

An Audit of the Data Needs of the DfEE Centres for the Economics of Education and the Wider Benefits of Learning

**Anna Vignoles with assistance from Tanvi Desai and
Estela Montado**



**CENTRE FOR THE
ECONOMICS OF
EDUCATION**

November 2000

Published by
Centre for the Economics of Education
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

© Anna Vignoles, Tanvi Desai and Estela Montado, submitted October 2000

ISBN 0 7530 1432 7

Individual copy price: £5

Executive Summary

The DfEE has established two new research centres, the Centres for the Economics of Education and the Wider Benefits of Learning, both examining the impact that education and learning have on people's lives. To be able to undertake such research, high quality data sets are needed that relate to education issues. This paper describes some of the key issues involved in obtaining such data. These issues are:

- Data Quality; data must be of a high quality, so that it can be relied upon to provide accurate estimates. Particular issues are response rates, representativeness, and attrition if dealing with longitudinal data.
- Identifying Data Gaps; it is necessary to identify certain areas where high quality data are in limited supply or totally absent, so that such shortages can be rectified in the future.
- Data collection; following on from the previous point, the centres should be able to suggest new strategies for data collection, so that the data sets of the future may be more useful for answering the types of questions that need answering.
- Linking data; at the moment, different data sets exist at different levels, for example at individual, school or LEA level. In an ideal world data at these different levels would be available in the same data set. We therefore identify separate data sets that can be linked together.

After outlining these general issues, the main body of this paper goes on to describe, in the light of these issues, existing data sets that should be useful in three areas of work:

- the production of education and skills
- lifelong learning
- estimating rates of return.

An Audit of the Data Needs of the DfEE Centres for the Economics of Education and the Wider Benefits of Learning

**Anna Vignoles with assistance from Tanvi Desai and
Estela Montado**

Introduction	1
Section I – Key Data Issues	1
Section II – Existing Data Sources	2
Section III	2
Section IIIa. Production of Education and Skills	2
Section IIIb. Lifelong Learning	9
Section IIIc. Rates of Return	10
Conclusions	12
Table	13

The Centre for the Economics of Education is an independent research centre funded by the Department of Education and Employment. The view expressed in this work are those of the authors and do not necessarily reflect the views of the Department of Education and Employment. All errors and omissions remain the authors.

Acknowledgements

Anna Vignoles is a member of the Centre for the Economics of Education and the Centre for Economic Performance, London School of Economics. Tanvi Desai is Data Manager at the Centre for Economic Performance, London School of Economics. Estela Montado is a member of the Centre for Economic Performance.

An Audit of the Data Needs of the DfEE Centres for the Economics of Education and the Wider Benefits of Learning

**Anna Vignoles with assistance from Tanvi Desai and
Estela Montado**

Introduction

The objective of this paper is to identify the primary data needs of the two, newly established DfEE Centres and, to summarise the existing data sources. The paper therefore gives background information on the major existing data sources, as well as ideas for further exploitation of such data.

Section I discusses the main data issues facing the two DfEE research centres. In Section II we then provide a brief data audit of the key existing data sets (see Table 1). The audit provides basic information on the contents of the various data sets, as well as details of access issues, sampling procedures *etc.* Section III then tries to identify the data sets that will be particularly useful to address the particular research questions being investigated by the two DfEE research centres, and provides still more background information on certain data sets. To assist the reader, in Section III the main sources of data are organised into three conceptual categories: *the production of education and skills, life-long learning and the returns to education.* The first theme includes all data related to the production of formal education and skills in the different educational stages (primary, secondary, *etc.*). The second theme of life-long learning, includes any data set that gives information on any formal education and training acquired during adulthood. The third theme on the returns to education includes all data sets relating to the economic and non-economic returns to education.

Section I – Key Data Issues

Several data issues emerged, following discussions between the two centres, as being of particular importance.

- **Data Quality** - there is a clear need to check data sets against each other, and against other established sources of information, to gauge quality. Certainly response rates, attrition and sampling procedures vary between data sets. Hence a systematic evaluation of the available data is needed to provide some guidance on quality. To this end, some quality related information is included in this working paper. Specifically the paper indicates the type of data included (*e.g.* longitudinal), the coverage of the data (*e.g.* whether the data include family background variables and information at different levels pupil, class, school, *etc.*), as well as technical information on response rates and attrition problems. To aid researchers, this paper also indicates, where possible, key references relating to each data set.

- **Specific Data Gaps** - it has been noted that the data are very limited in some areas, *e.g.* data that can be used to study lifelong learning. Other areas of weakness are data on FE qualifications and outcomes, under-fives data, cost and expenditure data on education and linked student-teacher data.
- **Strategies for New Data** - it has been recognised that an important role for the two Centres is to influence new data collection and to identify gaps in data sources. For example, many existing data sets, which give information on schools, lack robust information on groups within schools (classes) and contextual data. This paper aims to provide the knowledge base, from which the two Centres can devise a strategy to influence new data collection.
- **Linking Data** - whilst in some instances new data collection will be essential, the usefulness of existing data sets could be increased by linking between different data sources. For example, one might use the longitudinal data sets as a base, and link in cross-sectional data, possibly at LEA level. This working paper indicates where linking might overcome existing data limitations, although access and confidentiality considerations may limit the extent to which data sets can be linked in practice.
- **Access** - this is a crucially important question that cannot easily be resolved. For example, it appears that the DfEE's access policy does not enable researchers to lodge cleaned DfEE data with the ESRC archive. However, it has been agreed that researchers from the two centres will collaborate so that access to a particular data set would only need to be agreed once with a representative from the two centres, rather than for every project. This will ensure that the burden on data producers is kept to a minimum. This paper indicates where there are possible access restrictions.

Section II – Existing Data Sources

Table 1 below summarises the main data sources that could be used to study a variety of educational issues. The table describes each data set, giving information on the coverage, methods of collection and other technical points, access issues and key references. Section III gives more detail for each data set.

Section III

This section is organised into the three themes discussed above; namely, the production of education and skills, life-long learning and the returns to education. Clearly there are many educational issues that do not fit neatly into these themes, or that may cut across more than one theme. Furthermore, the same data set may be used to address issues that fall under all three themes. However, we attempt to provide a useful categorisation of the data sets to enable researchers to identify suitable data sets, or indeed gaps in data.

Section IIIa – Production of Education and Skills

This section outlines the main data sets that might be used to investigate the production of education and skills. The key criteria for including a data set in this section are that it must contain information about family background, some information on the learning experience itself (*e.g.* class size) and some kind of outcome variable (*e.g.* qualifications acquired). We

also include Administrative data sets at the end of this section which may not contain all of these variables but which could be linked with other data sets that do.

We start by considering in more detail the large longitudinal data sets, which will continue to form the core data resource for much of the socio-economic and education research in the UK.

1. National Child Development Study (NCDS)

This longitudinal study follows the lives of those born in Great Britain between 3rd and 9th March 1958. The time dimension is 1958 - 1991. Follow-up sweeps were carried out in 1965, (when they were age 7), in 1969 (when they were age 11), in 1974 (when they were age 16), in 1981 (when they were age 23) and in 1991 (when they were age 33). In addition, in 1978, schools and further education colleges were contacted and information collected. The data set includes the following information:

At birth: background information on the parents, including education, social class and financial position of the family.

Children at age 7: medical information, educational attainment test scores, details of child's early life and environment, medical history, some schooling information including indicators of quality, such as class size.

Children at age 11: medical information, educational attainment test scores, educational progress and aspirations, activities outside school, hobbies, family living conditions and other background variables. Includes schooling information such as class size.

Children at age 16: medical information, educational attainment test scores, necessity for SET, educational progress/qualifications, home environment, living conditions, relationship between child and parents, behavioural difficulties, educational expectations and aspirations, career expectations and aspirations, spare-time jobs held, spare time activities, attitudes towards marriage, family size, sex education received at school, relationship with family, details of cigarettes smoked and alcohol consumption. Includes schooling information such as class size. Supplemented with external data on public exams results.

Individuals at age 23/33: mostly focused on labour market information but includes information on the respondent's participation on government special training schemes, apprenticeship and training, post school education.

The NCDS data set is therefore an extremely valuable resource, since it contains good individual level data on pupils and their outcomes, and also a significant amount of information on their schooling experience. Its main strengths are that it is longitudinal and extremely comprehensive. The 2000 follow-up survey of the NCDS sample will add further to the value of this data set for the purposes of educational research.

2. British Cohort Study, 1970 BCS70

This is another longitudinal data set that follows a cohort born in 1970, and it contains similar information to the NCDS. It consists of a birth survey, plus 4 follow-ups and 4 sub-studies. Data were collected at birth in 1970 and then follow-up surveys were conducted at ages five, ten, sixteen and twenty six. The survey therefore covers the period 1970-1996. As already indicated, there have also been four sub-sample surveys carried out. The first two were carried out in 1972 and 1973 (the British Births Child Surveys) and surveyed sub-samples of the original cohort at ages 22 months and 42 months. The sub-samples consisted of all twins in the original cohort, the small-for-dates and post-mature births, and a 10% random sample

of the original cohort. The third sub-sample survey was carried out in 1977 when 1917 non-respondents from the 5 year survey were traced and interviewed in an attempt to assess the effect of non-response. In 1991, when the cohort were aged 21 years, a 10% sample survey was carried out which focused on adult literacy and numeracy problems as well as the transition from school to work.

As the BCS covers much of the same material as the NCDS not all the key variables are detailed here. For the purposes of educational research, the survey includes parental background, health of child, child's ability test scores at ages 5, 10 and 16, school quality variables, including class size, *etc.*, details of respondents' educational qualifications and the skills they have acquired. The BCS shares many of the advantages of the NCDS, with the additional bonus of being more recent. Again the 2000 follow-up survey will add still further to the value of the data set.

3. Youth Cohort Study (YCS)

This is a longitudinal study of young people in England and Wales, designed to monitor the behaviour and decisions of a sample of young people aged sixteen and upwards as they make the transition from compulsory education to further or higher education, or into the labour market. To date the YCS covers nine cohorts and over twenty surveys. The first survey was carried out in 1985 and the most recent in 1998. The questionnaires have been designed, over the ten years, to be broadly comparable, but external changes and shifts in policy interest, have brought about changes - some minor, some fundamental. For Cohorts 1-6, cohort members were contacted by post three times, at yearly intervals, when they were 16-17, 17-18 and 18-19. For Cohorts 7, 8 and 9, however, the sweeps were carried out at two yearly intervals, instead of annually. Thus Cohort 7 Sweep 2 took place when the respondents were aged 18-19 whereas previous second sweeps have taken place at age 17-18. A fourth sweep for Cohort 3 was subsequently carried out, surveying cohort members at the age of 22-23.

For the purposes of educational research, the YCS is useful since it contains information about students' GCSE and other qualification attainments in school year 11; participation in full-time education, government supported training and students' labour market experiences. It also asks students about their educational aspirations at age 16, their experience of year 11 and about leaving education (contact with careers service, work experience, national record of achievement, youth credits, truancy). However, the data contain very limited school quality information, although this can be overcome with careful linking to other data sources. Certainly the school identifiers which exist in the data, will enable the YCS data to be linked to school information, for example from the OFSTED database or from Form 7 (see below for details of these data).

4. International Adult Literacy Survey (IALS)

This is a major cross section study and was the first multi-country, multi-language assessment of adult literacy. It was conducted in 8 industrialised countries and covered over 40,000 adults, over the period 1994-1996. Although the IALS data is not longitudinal, and therefore not as useful as the NCDS or BCS data sets, it does have the advantage that it was specifically designed to assess literacy and numeracy and is extremely useful for research into this issue.

The data set includes information on the educational programmes that have been followed by respondents and the qualifications they acquired, by type, sponsorship, duration and purpose. Most importantly, respondents' current reading and writing skills in their mother tongue were tested. Some background information is asked, including parents' or

guardians' education and socio-economic group. The survey also asked respondents' about the mathematics/reading/writing skills that they use at work (may be relevant for lifelong learning section).

5. A-Level Information System, ALIS

This is a National Survey of A level students, conducted by the University of Durham, which consults students about their experience of post compulsory schooling¹. The Basic questionnaire includes information on courses and grades achieved prior to A-level, full details of A level subjects studied, students' views on the different courses they've taken and what they expect to do after A-level and specifically, their preferences, in terms of career choice. Most importantly, the data set includes background information, such as parents' occupations and parents' years of schooling. Furthermore, the survey includes comprehensive information about the institution at which respondents are studying.

The extended questionnaire includes more detailed questions on students' educational qualifications, their other commitments (domestic and work related) and their views on the quality of their educational experience.

This data set is extremely useful since it provides rich information on the pupil's schooling experience, and, perhaps as importantly, the pupil's perceptions of their learning experience. However, the data is limited by the fact that there are no follow-up surveys of the students once they enter the labour market.

6. Survey of Postgraduates Funded by the Research Councils (SPG)

This is a survey of former postgraduates who were funded by the Research Councils and whose funding ended in the academic years 1987/8 and 1988/9. Although a niche data set, focusing on a relatively small education sector (post-graduate education) it has the advantage of following up students once they enter the labour market.

This data set includes information about postgraduate study and qualifications, and students' views about their post graduate studies. It also includes information about subsequent employment (sector and occupation), the extent to which knowledge and skills acquired during the course have aided career progression, and the effectiveness of the course in preparing students for employment. In addition the data set includes numerous background characteristics such as ethnicity and gender.

7. Universities' Statistical Record (USR)/ HESA

This data set consists of records of all undergraduate and postgraduates students, as well as data on HE institutions and staffing. It is a series of panels and is now held by HESA. The time dimension for the USR was 1972/3 - 1993/4, which covered UK universities only. More recent data of a very similar nature is held by HESA and covers all HE institutions.

Background variables include: age, sex; marital status, country/county of domicile, country of birth, occupation of parent or guardian, last full-time school attended, other full-time/part-time post secondary educational institution attended, GCE "A" level or Scottish certificate of education higher grade results, other entrance qualifications.

Educational information variables include: details of degree subject, institution of study, normal duration of course, type of course, year of course, date of enrolment, method of

¹ The authors wish to thank Professor Carol Taylor Fitz-Gibbon providing us with information on this data set.

study, qualification aimed for, source of fees, accommodation used. Also contains some institutional data and data on staffing.

Outcome variables include: qualifications obtained, class of degree, date of leaving, reason for leaving, first destination (labour market information).

Whilst this data set is useful because of the detailed information it provides on the HE experience, it only follows up students some six months after graduation and therefore does not provide a satisfactory longitudinal element.

8. Program for International Student Assessment (PISA)

This is a study, organised by the OECD, which started in the first quarter of 2000 in 32 countries. It aims to assess three 'literacy' domains of reading, math and science. There will be three cycles. The next two will be taking place in 2003 and 2006. The first cycle concentrates on reading, the second on math and the third on science. The sample size of 15-year-olds in school will be between 4k and 10k in each country with between 100 and 200 schools. In addition to the tests there is a student and school questionnaire collecting limited background information.

In the UK the study is being done by ONS and there is some enthusiasm to link the data with Key Stage test scores, but currently there are no resources to do this. There would also be a possibility that the sample could be re-surveyed at subsequent dates, again a possibility that is being currently explored by both DfEE research centres.

9. Schools Examination Results Analysis Project (SENAP)

These data include individual level linked information on pupil's GCSE and A Level scores, as well as details of the subjects studied and the school or college the pupil attended. Key stage 3 data for 1996/7 is also being entered, with the intention that students will be able to be tracked from KS3 onwards, through their O levels and on to their A levels. The data only cover English 15-20 year olds and English exam boards/awarding bodies.

With a sub-set of this data, the Institute of Education (IOE) is currently doing a study of 4 years of linked A-level and GCSE results, with about 750k students. The study is exploring issues such as points scoring systems, measurement errors and institutional differences. Potentially, information could be linked through identifying the institutions (schools, further education, *etc.*) and thus obtaining information related to small areas.

10. Qualification on Entry Data Set (part of FE Individual Student Record)

These data, which form part of the FE Individual Student Record, provide a potentially valuable resource on the achievements of students participating in FE. However, the data suffer from low response rates and it is difficult to track students as they progress through FE. The Further Education Funding Council currently holds the data, although the new Learning Skills Council is likely to have responsibility for this data in the future. The LSC may also, of course, change the format of the data.

11. Various Administrative Data Sets

OFSTED²

OFSTED collect a variety of school level data in the course of their inspection process. Furthermore, OFSTED inspections make a qualitative judgement of a school using a common

² Many thanks to Ros Levacic for this information.

set of published criteria. Inspection covers three areas; educational standards achieved by pupils at the school; quality of education provided; management and efficiency of the school. Specifically Inspectors also make judgements about the relationship between aggregate levels of expenditure and school outcomes, and provide an indicator of value for money. In addition the data contain the following: background variables include socio-economic variables taken from the population census, data on the proportion of pupils known to be eligible for free school meals. Attainment of pupils on entry is derived from LEA and school data, and so will vary in detail and quality between schools.

School level information includes expenditure per pupil and numerous other measures of school resourcing (full details are available from the “Headteachers form”).

The outcome data available include OFSTED judgements about school quality, examination results and some value-added measures of pupil progress.

The OFSTED data, when linked to other information about schools from Form 7, and to individual level pupil data on examination performance, should provide a useful resource to investigate the impact of school characteristics on pupils’ learning outcomes. However, as is the case for most official sources of data on schools, questions remain about the quality of the financial data available in the OFSTED data set.

Proposed Basic Common Data Set

It has been proposed that an attempt be made to link the large number of administrative data sets that relate to education. To link all these data, each pupil will be given a unique identifying number, and will be tracked through the education system. The linking of these data sets will provide a number of very significant advantages. Primarily, it will facilitate proper panel studies using pupil level data on prior attainment and eventual educational performance.

It is still not clear exactly what this data resource will contain, however (Mayston and Jesson, 1999³). However, from the information given to us, it appears that once the CBDS comes fully on line, with unique pupil identifiers linking each pupil’s achievement through the system and *then* once these linked pupil achievement records have been merged with Form 7, OFSTED and data on school returns to LEAs etc, this will provide high quality performance data for research purposes. However, the CBDS will not overcome the key problem faced by researchers, namely poor data on school resourcing. The lack of good financial data in particular, will continue to limit researchers’ ability to reliable provide evidence on the relationship between resource allocation and performance between schools. We note that the CBDS was never expressly designed to provide evidence on the allocation of resources within schools, and it may therefore not enable researchers to evaluate the impact of, for example, expenditure per pupil.

The CBDS will also provide only limited information on teachers. This is because a) current information from the DfEE suggests that it will not be possible to identify each individual pupil’s teacher from the records in the CBDS and b) even if one can link pupils and teachers, the CBDS will not contain sufficient detail about each teacher to enable researchers to investigate issues such as the impact of teachers’ education level on pupil outcomes. There are however, ways around this problem. Using unique teacher identifier numbers and merging information from the Database on Teachers Records (DTR – see Table 1) would give researchers full background information on each teacher, including education level. It is clear though that this merging would be resource intensive, given the current

³ D. Mayston and D. Jesson (1999), *Linking Educational Resourcing with Enhanced Educational Outcomes*, Department for Education and Employment Research Report RR179.

format of the DTR. Updating the DfEE Schools Staffing and Curriculum Survey (see below) and again using unique teacher identifier numbers to merge it in with the CBDS/Form 7 data would perhaps be an easier way to proceed.

Form 7

Form 7 is essentially an annual census of schools. This data set provides quite detailed information on the teaching resource input into individual schools, and on some limited pupil characteristics – although data on pupils are currently only available at the school level. Specific variables contained in Form 7 include:

Pupil characteristics – age, gender, ethnicity, eligibility for free school meals, pupils with SEN.

Teaching resource input – numbers of teachers, number of job vacancies within the school, hours worked per week by staff, distribution of class sizes during census time period.

Secondary Schools Curriculum and Staffing Survey

These data come from a survey carried out by the DfEE on a random sample of around 550 secondary schools. Data is available for 1992 and 1996 and actual sample sizes are around 300. The data would have to be linked to other sources (see Basic Common Data Set above), since it does not contain students' prior attainment or indeed any educational performance information, nor information on pupils' backgrounds. Variables include the proportion of students studying named subjects by year group and the age, gender and qualification level of teachers taking each subject.

LEA Data – previously R01 Education Form and soon to be LEA Section 52 Outturn Statement

These data provide good LEA level data on educational expenditure levels, broken down into the following categories: teaching staff, support staff, other staff, equipment and materials, premises, other. As detailed in a recent report by Mayston and Jesson (1999) however, the existing data on individual school expenditure (from LEA sources) is less robust. This situation will improve once the new Section 52 budgeting arrangements come into force. These new arrangements will provide more detailed LEA level data as well.

Audit Commission Data Set

The Audit Commission collected some comparative information on a sample of schools over the period 1993-1996. This has been extended to 1997/98 for 1500 schools. An update for 1998/99 is also promised (see website). The data set contains the following:

Teachers' salary costs, pupil teacher ratios, teachers' time allocation, expenditure and income information, headteacher's position on the pay scale *etc.*

National Information Exchange

The Headteacher of Harrogate Grammar School set up this data set/data exchange in 1955 (Kevin McAleese – see Mayston and Jesson, 1999). It now includes comparative expenditure data (and income information) on more than 60 secondary schools. In particular it holds

detailed information on staffing costs. Clearly there are drawbacks in that this data set is not nationally representative and the sample size is small.

Funding Agency for Schools (no longer in existence)

The Funding Agency for Schools, which established this database, was, until 1999, responsible for the educational performance and use of resources by Grant Maintained Schools. The data set is useful in that it links together several administrative sources, including Form 7, examination performance data and other administrative information. It also has a panel element. However, the data is still only available at an aggregated level (school level). Furthermore, the sample is obviously not representative of all schools (see Mayston and Jesson, 1999 for a full discussion). The data set contains the following:

Pupil numbers, % eligible for free school meals and who are SEN, number of staff, size of premises, expenditure (on staffing, equipment, buildings, capital etc) and income.

Section IIIb – Lifelong Learning

This section looks at the data sets that might be used for a programme of research investigating lifelong learning. It is obvious that this is an area in which data are very weak. The key criterion for inclusion in this section was that the data set needed to have information on *when* training or education was acquired, so that adult learning might be investigated.

1. National Child Development Study (NCDS)

See Table 1 and Section IIIa for a description of the data set. It contains information on when qualifications were acquired (i.e. post school education - at age 33 - education and training, participation on government special schemes, apprenticeship and training. However, it suffers from the fact that only relatively small numbers of adults will have acquired education and training later in life, and hence this will limit any quantitative analysis.

2. British Cohort Study, 1970 (BCS70)

See Table 1 and Section IIIa for a description of the data set. The BCS70 has information on when education and training was acquired but like the NCDS, suffers from small sample sizes.

3. British Household Panel Study (BHPS)

This is an annual survey of each adult (16+) member of a nationally representative sample of more than 5.000 households, or approximately 10.000 individuals. It covers the time period 1991-98. The same individuals are re-interviewed in successive waves and, if they split-off from the original household, all adult members of their new household will also be interviewed. Children are interviewed once they reach the age of 16; there is also a special survey of 11-15 year old household members from Wave Four. Thus the sample should remain broadly representative of the population of Britain as it changes through the 1990s.

Information on the education acquired from wave to wave is known, so it is potentially a source of information on lifelong learning. The data also include some family background and personal characteristic information (gender, age, marital status, lifetime

relationship history, other demographic and health variables) and outcome information, particularly on labour market activities.

4. Universities' Statistical Record (USR)/ HESA

See Section IIIa for a description of the data set. It contains information on the age of students and whether the student is undertaking continuing education. Hence this data set may be useful for the purposes of looking at higher education lifelong.

5. Labour Force Survey (LFS)

This large-scale survey is discussed further below, since it is most useful for rates of return analysis. It may be of some use for analysing life long learning, although it has only very limited information on when individuals acquire their education/qualifications.

Section IIIc – Rates of Return

This section includes all the major data sets that might be used for a programme of research into the economic and non-economic returns to education. As for Section IIIa, the criteria for inclusion in this section is that a data set must have some family background information on the individual. In addition details of the educational qualifications acquired and subsequent labour market and non-economic outcomes must be available.

1. National Child Development Study (NCDS)

See Table 1 and Section IIIa for a description of the data set. These data, as has already been discussed, contain in depth information about individuals' family background and schooling. In addition, the data set includes individuals' employment history (sampled at ages 23 and 33) and economic outcomes such as family income, savings and investments. It also contains information on non-economic outcomes, such as housing, household details, health, malaise scores, parenting, and progress of any children (including ability test scores), political behaviour, voluntary activities, leisure activities *etc.*

2. British Cohort Study 1970 (BCS70)

See Table 1 and Section IIIa for a description of the data set. These data, like the NCDS, contain good background and educational information. Furthermore, at age 26, data were collected on employment, earnings, unemployment and periods out of the labour market. It also contains similar information to the NCDS on non-economic outcomes, such as housing, household formation, parenting skills *etc.*

3. British Household Panel Study (BHPS)

See Section IIIb for a description of the data set. It contains good information on respondents' employment history, earnings, occupational status and household finances. The BHPS also has some information on non-economic outcomes, including relationship history and marital status/household formation, housing situation, residential mobility, neighbourhood variables, measures of consumption, health, opinions and values of

respondents, especially their views about topics such as law and order, politics, jobs, marriage and family *etc.*

4. General Household Survey and Continuous Household Survey (GHS-CHS)

The General Household Survey (GHS) is an annual national survey, conducted quarterly by the ONS. It covers the time period 1973 - 1996/7 and includes approximately 8.000 households and 16.000 individuals. It is a multi-purpose survey, giving information on housing, employment, education, health and social services, for example.

The Continuous Household Survey (CHS) covers the period 1985 - 1991/92. It samples approximately 1% of households in Northern Ireland each year, giving a sample size of 3000. It is designed primarily to meet the information needs of government departments in Northern Ireland. The CHS is modelled on, and is similar in form and content to, the General Household Survey.

Both data sets contain information on a number of economic and non-economic outcomes, including various aspects of employment and unemployment, housing and migration, housing costs, ownership of consumer durables, job satisfaction, leisure activities and medical status (consultations with doctor, use of health and social services).

5. International Adult Literacy Survey (IALS)

See Table 1 and Section IIIa for a description of the data set. This data set has information on a number of economic outcomes; household income, full-time or part-time status, hours worked, employment status, industry, occupation and skills used at work (mathematical, reading and writing).

6. Labour Force Survey (LFS)

This large-scale survey contains both a cross-section and a panel element. It is useful in that it adheres to internationally agreed definitions of employment and unemployment and economic inactivity, and can be compared to similar surveys in other European countries. The survey was carried out every two years between 1975 and 1983, every year between 1984 and 1991, and since 1992 on a quarterly basis. The sample size is approximately 60.000 households and 120.000 individuals.

This survey is particularly useful because it contains some background information, educational qualifications and good quality information on economic activity and particularly income. It also contains some non-economic outcomes, such as household formation, housing information, health (particularly as it pertains to difficulties at work) *etc.*

7. Survey of Postgraduates Funded by the Research Councils (SPFRC)

See Table 1 and Section IIIa for a description of the data set. This survey includes details of these postgraduates' employment, including details of remuneration and with particular attention to their first substantive job and their current/most recent job since they finished their postgraduate studies. Since the data also include very detailed information on the post-graduate qualifications acquired by each respondent, it will be very useful for looking at differences in the economic outcomes between postgraduates. However, since the survey only includes postgraduates it has a limited applicability and cannot address wider questions about the rate of return (economic/non-economic) to different qualifications.

Conclusions

This paper serves as a resource to researchers who wish to undertake educational research that relates to either the economics of education and/or the wider benefits of learning. Table 1 in particular, serves as a brief audit of the available data. The commentary provided in this paper should also assist the reader in making choices about the most appropriate data to use for a particular research question. Inevitably there will be numerous sources of data that may be useful in this field of research that are nonetheless not mentioned here. In addition this resource will need to be kept updated, if it is to be useful to researchers in the future. For example, once the new Millennium Cohort Survey comes on line, this will provide an extremely valuable new data source for children born in 2001, similar to the NCDS and the BCS. Obviously the data audit will have to be updated to reflect such developments. Hence anyone wishing to contribute to the development of this resource should contact the *Centre for the Economics of Education at cee@lse.ac.uk.*

1.1.1 Table 1: Main topics, sampling and data collection procedures, related studies, limitations

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
A-level Information System (ALIS)	<p>This is a national survey of A level students, conducted by the University of Durham, which consults students about their experience of post compulsory schooling. The basic questionnaire includes questions on GCSEs courses taken, grades achieved prior to A-level, mother tongue, ethnic group, students' views of the courses they've taken and their career expectations and some family background information.</p> <p>The extended questionnaire includes questions on additional qualifications and other commitments during study (domestic and work related) and students' satisfaction with their educational experience.</p>		Survey.		
The Careers Service Activity Surveys	<p>These surveys provide information on the early career choices made by young people in England, after completing compulsory schooling. The surveys give some background information on pupils, such as gender and ethnicity, as well as details of the jobs and training options chosen by school leavers.</p>	No sampling used.	Careers Service database of information.	<p>The Careers Service, Department for Education and Employment (1999) Moving On - 1998: Pathways Taken by Young People Beyond 16, REF: MO98, ISBN: 1 84185 077 2.</p>	
1970 British Cohort Study	<p>The 1970 British Cohort Study (BCS70) began in 1970 when data were collected about the births and families of babies born in England, Scotland, Wales and Northern Ireland in the week 5th-11th April 1970.</p>	All children born in one week in 1970.	Postal and face to face survey. Attrition is an issue.	<p>Butler, N.R. Golding, J. and Howlett, B. (1986) From Birth to Five: a study of the health and</p>	

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
	Part of the aim of this survey was to make comparisons with the 1958 National Child Development Study. With each successive follow-up of the cohort, the scope of the enquiry broadened from a strictly medical focus at birth, to encompass physical, educational and social development and eventual labour market success.			behaviour of Britain's five year-olds., Oxford: Pergamon Press. Ekinsmyth, C. and Bynner, J. (1994) The Basic Skills of Young Adults, Some findings from the 1970 British Cohort Study., London: ALBSU.	
British Household Panel Study	The main objective of the survey is to further our understanding of social and economic change at the individual and household level in Britain, to identify, model and forecast such changes, their causes and consequences in relation to a range of socio-economic variables.	Annual survey on a nationally representative sample of households. Same individuals re-interviewed at subsequent waves and followed into new households if necessary.	Survey.		Includes some special surveys of BHPS children.
Continuous Household Survey	The Continuous Household Survey samples approximately 1% of households in Northern Ireland each year. It aims to provide accurate information about the social and economic conditions of the population of Northern Ireland, on an ongoing basis, and thus provide a means of examining relationships between the main variables with which social policy is concerned and in particular of monitoring changes in these associations over time.	Modelled on the GHS. 1% sample of households in Northern Ireland.	Survey.		Specifically designed to meet information needs of government departments in Northern Ireland
Database of Teacher Records	Individual records of each teacher recognised as qualified in England and Wales, from date of award of QTS. Contains, pay, qualifications at time of	Not appropriate	Annual Returns from employers to the Teachers Pension Scheme. Some	Statistics of Education: Teachers: England and Wales: 1999 Edition, published by Stationery	Considerable amounts of dirty or potentially confusing data that need to be handled with care.

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
	QTS, age, sex, length of service, current service, school and LEA etc.		employers do not provide all records until after the data is extracted in a particular year.	Office. Publications also for earlier years.	Expert support is needed. The dataset is large, about 1.5m records of existing and out of service teachers and is accessed using Q Stat.
Eurostat Continuing Vocational Training Survey	These data cover continuing vocational training issues, such as cost, hours spent on training, and details on participants. Also included is data on training systems, participation rates and apprenticeships. The Continuing Vocational Training System (CVTS) was the first survey on continuing vocational training in enterprises carried out at the EU level in a co-ordinated form (outline questionnaire, common definitions, common recommendations with the respect to the fieldwork).	Random sample of enterprises.	Surveys with additional field work.		Also covers employees in trainer/non-trainer enterprises by industrial sector, proportion of enterprises providing training, type of training provided and details on participants (including data on training by gender and occupation).
Family Expenditure Survey	The UK Family Expenditure Survey (FES) is a continuous survey (of about 1 in 2000 of all United Kingdom households). It provides information about the household, household and personal incomes, certain payments that recur regularly (eg rent, gas and electricity bills, telephone accounts, insurances, season tickets and hire purchase payments). Respondents also maintain a detailed expenditure record for 14 consecutive days. The main purpose of the survey is to provide the weights for the United Kingdom Retail Price Index (RPI).	Random survey of households.	Information for most of the questions is obtained via interviews from the head of household or housewife, but certain questions of a more individual character are put to every spender aged 15 or over (or 16 or over from 1973 onwards).		There are some very early surveys available dating back to 1953/54.
Family and Working Lives Survey	The aim of this study was to examine people's working patterns in order to inform	Random sample plus booster sample (2000 adults) for	Interviews.		This study updates and expands on the Women

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
	<p>policy on equal opportunities and labour market flexibility. Contains longitudinal information on the ways in which people participate in the labour market, and in training. Main subject areas include: life and work history data; socio-economic, personal and household characteristics; housing history since leaving parental home; training and education including on and off the job training; detail on current job and key past events (e.g. redundancy); spells of unemployment; pension provision and retirement planning; details of benefits claims; childcare arrangements; caring for elderly and disabled adults; past and current disabilities.</p>	<p>minority groups.</p>			<p>and Employment Survey, 1980 (held at The Data Archive under SN:1746) conducted by the then Employment Department (ED) and Office of Population Censuses and Surveys (OPCS).</p>
General Household Survey (GHS)	<p>Annual cross-section survey carried out in Britain.</p> <p>Contains socio-economic data on a wide range of topics. The data of most interest are on education - age on leaving school, age on leaving last place of full-time education, type of educational establishment last attended full-time and qualifications obtained.</p> <p>Limited information on vocational training.</p> <p>Considerable amount of labour market information, for example, labour market status (employed, unemployed, inactivity etc) income, usual hours of work, tenure, industry, occupation, public/private trade union membership.</p>	<p>Since 1984, postcodes have been used as the primary sampling units.</p> <p>Postcodes stratified firstly by region, and then within regions by household and economic indicators.</p> <p>Households randomly chosen from within these strata, and then individuals randomly chosen from within households.</p>	<p>Face to face interviews.</p> <p>Occasionally, proxy interviews conducted when sampled individual not present at time of interview.</p>	<p>Annual report and summary of key results published by the Stationery Office (previously HMSO).</p>	<p>Useful for providing long time-series of educational attainment.</p> <p>Qualifications reasonably disaggregated, though not as disaggregated as currently found in the Labour Force Survey (LFS).</p> <p>Is the longest running survey to include both qualifications and income, and so provides the longest time series on the returns to education.</p> <p>Note that changes in the measurement of hours of</p>

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
	Usual demographic characteristics of the individuals included, plus family members. Detailed health information.				work make it difficult to obtain a consistent series for hourly earnings.
Graduates Data Sets	This series of three surveys was designed to track the labour market outcomes of individuals who graduated in 1960, 1970 and 1980 (the survey was conducted six years after graduation for each cohort). Although the 1986 survey is the most detailed, each survey was modelled on the previous one so there are consistent questions over time.	Random samples of cohort.	Mail surveys.	Dolton and Makepeace (1992) Dolton, P and Vignoles, A (2000) , <i>The Incidence and Effects of Overeducation in the UK Labour Market, Economics of Education Review</i> .	
Hampshire LEA (subset 1)	Variables include: Outcomes: Year 2 (KS1) scores Baselines: Entry to reception Background: SEN status, SEN need, attendance, income, teacher turnover, class size, non-teacher adults in class, school type, school make-up, pupil turnover in school	All Students	By LEA		
International Adult Literacy Survey (IALS)	Survey carried out in twelve countries (including Britain). Each respondent undertook tests to assess their literacy and numeracy skills. Additional questionnaire contains information on labour force status, wages, job characteristics, job requirements, education, parents, plus a very detailed	Different procedures in each country. In Britain, all individuals of working age stratified by region and socio-economic group, and clustered into postcode sectors and addresses.	Face to face interviews and self-completion of numeracy and literacy tests.	<i>Literacy, Economy and Society: Results of the First International Literacy Survey</i> , OECD and Minister of Industry, (Statistics Canada catalogue no. 89-545-XPE), 1995.	The comparability of the literacy and numeracy scores across countries has been questioned - see forthcoming ONS report by Harvey Goldstein. Using the data in a Britain-only context, however, should

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
	section on vocational training.			<p><i>Literacy Skills for the Knowledge Society: Further Results from the International Adult Literacy Survey</i>, OECD and Human Resources Development Canada, 1997.</p> <p><i>Adult Literacy in Britain</i>, Carey, S., Low, S. and Hansbro, J., London: The Stationery Office, 1997</p>	provide some information on relative skills levels.
Labour Force Survey	The Labour Force Survey uses international definitions of employment and unemployment and economic inactivity, together with a wide range of related topics such as occupation, training, hours of work and personal characteristics of household members aged 16 and over.	1984-1991 in 1984 sample rotation was introduced along with a panel element, the quarterly survey, which uses a two-stage clustered sample design. 1992- Simple random sample: 4 sampling frames used; GB South of the Caledonian Canal, GB North of the Caledonian Canal, Residents in NHS accommodation, and NI. Households are interviewed on 5 occasions at quarterly intervals thereby introducing a panel element to the survey.	<p>Face-to-face interview: first interview;</p> <p>Telephone interview: subsequent interviews where possible.</p>	<p>[DASN: 33132] Labour Force Survey, 1975-1991.</p> <p>[DASN: 33286] Quarterly Labour Force Survey : Local Area Data, 1992-</p> <p>[DASN: 33298] Labour Force Survey Household Datasets</p>	Income data only available from 1992q4
National Adult Learning Survey (NALS)	Survey of people's experiences of and attitudes towards learning, intended to provide a basis for future monitoring of adult learning. In particular, the survey aimed to identify the extent to which people	Cross-sectional (one-time) study		This survey followed three earlier studies: Individual Commitment to Lifetime Learning : Individual's Attitudes	

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
	were taking part in different types of learning (both vocational and non-vocational), the cost of doing such learning, people's reasons for doing some learning, problems experienced, perceived benefits of learning and barriers towards taking part in learning.			Park, SCPR (1994) (held at the Data Archive under SN:3305); Monitoring Individual Commitment : Question Development Campanelli and Rutherford, SCPR, (1995); Feasibility Study for an Adult Learning Survey Lewis and Smith, SCPR (1996).	
National Child Development Study	Longitudinal survey of sample born in 1 week in March 1958. Followed up at ages 0, 7, 11, 16, 23, 33. Key topics include family and health background, education and schooling, early labour market experience, labour market progress at age 33.	Longitudinal follow-ups of original cohort.	Face to face and postal surveys	Ferri, E (ed). (1993). <i>Life at 33: The Fifth Follow-up of the National Child Development Study</i> , National Children's Bureau and City University.	
Newcastle Alumni	This data was commissioned and supplied by the University of Newcastle-upon-Tyne, and funded by Peter Dolton and the Alumni Office at the University of Newcastle-upon-Tyne, and the Government Office North East. The survey is designed to track the labour market progress of Newcastle Alumni, focusing particularly on wages, mobility and skill utilisation.	Sample obtained from Alumni Office database.	Sample obtained from Alumni Office	See Graduate data sets also described in this table.	
OFSTED	The OFSTED database contains	The timing of schools	Data entered by	OFSTED (1995a),	Data include information

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
	information from the Headteacher's Form and Statement, and from the inspectors' judgement recording grades for the inspection schedule. It contains data for all schools inspected in a given year. Hence each school only appears in the database every fourth year.	inspected (and therefore the number of times the school is included in the sample) is decided on the basis of the school's performance.	OFSTED. Missing values for individuals is a problem but the data have 100% response rates for schools.	Guidance on the Inspection of Nursery and Primary Schools, London: HMSO OFSTED (1995b), Guidance on the Inspection of Secondary Schools, London: HMSO OFSTED (1995c), Guidance on the Inspection of Special Schools, London: HMSO OFSTED (2000), The Annual Report of HMCIS of Schools, London, Stationery Office.	Number of pupils on roll Pupil background variables (at school level) Inspection grades (for schools inspected since April 1996) Expenditure data from Headteacher's Form Staffing numbers Pupil attitudes/disaffection The Curriculum Pupil outcomes at school level
OME : Survey of Teachers Hours and Conditions	Cross section of teachers which focuses on their work hours and conditions in a given week.		Postal Survey.		At present the school codes are not being made available to facilitate merging with school data.
ONS Longitudinal Study (LS)	A longitudinal dataset based upon linked census data from the 1971, 1981 and 1991 Censuses and vital events (births, cancers, deaths) recorded by ONS for 1% of the population of England and Wales (approximately 500,000 individuals at any one time point). It offers opportunities for a wide range of analyses in the areas of sociology, demography, geography and epidemiology.	Sample based on individuals giving one of four birthdates at census or event registration. (The dates are equally spread throughout the year to minimise any seasonal bias.)	Linkage of census and event data routinely collected by ONS.	Many other countries (inc. France, the Scandinavian countries and the USA) maintain data linkage studies similar to the LS. See ONS/CLS Longitudinal Study Newsletter no. 20 (June 1999): Using the LS for making	The educational data in the LS is restricted to the details of "higher qualifications" obtained after age 18 recorded by the 1971, 1981 and 1991 Censuses. In 1981 only the level of highest qualification is recorded - in 1971 and 1991 the

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
			international comparisons, pp 2-4		level of EACH of the first six qualifications is recorded, along with the relevant subject area.
Programme for International Student Assessment (PISA)	The PISA study is organised by OECD and went into the field in the first quarter of 2000 in 32 countries. It aims to assess the three 'literacy' domains of reading, maths and science. In addition to the tests there will be a student and school questionnaire collecting limited background information.	Interviews.			
Qualification on Entry Data Set (part of FE Individual Student Record)	Data set reference (= student ID) Qualification on entry reference code Grade Date awarded	Proportion of those covered by main Individual Student Record	Low and erratic response rates on these items		Includes GCSEs etc but quality of information unclear: no attempt made to cross-reference to SENAP
St Thomas's UK Adult Twin Registry	The Twin Registry contains mostly medical data on identical and fraternal twins in the UK. In addition to a wealth of medical data, a recent questionnaire included socioeconomic, education and earnings questions.	Sampling procedures depend on the medical research project being undertaken at the time .	Face to face individual interviews/medical exams at the unit: Telephone interviews and postal surveys:		Income/education data only available for female twin pairs
Social Change and Economic Life Initiative	This survey was specifically designed to illicit respondents' assessments of the skills they use at work, and the educational requirements of jobs.	One-stage stratified random sample.	Mail survey.	Gallie, D. (1998) The Social Change and Economic Life Initiative, Working Paper 1, Oxford.	

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
Schools Examination Results Analysis Project (SENAP)	<p>(a) Individual level data matching an individual's GCSE and A Level scores and subjects + school/college affiliation, on all occasions when exams were taken: covers English 15-20 year olds and English exam boards/awarding bodies. Although all English boards covered since 1990, early years don't match backwards very thoroughly.</p> <p>(b) For cohorts since 1997, GNVQ results in the process of being added to GCSE and A levels.</p> <p>(c) For 1996/7 on, KS3 data is being entered, will be matched to GCSE and eventually A level/GNVQs. Only one GCSE cohort so far (i.e. last year's) that really useable.</p> <p>(Some N. Ireland data but no Welsh data held. However, Welsh data do exist and <u>may</u> include matched KS3 results in the near future)</p>		Data from exam boards	(Used by some Awarding Bodies for internal research studies.)	Ongoing study at Institute of Education looking at GCSE-A level value - added. The Institute study uses a subset of these data: results due out 2000
Skills Survey	The dataset contains information on people in work and their jobs. The topic areas covered include: job details; job analysis; attitudes and management skills; competence; transferability of skills, pay and qualifications; job held five years ago; demographics.				

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
Survey of Postgraduates	<p>A survey of former postgraduates funded by the Research Councils whose funding ended in the academic years 1987/8 and 1988/9. The aim of the survey was to supplement data on first employment destinations of students and to identify the benefits of postgraduate study to former students and the economy (in the widest sense) over the longer term. Data were specifically collected on individuals' employment outcomes and the impact of their course(s) on their outcomes.</p>	Cross-sectional (one-time) study	Survey.		Only covers postgraduates funded by research councils.
Universities' Statistical Record	<p>TheUSR consists of records of undergraduate students on courses of one academic year or more; postgraduate students on courses of one academic year or more; academic and related staff holding regular salaried appointments, and finance data for all UK universities.</p>	Full population survey.			
Workplace Employee Relations Survey	<p>The 1998 Workplace Employee Relations Survey (WERS 98) is the fourth in a series of surveys that began in 1980. Its aim to provide statistically reliable, nationally representative data on the current state of workplace relations and employment practices in Britain.</p> <p>Topics include recruitment and training; consultation and communication; worker representation; payment systems; equal opportunities; health and safety; flexibility and performance; change; attitudes to work</p>	<p>Simple random sample: survey of employees questionnaire; One-stage stratified or systematic random sample: cross-section sample for the management questionnaire</p>	<p>Face-to-face interview: management and worker questionnaire collected via CAPI; Self-completion: survey of employees questionnaire</p>	<p>WIRS80, WIRS84, WIRS90, WERS98, WIRSpane80-84, WIRSpanel84-90, WERSpanel90-98</p>	

Data Name	Description and main topics	Sampling procedures	Method of data collection/ response rate issues	References to related studies	Comments/ Limitations
Youth Cohort Study	<p>The longitudinal data set contains information on pupils' achievement (GCSE attainment) at school and their subsequent activities after age 16, including participation in full time education, labour market experiences and training programmes.</p> <p>Individuals are also asked about their experiences of Year 11 and their educational aspirations. Individuals are followed up at ages 16/17, 17/18 and 18/19. One cohort was also followed up at age 22/23.</p>	<p>Representative samples of 16 year olds.</p>	<p>Postal surveys.</p>	<p>Dolton, P., Makepeace, G., and Treble, J. (1994) The Youth Training Scheme and the School to Work Transition, <i>Oxford Economic Papers</i>, Vol 46, pp 629-657.</p>	<p>Now 9 cohorts of data.</p>

Table 2: Organisation with responsibility for data, access details, population, years covered, spatial unit and key people

Data Name	Organisation with Prime Responsibility for Data	Access Details	Population	Years covered	Spatial Unit	Key People
A-level Information System (ALIS)	University of Durham	Restricted - need to contact University of Durham.	A level students.	Annual.	Individuals	Carol Fitz-Gibbon, University of Durham.
1970 British Cohort Study	Centre for Longitudinal Studies at the Institute of Education	Available from the Data Archive at Essex University additional data available from CLS.	Approx 10,000	follow-ups in 1975, 1980, 1986 and 1996	Individuals	John Bynner
British Household Panel Study	Institute for Social and Economic Research	Available from the Data Archive at Essex University	Approx. 5,000 households, 10,000 individuals	1991-1998	Households/ individuals within households	ESRC Research Centre on Micro-social Change
The Careers Service Activity Surveys	DfEE	Not known.	17/18 year old population in England, as well as Year 11 completers.	Annual since 1980s Data is consistent since 1996.	Individuals	Brenda Alexander, DfEE
Continuous Household Survey	Office for National Statistics	Available from the Data Archive at Essex University	Approx 3,000 households	1985-1991/92	Households/ individuals within households	
Database of Teacher Records	DfEE	Analyses available from DfEE. Data only available to staff working for DfEE to	All teachers awarded QTS in England and Wales All staff who are or have been	Data extracted twice a year: Summer and late Autumn.	Individuals	John Pascoe, Andy Clarke, Dean Franklin.

Data Name	Organisation with Prime Responsibility for Data	Access Details	Population	Years covered	Spatial Unit	Key People
		ensure confidentiality.	members of the Teachers Pension Scheme All regular teachers in schools and full-time staff in FE			
Eurostat Continuing Vocational Training Survey	EUROSTAT	Available from the Data Archive at Essex University	European states.	1993-1996	Firms.	
Family Expenditure Survey	Office for National Statistics	Available from the Data Archive at Essex University	Approx. 5,000 households, 10,000 individuals	1961-1997/98, 1964-1967 currently inaccessible	Households/ individuals within households	
Family and Working Lives Survey	DfEE..	Available from the Data Archive at Essex University	Adults aged 16-69 years in Great Britain and their partners. Target was 12,000 adults.	1994-1995	Individuals + partners	
General Household Survey (GHS)	Office for National Statistics	Available from the Data Archive at Essex University Held at CEP	Annual cross-section of approximately 20,000 individuals aged 16 or over, in approximately 10,000 households.	Each year, from 1971 to the present	Individuals	Steve Machin UCL and CEP Susan Harkness Sussex University.
Graduates Data Sets	DfEE	Department for Education and Employment.	Random samples of UK graduates and diplomats. (1 in 4 or 1 in 6)	1966, 1976, 1986.	Individuals	Peter Dolton, University of Newcastle and Institute of Education.
Hampshire LEA (subset 1)	Institute of Education	Not available for general use:	Covers 161 schools; 6400 pupils	1993 onwards 4 cohorts to date	Individuals	Harvey Goldstein, Institute of Education

Data Name	Organisation with Prime Responsibility for Data	Access Details	Population	Years covered	Spatial Unit	Key People
International Adult Literacy Survey (IALS)	Special Surveys Division, Statistics Canada,	Available free of charge direct from Statistics Canada	Between 2000 and 4200 respondents, aged 16-65, in each country (3800 in Britain).	Survey carried out in two sweeps, in 1994 and 1995. British cross-section 1995.	Individuals	Harvey Goldstein Institute of Education Siobhan Carey ONS
Labour Force Survey	ONS Socio-Economic Statistics and Analysis Group	Available from the Data Archive at Essex University Held at CEP.	Approx. 150,000 individuals	1973-1983 (biennial), 1984-1991 (annual), 1992 (quarterly)	Individuals within standard regions	
National Adult Learning Survey (NALS)	Social and Community Planning Research	Available from the Data Archive at Essex University	5,653 Adults	Main data: January 1994 to April 1997. Learning and life history data: January 1987 to April 1997.	Individuals	
National Child Development Study	CLS at the Institute of Education	Available from the Data Archive - some additional data available from CLS.	Cohort of 17,000 individuals born in the UK in March 1958. Just over 11,000 by 1991 sweep.	1958-2000 irregular follow-ups	Individuals	John Bynner
Newcastle Alumni	University of Newcastle upon Tyne	Available from Peter Dolton, at the University of Newcastle and Institute of Education.	2200 sample of Newcastle graduates.	Survey 1998 - but includes graduates from as far back as the 50s.	Individuals	Peter Dolton, University of Newcastle and Institute of Education.
OFSTED	OFSTED	Permission required from OFSTED for use.	All maintained primary and secondary schools once every 4 years	Full data for 1996 onwards	Schools	Christine Agambar (Head of Research, OFSTED)

Data Name	Organisation with Prime Responsibility for Data	Access Details	Population	Years covered	Spatial Unit	Key People
ONS Longitudinal Study (LS)	Office for National Statistics	Access via LS User Support Programme, Centre for Longitudinal Studies,	1% sample of the population of England and Wales enumerated at the 1971, 1981 and 1991 Censuses.	N/A	Primarily individuals	Jillian Smith, LS Unit, Professor Heather Joshi, LS User Support Programme,
OME : Survey of Teachers Hours and Conditions	Office of Manpower Economics	Contact the OME, DTI	Approx. 4,000 teachers at each wave.	1994, 1996, 2000	Individuals	
Programme for International Student Assessment (PISA)	OECD. (ONS in the UK)	Not available as yet.	The sample size of 15 year olds in school will be between 4k and 10k in each country with between 100 and 200 schools.	2000, 2003, 2006.	Individuals within schools	
Qualification on Entry Data Set (part of FE Individual Student Record)	Further Education Funding Council	Apply to FEFC: currently restricted to other governmental institutions (including FEDA)	FE population.		Individuals	Terry Rudden Andrew Morris (FEDA)
St Thomas's UK Adult Twin Registry	Twin Research and Genetic Epidemiology Unit , St Thomas' Hospital	See Key People	Identical and fraternal twins 1992-1996 500 female pairs From May 1996 approx. 4500 male and female same-sexed pairs over 18 years old.	original sample: 1992-1996 extended sample: 1996-2000	Twin Pairs	Dr T Spector (Twin Research Unit) Prof J Haskel (QMW)
Schools Examination Results Analysis Project (SENAP)	University of Bath. International Centre for Higher Education Management.	Access only with DfEE permission	All KS3, GCSE, and A level candidates taking exams with English boards (up to age 20)	Annual	Individuals	University of Bath: Richard Mawditt Gordon Brown DfEE: Malcom Brittain, Bridgit Frost,

Data Name	Organisation with Prime Responsibility for Data	Access Details	Population	Years covered	Spatial Unit	Key People
						Alan Short Institute of Education: Geoff Woodhouse Min Yang
Skills Survey	DfEE	Available from the Data Archive at Essex University	2,467 individuals. Weighted sample size - 2501.	January 1997 to May 1997	Individuals	Francis Green, University of Kent at Canterbury
Social Change and Economic Life Initiative	ESRC/Nuffield College (Oxford)	Available from the Data Archive at Essex University	Employers, individuals aged 20-60 and households in Aberdeen, Coventry, Kirkcaldy, Northampton, Rochdale and Swinton.	1986	Travel to work areas	D. Gallie, Nuffield College, Oxford.
Survey of Postgraduates	Social and Community Planning Research	Available from the Data Archive at Essex University	Approx. 3,000 individuals		Individuals	
Universities' Statistical Record	UCAS (now HESA)	Available from the Data Archive at Essex University	All undergraduates	Annual 1972/3-1993/4	Institutions/ individuals	
Workplace Employee Relations Survey	Department of Trade and Industry	Available from the Data Archive Held at CEP	All British workplaces. Employees in workplaces with at least 10 employees. Some WERS98 files are currently restricted.	WIRS80 WIRS84 WIRS90 WERS98	Government Office Regions	
Youth Cohort Study	DfEE	Available from the Data Archive at Essex University,	14,662 (obtained) year 11 students.	1985-1998	Individuals	Peter Dolton, University of Newcastle and Institute of Education.