes

1. Each cohort is labelled by the year in which members reached age 16, or completed the final year-stage of compulsory schooling, which was Year 11 in England and Wales and S4 in Scotland.
2. Years shown in brackets indicate the year in which survey sweeps took place. The first cohorts had three annual sweeps, at ages 16/17, 17/18 and 18/19, but subsequent cohorts left out the sweep at 17/18 and added a further sweep at age 23 approx.
3. The letter "a" after a date shows the survey took place in the autumn of that year – all other surveys took place in the spring.
4. The 1984 YCS cohort did not include independent schools. Samples for YCS cohorts from 1984-90 were based on stratified samples of schools.
5. Samples for YCS cohorts 1991 onwards targeted all schools but encountered school-level non-response (and to deal with this there was further sub-sampling to make the sample representative by school-type, sex and region).
6. The Scottish cohorts included every secondary school in Scotland. Target sample for cohorts 1984-96 were 10% random samples, and 20% for 1998. The 1992 cohort was redesigned as three annual surveys of school leavers, from which the cohort was "reconstructed". There is a gap in the series because the 1994 cohort was not surveyed. The 1996 and 1998 were redesigned as cohort surveys.
7. Response rates for the YCS have declined from 77% in 1986 to 55% in 1999, and for SSLS have declined from 81% in 1984 to 63% in 1988.
8. Design weights have been derived by the survey contractors to compensate for non-response bias.

Annex 2: Social class variables created for the EYT time-series

The construction of social class variables for government surveys has changed following the ESRC review (Rose and O'Reilly 1998): the National Statistics Socio-economic Classification (SEC) has replaced the former Registrar General's Social Class. (SC) and Socio-economic Groups (SEG). In the light of this we attempted to derive SEC for the EYT time-series.

The derivation of SEC requires that occupation is coded to the detailed Standard Occupational Classification (SOC90 or SOC2000) and Employment Status. The National Statistics website provides full details of the method of deriving SEC, including down-loadable spreadsheets that give the appropriate value of SEC for each combination of SOC and employment status (http://www.statistics.gov.uk/methods_quality/ns_sec/).

There are a number of problems in deriving SEC for historic surveys:

- YCS cohorts from 1984 to 1988 did not include detailed occupational codes. We derived an approximate 4-class SEC from the existing SEG variables, but these are very unsatisfactory. The 1986 cohort did not ask about parental occupation until sweep 2, so the variable is missing for non-respondents.
- Scottish cohorts from 1984 to 1988 have detailed occupation codes, but these are classified by the pre-1990 Classification of Occupations (CO80). There is no official recode from CO80 to SOC90, but an approximate mapping is available, and we used this to recode occupations of the early Scottish cohorts to SOC90.
- Information about employment status in the youth cohort surveys is very limited and inconsistent. The only measure of employment status consistently asked in YCS and SSLS is about self-employment, with answer categories "yes" or "no". Consequently, our method of deriving SEC was to select the "simplified" or default SEC for each category of SOC90, and in cases where the answer to "self-employed" was "yes" we selected the value of SEC for "self-employed with no employees".
- There are large numbers of cases for which it is not possible to derive SEC because of insufficient or missing information. These cases we have coded "unclassified".
- Different survey contractors have different methods of occupation coding – some manual and others electronic – which caused differences in distributions of social class groups. As secondary analysts we have no way of overcoming this problem, but the government departments should require a higher quality of coding from their survey contractors.

The measure of parental SEC used for the EYT analyses takes either mothers' or fathers' SEC, whichever is the highest status. It has four values:

1. Managerial and Professional

2. Intermediate

3. Working Class

4. Unclassified

However, we must make it clear that the derived social class variables are not as robust or reliable as we would wish because of severe limitations with the survey data.

Annex 3: Equivalence between qualifications for the EYT time series

For the EYT time series we have attempted to derive variables on qualifications that are comparable across cohorts and national systems. The task was complicated by:

- changes over time in qualification systems, especially introduction of GCSE and SCE Standard Grade after 1986;
- differences between national qualification systems, especially post-16 qualifications;
- differences in survey questions and data sources between surveys and cohorts.

Age 16 qualifications

In order to derive variables that are compatible over time and national systems, we have assumed formal equivalences between the following levels of award.

England and Wales				Scotland		EYT time-series summary variables	
GCSE	GCE	CSE	16+	SCE Standard Grade	SCE Ordinary Grade	Award at A-C	Point score
A, A*	A		A	1	A	yes	7
B	B		B	2	B	yes	6
C	C	1	C	3	C	yes	5
D	D	2	2	4	D		4
E	E	3	3	5	E		3
F				6			2
G				7			1

Age 18-19 Qualifications

In order to derive variables that are compatible over time and national systems, we have assumed formal equivalences between the following levels of academic award, based on the tariff developed for the Universities and Colleges Admissions Service (UCAS).

Tariff	A-level	CSYS/Adv Higher	Higher	GCE AS	Int 2
120	A	A			
100	B	B			
80	C	C			
72			A		
60	D		B	A	
50				B	
48			C		
42					A
40	E			C	
38					
35					B
30				D	
28					C
20				E	

Highest level of qualification

We have used the National Qualification Framework to derive measures of highest level of qualification. The levels are as follows:

Level	Academic quals	Vocational quals
NQF Level 3	▪ CSYS ▪ Advanced Higher ▪ A-level or AS level ▪ Higher	▪ SVQ 3, NVQ 3 ▪ GSVQ or GNVQ advanced
NQF Level 2	▪ Ordinary Grade A-C or 1-3 ▪ Standard Grade 1-3 ▪ GCE or GCSE A-C ▪ Intermediate 2	▪ SVQ 2, NVQ 2 ▪ GSVQ or GNVQ Intermediate
NQF Level 1	▪ Ordinary Grade D-E or 4-5 ▪ Standard Grade 4-5 ▪ GCE or GCSE D-E ▪ Intermediate 1	▪ SVQ 1, NVQ 1 ▪ GSVQ or GNVQ Foundation

References

Croxford, L. and Paterson, L. (forthcoming) Trends in social class segregation between schools in England, Wales and Scotland since 1984, (accepted for publication in *Research Papers in Education*).

Croxford,L. and Raffe,D (2006) Education Markets and Social Class Inequality: A Comparison of Trends in England, Scotland and Wales, chapter prepared for Teese, R. (ed) *Inequality Revisited,* (forthcoming).

Croxford,L., Howieson, C., Iannelli, C., Ozga, J., Raffe,D. and Shapira, M. (2006) Trends in education and youth transitions across Britain 1984-2002, Background paper for Conference: Education and Social Change: England, Wales and Scotland 1984-2002, to be held in Edinburgh on 12th May 2006.

Raffe, D., Brannen. K., Croxford, L. and Martin, C. (1999) Comparing England, Scotland, Wales and Northern Ireland: the case for 'home internationals' in comparative research, Comparative *Education,* 35(1), pp.9-25.

Raffe, D. and Byrne, D. (2005) Policy Learning from 'Home International' Comparisons. *CES Briefing No 34*.

Raffe, D and Spours, K. (2006) *Policy Learning in 14-19 Education,* Bedford Way Papers, University of London Institute of Education (forthcoming).

Rose, D. and O'Reilly, K. (1998) *The ESRC review of Government Social Classifications*. London: Office for National Statistics.

CENTRE FOR EDUCATIONAL SOCIOLOGY
UNIVERSITY OF EDINBURGH
ST JOHN'S LAND
HOLYROOD ROAD
EDINBURGH EH8 8AQ
SCOTLAND

TEL: 0131 651 6238
FAX: 0131 651 6239
EMAIL: ces@ed.ac.uk
http://www.ces.ed.ac.uk/