

Quantitative research:
a scoping study for the
Learning and Skills Research Centre
Building effective research: 2

learning
& skills
research
centre

This publication responds to widespread concerns about the state of quantitative studies in post-16 education and training and is aimed at people concerned with planning and designing research.

Building effective research: 2

Series editor:
Andrew Morris

Quantitative research:

a scoping study for the Learning and Skills Research Centre

Michael Rendall

Institute of Education, University of London
and Office for National Statistics

Building effective research

Series editor:
Andrew Morris

The series is published by the Learning and Skills Research Centre as part of its strategy to increase the effectiveness of research on, and for, the sector. The strategy includes initiatives to develop methods, secure impact and build capacity. Publications in the series draw on these three themes, and aim to inform practitioners and policy-makers who make use of research, as well as those who undertake it.

The Learning and Skills Research Centre is supported by a grant from the Learning and Skills Council and the Department for Education and Skills

Published by the
Learning and Skills Research Centre
www.LSRC.ac.uk

Feedback should be sent to:
Andrew Morris
Learning and Skills Research Centre
Learning and Skills Development Agency
Regent Arcade House
19–25 Argyll Street
London W1F 7LS

Tel 020 7297 9104
Fax 020 7297 9190
amorris@LSDA.org.uk

Copyedited by Helen Lund
Designed by sans+baum
Printed by Digital Brookdale Print,
Southend on Sea, Essex

1416/05/03/1500
ISBN 1 85338 853 X

© Learning and Skills Research Centre
2003

All rights reserved

Planned titles

1

**From idea to impact:
a guide to the research process**

Andrew Morris

2

**Quantitative research:
a scoping study for the Learning and Skills
Research Centre**

Michael Rendall

3

**Collaborative research:
the experience of the regional Learning and Skills
Research Networks**

Lin Norman (ed)

4

**Models of research impact:
a cross-sector review of literature and practice**

Sandra Nutley, Janie Percy-Smith and William Solesbury

Contents

	Acknowledgements
	Editor's foreword
	Preface
	Executive summary
1	Section 1 Introduction
3	Section 2 Participation and demand studies
	Quantitative framework and concepts
	Review of Learning Skills Council (LSC) and
	Department for Education and Skills (DfES) research reports
	Review of data resources
	Statistical methods for quantitative research on participation
9	Section 3 Longitudinal (life course) studies
	Quantitative framework and concepts
	Review of LSC and DfES research reports
	Review of data resources
	Statistical issues and methods for longitudinal quantitative research
15	Appendix 1 The National Statistics StatBase system
16	References
17	Abbreviations

Acknowledgements

This publication has been written for people concerned with planning and designing research in post-16 education and training. It has been written in response to widespread concerns about the state of quantitative studies in post-16 education and training. The development of this report has been made possible through the partnership between the Learning and Skills Research Centre (LSRC) and the Bedford Group for Longitudinal Studies and Statistical Research at the Institute of Education (IoE) at the University of London. John Bynner, Alison Wolf and Harvey Goldstein at the IoE responded enthusiastically to the call from the LSRC to help it to address the issue, engaged in discussion about what needed to be done and recommended Michael Rendall to undertake the study.

The views expressed in this study are those of the author, and do not necessarily reflect those of the IoE or the Office for National Statistics.

In October 2001, the Learning and Skills Research Centre (LSRC) was launched by the minister for higher education and lifelong learning, Margaret Hodge. Its remit was to create a strong body of evidence from research for the long-term development of post-16 learning. It is the first such centre to focus on the needs of the learning and skills sector with its 4m learners and £7bn budget.

In recent years, there has been considerable debate about the nature and capacity of education research in the UK (Hillage *et al.* 1998; NERF 2000). Some studies have addressed, among other things, the issue of quantitative research. A report on research capacity for the Economic and Social Research Council's (ESRC) Teaching and Learning Research Programme (TLRP), for example, referred to evidence of 'relatively limited capacity in [university education departments] for quantitative research into teaching and learning' (McIntyre and McIntyre 1999).

This general state of affairs is reflected in some of the previous work of the Learning and Skills Development Agency (LSDA), formerly the Further Education Development Agency (FEDA). Relatively few staff had quantitative skills and relatively little use was made of numerical data. As a result, the quantitative dimension may have been underdeveloped in the design of some projects. Steps were taken to redress this imbalance. Staff with training in statistics were introduced, access to the Further Education Funding Council (FEFC) ISR database was negotiated and analytical software was purchased. As a result, significant new research lines were opened up, particularly in exploring relationships between learner achievement and socio-economic background (Davies and Rudden 2000).

With the establishment of the LSRC in 2001, extra impetus was given to the development of quantitative research, leading to further attempts to recruit staff with quantitative skills. This proved difficult for LSDA, as it has for others, and bore out the need for targeted capacity-building measures. The ESRC TLRP is addressing this issue through its dedicated research capacity-building network, based at Cardiff University (TLRP 2002).

These experiences may well typify a problem that runs across the sector as a whole, and which could threaten the balance of the LSRC's work. In anticipation of this, the LSRC took advice from quantitative research teams at the University of London's Institute of Education. A steering group, comprising John Bynner, Harvey Goldstein and Alison Wolf of the IoE together with Ursula Howard and Andrew Morris of the LSDA, devised a preliminary project aimed at setting out the scope for wider use of quantitative data and methods by the LSRC. Michael Rendall undertook the project between December 2001 and June 2002. Many of the issues raised in his report are of significance for organisations throughout the learning and skills sector and perhaps beyond. For this reason, his report is being made available publicly in the hope that it will encourage wider debate and action.

Many issues of quantitative research faced by the LSRC also concern research providers and users more generally. A wide range of organisations are involved in research relating to the learning and skills sector, including learning providers, universities, the DfES, the LSC, Regional Development Agencies (RDAs) and a host of independent research companies. This study addresses key questions for all of these.

- Is the full range of data sources relevant to the sector sufficiently widely understood?
- Is the range of data sources sufficiently exploited?
- Are relevant data sources sufficiently accessible?
- Are there ways in which data sets could be made more amenable to research?
- Are the skills to exploit data sources available where they are needed?
- Are the skills required to design projects with a quantitative dimension sufficiently available?
- Can understanding of the comparative merits and limitations of quantitative and qualitative methods be developed for funding bodies, designers and users of research as well as the research community itself?
- Should a wider range of quantitative analytical techniques be encouraged?

The report is also of importance for people who make use of research – policy officers, managers, teachers, trainers and support staff, for example. It provides background information and ideas relevant to some of their questions.

- Where can I obtain relevant numerical information to inform my work?
- Does my organisation have the skills to interpret and make use of numerical information?
- How do I know whether research findings are reliable or valid?
- How can I commission research that will be useful to my organisation's decision-making or to the improvement of practice?

For all those engaged with research, whether as researchers, practitioners or policy-makers, a number of concerns stand out. There is a widespread lack of awareness of what data is available, what influence it could have on our understanding of matters of practical and strategic concern, and what should be done to make use of it. For example, labour market economists make extensive use of the Labour Force Survey, but do college planners? The Census contains unique information about the population from which learners are drawn, but is it exploited? The birth cohort studies contain valuable information about what happens to people over their life courses, but do they inform advice and guidance services sufficiently?

Educational research is connecting increasingly with issues of policy and practice. The quality of planning, policy development and educational interventions is enhanced when they are informed by research-based evidence. To achieve this will require considerable changes in culture and many small steps in practice. In the area of quantitative research, this report takes one tentative step in this direction. It is hoped that it will be of use beyond, as well as within, the LSRC.

Andrew Morris
Learning and Skills Development Agency

Preface

This is an unparalleled time for embarking on a richer programme of quantitative study of post-16 education in the UK. The 2001 Census of Population includes education qualifications at a level of detail not seen since the 1971 Census. The LSC-coordinated administrative records of post-16 sector students, staff and qualifications have developed since the mid-1990s into a user-friendly integrated database system of near-complete institutional coverage. The LSDA itself is in its third year of collecting student and staff satisfaction surveys from a large number of post-16 institutions. There is a rich mix of nationally representative 'longitudinal' sample surveys that follow individuals' life courses from the cradle (their families of origin), intensively through ages 15 to 20, and afterwards into the labour-force and later ages. These surveys continue to add more recent data, allowing longer periods of life to be observed and more comparisons across cohorts; and increasingly include targeted sampling of groups at greater risk of social exclusion. There are also ongoing 'cross-sectional' surveys of persons of post-16 and labour-force age, which allow for detailed analyses of the relation between further education and labour market outcomes.

Executive summary

This report throws up a number of key points for researchers, research users and planners across the sector. A list is given below of information points and suggestions for action extracted from the report. They need to be read and debated widely. For ease of reference, they have been divided into those that focus on participation and progress (Section 2), and individual and cohort life courses (Section 3).

Participation and progress

- Statistics on participants and population rates of participation, from Learning and Skills Council (LSC) compilations of the individualised learner record (ILR) and Office for National Statistics (ONS) population estimates, may be exploited to preface studies. Use of the National Statistics StatBase is the recommended starting point for this.
- Full use should be made of the Department for Education and Skills (DfES) and LSC published statistical reports on the progress of student cohorts.
- The ILR and associated databases, Staff Individualised Record (SIR) and Qualifications Database (QDB), within the FE system remain the basic sources of data on students, staff and qualifications in the sector.
- The ILR may be used alone for studies of **participants**, but cannot be used alone for **participation**. For use in participation studies, a separate denominator source must be used. This will be either the decennial Census of Population or official population estimates made between censuses.
- The 2001 Census may be used alone for the estimation and analysis of participation in small areas defined, for example, by indicators of economic and social exclusion. Multilevel modelling is often an appropriate statistical method for analyses involving more than one institution or region.
- 2001 Census data also enables analysis of the stock of post-16 qualified individuals, providing information on their current employment and occupation. Aggregate tabulations from ONS and individual data kept on the University of Manchester's Samples of Anonymised Records (SARs) system may be exploited for this purpose.
- The ONS Longitudinal Study (LS) census microdata will be available from early 2004 with the data from the 2001 Census included. This is currently the only longitudinal data source with sufficiently large samples for analysis of special sub-populations such as ethnic groups or older students.
- The Labour Force Survey (LFS) is the best survey data source for estimation of post-16 participation rates across ages and over time. As it is a national sample survey, it would not normally be appropriate for the estimation of smaller geographical areas or for the study of very specific groups. However, recent 'booster' samples have been designed to allow for some estimation, geographically, at Local Education Authority (LEA) level as part of the 'Local area LFS' (LLFS).
- The LSDA-coordinated student and staff surveys may be used to explore whether the satisfaction of students with their courses, or of teachers with the institution and its management, predicts future rates of retention and levels of performance. Improvement of the statistical sampling of the LSDA surveys is necessary for this.

- The DfES's forthcoming Student Income and Expenditure Survey is destined to become the main survey source for student expenditure. The main UK expenditure survey, the Family Expenditure Survey does not break down 'education fees' by educational sector, with the exception of the category 'Leisure classes – fees'.
- The Learning and Training at Work (LTW) survey of employers may be used for matching demand and supply.

Individual and cohort life courses

- Life trajectories may be analysed with data from retrospective surveys and from surveys of a sample of individuals (a 'panel') who are interviewed at repeated intervals in their lives. Panel data is usually considered to be of better quality than retrospective data, as it does not rely on the accurate recall of respondents about past events.
- The main statistical methods for analysing panel data are 'survival' or 'spells' analysis. They may be used to provide measures of staff retention and of student retention and achievement. These analyses can be extended to consider multiple periods (spells) in full-time or part-time education.
- Comparison of student retention rates between the administrative and survey sources should be explored. Administrative and survey sources may be combined to analyse institutional and individual factors affecting student retention rates.
- There are three main sets of sample survey panel studies in Britain:
 - the DfES's Youth Cohort Study (YCS)
 - the Institute of Education's Cohort Studies, National Child Development Study (NCDS) and British Cohort Study (BCS)
 - the University of Essex's British Household Panel Survey (BHPS).
- The YCS has, since 1988, collected data on groups (cohorts) of 15 year olds as they pass through their next five years of schooling or entry to the labour force. This may be used to describe cohort changes and group differentials in the progress and pathways of learners through post-16 institutions. Pooling different YCS cohorts would allow for sufficient numbers for the main ethnic groups to be analysed.
- The BHPS allows for comparison of labour-market trajectories of persons with and without post-16 educational experiences.
- The Institute of Education's Cohort Studies (the 1958 NCDS and the 1970 BCS) provide opportunities to study the influence of family and schooling background on participation and qualification. Use of the qualification-on-entry fields of the ILR may be used to complement survey data on routes of entry.
- A forthcoming Cohort Study of Young People will follow individuals from age 14 to age 25, and include an 'oversample' of ethnic minorities.

Section 1

Introduction

This report addresses the following objectives set out by the LSDA and the Institute of Education's steering group on scoping quantitative research in the learning and skills sector:

- to report on the relevance of data sources both inside and outside the education system
- to report on the range of appropriate statistical methods for analyses of this data.

Three ways of using quantitative data are considered in the report:

- extracting from appropriate prior tabulations
- re-analysing existing data ('secondary data analysis')
- collection and analysis of data ('primary data analysis').

The report draws on a review of the LSDA's project database; data inside and outside the education system; and quantitative statistical production and analysis from the LSC, DfES and ONS. Two 'frameworks of quantitative analysis' are considered separately: participation and demand studies, and longitudinal (life course) studies. The report reviews concepts, existing reports, data resources and statistical methods for each.

Notes on terminology

- The term 'post-16' refers to 'post-16 education excluding higher education' – that is, to all skill-learning environments within the scope of the learning and skills sector.
 - The individualised student record (ISR) in the FE subsector is being developed into the individualised learner record from 2002, and over time will include work-based and adult and community-based learning.
-

Section 2

Participation and demand studies

Quantitative framework and concepts

The data required for studies of participants and of participation is contrasted. Studies of student numbers and their composition, for example, give information on participants. Numbers of persons in a given sub-population of potential participants are additionally needed for inference about participation. Participation rates consist of a numerator of participants with a given set of individual characteristics and geographical location, and a denominator consisting of all persons with that same set of characteristics and location (for example, 16–18-year-old FE students in England divided by all 16–18 year olds in England).

Participation rates may be calculated from separate data sources for the numerator and denominator, or from a single source that supplies data for both the numerator and denominator. For example, participation rates may be calculated using college administrative data for the numerator (number of students by age and sex in a given location) and official estimates of population for the denominator (number of individuals by age, sex and location) that are based on population counts from the most recent census. Alternatively, statistical sample surveys in which individuals give information about their current student status may be used alone to estimate participation rates.

Cross-sectional surveys sample individuals from a sampling frame approximating all individuals of the contemporaneous population. This contemporaneous sampling scheme enhances the degree to which cross-sectional surveys are representative of the population at a given point in time. This is one major reason why they are generally superior to panel surveys (discussed below in Section 3) for the purpose of estimating participation rates or other single-point-in-time quantities. The second reason is that they also typically have larger sample sizes than do panel surveys. This allows for statistically reliable estimates of participation in population sub-groups. Nevertheless, lack of estimation detail, especially with respect to geography, is the single most important disadvantage of surveys as compared to participation rates estimated from population counts of the numerator and denominator.

As a practical example of the data-related issues that arise in studies of participation, to understand the performance of an FE college in providing broad access, data on potential participants is needed in addition to data on actual participants. Defining the geographical area from which these potential participants might come (the region, the sub-region, the college's local area) is an important issue. By comparing participant composition data across colleges in a sub-region where the population composition is known, for example, the implicit assumption is that all colleges are drawing from the same population. This assumption can be challenged. Further, changes in the character and geographical distribution of that population over time require monitoring to understand changes in participation.

Review of LSC and DfES research reports

Participant numbers and composition are reported annually by the LSC. Reporting of participation by the LSC is very fast. For example, the Statistical First Release of student numbers as at 1 November 2001 was released on 28 March 2002 (LSC 2002a). The availability of up-to-date results is a common advantage of administrative data over survey data, and over estimation from a combination of administrative and population data.

FE participation descriptions by the DfES use mid-year population estimates provided by the ONS as denominators for the participation rates (DfES 2002a). The need for independently generated denominators makes participation rates available significantly (about a year) later than participant numbers. Further, the only individual demographic dimensions for which estimates between censuses are regularly available, at all geographical levels, are age and sex. These are available down to the LEA level. The issue of whether institutions draw students from multiple local areas, however, arises here. Thus the DfES reports participation rates for inner London as a whole, and not for local LSCs within inner London.

Published and electronic forms of tabulations on participation

Tabulations from ISR and Census data of participants and participation are available in both published and electronic form. They may be exploited to preface LSDA studies with a general quantitative overview. Use of the National Statistics StatBase system is a recommended starting point for this (see Appendix 1). The DfES and LSC websites (www.dfes.gov.uk/statistics and www.lsc.gov.uk) may also be queried directly. The DfES website is the easier to navigate and includes many of the LSC statistical reports.

Review of data resources

Administrative records of the ILR and SIR databases

For data collected inside the education system, the LSC manages databases of learners in post-16 institutions and in combined work–study environments, and of teaching staff in post-16 institutions (LSC 2001a, 2001c, 2001d). The system of individualised learner records may be viewed as a census of participants. Its precursor in further education, the individualised student record, was introduced in 1994/95 (LSC 2001a). An analogous record (the staff individualised record or SIR) is kept by institutions and compiled statistically by the LSC for all teaching staff.

The ILR is integrated with the Qualifications Database (QDB) that provides details on the particular sets of qualifications being sought. The QDB's coverage of awarding bodies accounted for 92% of National Vocational Qualifications (NVQs) and 100% of General National Vocational Qualifications (GNVQs) in 2000/01. Statistics for other vocational qualifications are currently obtained only from the three largest awarding bodies: Edexcel, City and Guilds, and Oxford, Cambridge and RSA (OCR); and so cannot be considered to approach full coverage. Participation of NVQ awarding bodies in the QDB increased substantially in the late 1990s, such that the DfES has produced national estimates of qualifications awarded, adjusted up for non-coverage, from 1996/97 onwards (DfES 2002d).

A work-based learning (WBL) database (LSC 2001a) exists in parallel with the ISR database, covering learners on Advanced and Foundation Modern Apprenticeships, NVQ learning, Life Skills, Preparatory Learning and Co-Financed Delivery (European Social Fund). It also integrates the learner record with the QDB.

The ILR may be used alone for studies of participants, but cannot be used alone for participation. For use in participation studies, a separate denominator source must be used. This will be either the decennial Census of Population or official population estimates for years between censuses.

Official population estimates as FE participation rate denominators

The Office for National Statistics of England and Wales (ONS) and the corresponding statistical organisations for Scotland and Northern Ireland estimate population numbers by age and sex at regional and district levels (ONS 2001; GRO Scotland 2001). Estimates are available on the ONS StatBase system (see Appendix 1). Population estimates (eg of 17 year olds in a particular district) form the denominators used by the DfES for its estimates of post-16 participation, where the numerators (eg number of students aged 17 in a particular LEA) are derived from the ILR system. Small adjustments are necessary to account for the ILR being drawn from at year end, while the population is estimated at mid-year.

As participation rate denominators, population estimates made between censuses are inferior to census counts in two ways: the estimation methods are less accurate than actual counts; and they give no individual social or demographic detail beyond age and sex.

The 2001 Census of Population for participation rate denominators

Extra census detail is usable in a participation rate when that same detail is provided in the ILR. An important example of the usefulness of the 2001 Census for participation rate denominators is that of ethnicity. For 2001/02, the ILR's ethnicity codes have been updated to match those of the 2001 Census (LSC 2001c). Depending on the degree to which accurate information for the ethnicity field has been entered in the ILR, this allows for an important new opportunity to analyse participation by ethnic minorities both nationally and locally. Checks on the prevalence and accuracy of entry for the ethnicity field in the 2000/01 ILR should precede the estimation of the 2001 participation rates by ethnicity.

A further possible dimension to consider for participation studies that use the 2001 Census as a denominator is that of place of residence. The ILR already includes a 'Widening participation category' field that includes 'from deprived area' as a category (see www.lscdata.gov.uk/data/wideningparticipation.html). 'Home postcode' is also found in a separate field of the ILR. With this data, the 2001 Census provides an opportunity for the estimation and analysis of participation in small areas defined, for example, by indicators of economic and social exclusion. For a full range of local geographic indicators, see www.neighbourhood.statistics.gov.uk/catalogue.asp.

For information about the content and timing of the various releases of 2001 Census data, see www.statistics.gov.uk/census2001/default.asp. The data is currently scheduled to be released first in aggregate form between October 2002 and July 2003, and second in the form of anonymised individual records from the University of Manchester's SARs system later in 2003 (see www.ccsr.ac.uk/sars/). A third form of release, as longitudinal records linked across censuses, is discussed under Longitudinal studies (Section 3) below.

The quarterly Labour Force Survey for participation studies

The ONS quarterly Labour Force Survey is the best single data source in Britain for estimation of participation rates in post-16 education at all ages over 16 (see www.statistics.gov.uk and the *Labour Market Trends* journal). This data is available through the user-friendly Nomis interface (www.nomisweb.co.uk; see also Sutherland and Blake 2000).

The LFS asks a detailed set of questions about current studies and past qualifications (ONS 1997). Variables of importance for post-16 participation include: QUALNOW – Whether working or studying towards a qualification; HITQUAL – Highest qualification being worked for; and MODAPP – Whether apprenticeship forms part of a Modern Apprenticeship scheme. Variables as detailed as these about current education activity are found in the LFS, along with much detail about employment and basic demographic information including family and household structure. This information allows for more detailed information about participants and especially participation than the ILR and ILR/Census sources.

The richness of these additional variables is thus the main source of advantage of the LFS over the combination of ILR, census data and official population estimates made between censuses for estimating and analysing participation in post-16 education. The second source of advantage of the LFS is that there is no problem of match between numerator and denominator individuals, as the same LFS individuals contribute to both numerator and denominator.

The main source of disadvantage of the LFS is that as a sample survey, it would not normally be used in the estimation of smaller geographical areas (such as LEA areas) or for studying very specific groups – defined, for example, by a combination of specific age, sex, and ethnicity. For such studies, the researcher would need to pool several contiguous quarters or years of the LFS to overcome sample size limits.

An important development is that, in partnership with the DfES, the Department for Work and Pensions (DfWP) and the National Assembly for Wales, the ONS has recently added 'booster' samples to the English and Welsh LFS (Hastings 2002; ONS 2002a). These have been designed to allow for some estimation at the LEA geographical level as part of the 'Local area LFS'. A corresponding boost is planned for Scotland, beginning with the 2003/04 year. The LLFS achieves its large sample size by pooling across four successive quarters over a 2-year period of data collection. One point the user must bear in mind when pooling quarters of the LFS, however, is that attention must be paid to variables that are not available in every quarter: QUALNOW, for example, has been available in spring and autumn quarters only.

Other data sources for participation studies

Economic studies enquire more broadly about factors related to participation. Costs and benefits are the main concepts for which quantitative data is needed. Unfortunately the main survey of expenditure in Britain, the Family Expenditure Survey (ONS 2002b), does not break down 'education fees' by level of education. 'Leisure classes – fees' is perhaps the only expenditure category that may be of use in post-16 studies. The forthcoming Student Income and Expenditure Survey (www.dfes.gov.uk/research) is likely to be the main source of surveyed expenditures. Its uses, in parallel with ILR data on course costs and funding for the student, should be explored. Finally, the Learning and Training at Work (LTW) survey of employers (DfES 2003) may be considered for matching demand and supply with respect to occupation and related qualifications.

Statistical methods for quantitative research on participation

The discussion above (under Quantitative concepts and frameworks) on the role of colleges in encouraging diversity of access illustrates a common situation for which multilevel modelling may be an appropriate statistical method (Goldstein 1995; Yang *et al.* 2002). The college constitutes one level of cause of an individual's participating or not. Multilevel modelling may be considered in other types of study involving comparisons between institutions or teachers within institutions.

Another application of the multilevel methodology is to the combining of different data sources, with each source constituting a level of the model (Heady, Ruddock and Goldstein 1997). This method may be applied to the estimation of education statistics for local areas, where data is available from both local-level and national-level sources. Being able to use methods to combine data sources also has implications for the design of surveys. Where those surveys are designed to interface with existing data collections, greater benefits can be realised from the subsequent combining of sources at the analysis stage.

Section 3**Longitudinal (life course) studies****Quantitative framework and concepts**

Data that is collected for more than one period of an individual's life course is referred to as 'longitudinal'. Longitudinal data includes both 'retrospective' and 'panel' types of collection. Retrospective collections ask respondents to report life histories, for example, of their entry to, duration in, and qualifications obtained during, their educational 'career'. Panel data is collected for a set of individuals who are interviewed at different times. Respondents are asked to report on their contemporary status (eg current employment or student status) and activities since the previous interview (eg qualifications obtained, time in and out of employment). Panel data is usually considered to be of better quality than retrospective data as it does not rely on the accurate recall of respondents about past events.

Panel studies collect data in 'sweeps' (or 'waves') in which the individual is re-interviewed. The dimensions across which panel studies differ are as follows. Different panel studies have different lengths of time between sweeps, different ages and criteria for inclusion in the panel for its first interview, different lengths of time between the first and last interview ('panel duration'), and different lengths of time between the beginnings of collection of data on new panels. In common with cross-sectional sample surveys, the main initial tasks are the selection of the 'sampling frame', consisting of all members of the population of interest (the study's 'universe'); choosing a method of sampling from this sampling frame; and obtaining the maximum possible number of selected individuals who actually respond (measured by the initial response rate). Sampling methods may give equal probability of selection to all individuals, or they may 'oversample' (give greater probability of selection to) sub-groups in the sampling frame. Sample 'weights' are used to adjust estimates for different probabilities of selection.

The main challenge that is unique to panel studies is to minimise respondent attrition from one interview to the next. The 'attrition' of sample respondents in the course of successive sweeps must be monitored and evaluated for its effect on sample representativeness. Modifications to individuals' sample weights in the course of successive sweeps reduce the drift of estimates from the sample away from those expected for the population. It is, nevertheless, this drift that makes panel studies less representative of the contemporaneous population than are cross-sectional surveys.

Data on the background of students and their subsequent entry or non-entry to post-16 learning may be used to investigate entry. Data on the subsequent life courses of students, especially when compared with individuals who did not participate in post-16 learning, may be used to investigate outcomes of that learning. This data is likely to come from a variety of sources.

Review of LSC and DfES research reports

The LSC publishes annual analyses of ILR data by institution, including two longitudinal measures: student year-on-year retention rates and final-year qualification achievement rates (LSC 2002b).

The DfES publishes analyses of survey results from its Youth Cohort Study (DfES 2002b, 2002c), discussed in more detail below. Two remarks may be made about these research reports. One is that they do not make as much use as they could of the longitudinal aspects of the survey – which would allow analysis of transitions out of, between and back into FE and other education institutions. Second, the reports are mostly descriptions of one cohort only, missing the opportunity to compare cohorts.

Opportunities for comparison of student retention rates between the administrative and survey sources, and for complementary analyses of institutional and individual factors affecting student retention rates, might be explored more than is currently seen in the respective DfES and LSC reports. Use of the qualification-on-entry fields of the ILR may be used to complement survey data on routes of entry.

Review of data resources

Retrospective and panel data from the Census and Labour Force Survey

As noted above, any data set that collects information at more than one time point in an individual's life may be described as 'longitudinal'. In a more comprehensive use of the term, any data collection that asks respondents to report past achievement of educational qualifications is 'longitudinal'. It is in this sense that the use of the 2001 Census and the LFS will now be considered.

The 2001 Census asks each individual aged 16 or older to indicate which of 12 categories of qualification he or she holds (Moss 1999; www.statistics.gov.uk/census2001/censusform.asp). These categories include NVQ Level 1, Level 2, Level 3 (or Advanced GNVQ) and Levels 4–5 (or HNC, HND), and other qualifications (vocational). No such level of educational detail was available in the previous two censuses, in 1991 and 1981 (Dale 2000). This data permits analyses of Britain's stock of post-16 qualified individuals. It may also allow for 'progression and pathway' studies of the numbers of FE-qualified individuals who go on to obtain higher education (HE) qualifications, since the census respondent may 'tick' both sets of qualification. Further, the census provides information on the current employment, self-employment, unemployment or non-participation in the labour force of respondents. Use of the anonymised SARs census microdata records (www.ccsr.ac.uk/cmuc/index.htm) is encouraged for analyses of this data, as relationships between FE qualifications, employment and other variables can then be explored through specialised cross-tabulations and multivariate statistical techniques.

More detailed analyses of qualifications achieved in relation to current employment status may be made using the LFS, although for less specific populations, owing to its being a sample survey rather than a census. Economic studies of the individual payoff to education require the kind of detailed employment and wage data found in the LFS. Economic studies of the public costs and benefits of education require financial information on costs at the institutional level. Administrative data from within the system is required for this. A panel element is also now available in the LFS through household interviews in five successive quarters (ONS 2002a).

Studies of inter-generational educational mobility stand to profit from a third set of 2001 Census data still to be released: the ONS's Longitudinal Study (LS), scheduled to be available from early 2004 through the Celsius user support service (www.celsius.lshtm.ac.uk). The LS links individuals' census records between the 1971, 1981, 1991 and 2001 Censuses, and it links women's birth registrations. As such, it is a true panel design. As a random sample of 1% of the population of England and Wales, this is the only longitudinal data source with sufficiently large samples to conduct analyses of regions and special sub-populations such as ethnic groups or older students.

A particular opportunity for the study of the role of post-16 learning in inter-generational educational mobility comes with the more detailed coding of education in the 2001 Census. For persons born from the mid-1950s up to the time of the 1971 Census, their father's and mother's education, as recorded in the 1971 Census, will also be available. As noted above, 1971 was the most recent census before 2001 to record all levels of education.

Panel data from the Youth Cohort Study, the Cohort Studies and British Household Panel Survey

There are three main relevant sets of sample survey panel studies in Britain: the DfES's Youth Cohort Study, the Institute of Education's (IoE) Cohort Studies and the University of Essex's British Household Panel Survey. Initial sample sizes are approximately 20,000 for each (annual) cohort of the YCS, and between 10,000 and 15,000 for each of the two IoE Cohort Studies and for the BHPS.

The IoE's Cohort Studies and the BHPS were recently reviewed for their sample attrition and for bias resulting from this attrition (Nathan 1999). The conclusions of this review were that all the studies had successfully maintained their representativeness of the population, and so appear to be little affected by attrition biases. Rates of sample attrition are substantially higher in the YCS, probably because it is a telephone survey. For example, approximately one-third of the cohort of 19 year olds in 2000 had been lost in each of the previous sweeps (at ages 16, 17 and 18), resulting in only 20% of the original sample remaining at age 19 (LSC 2001b). The year-on-year results of the YCS may be compared with the LSC's (2002b) student retention and achievement rates, to assess bias through non-random attrition. The LSC (2001c) reports some bias towards overestimating participation rates in full-time education.

The YCS (DfES 2002a) has, since 1988, collected data on annual panels consisting of 15 year olds in compulsory schooling. Detailed variables of education type and employment status, together with family background in the first sweep, make this an invaluable source of data on the first post-16 years. Statistically, the YCS comprises simple random samples of approximately 20,000 individuals (before attrition) in each panel. Response rates of about two-thirds are achieved, on average, both initially and at each successive sweep. This means that each individual panel has too few individuals from ethnic minorities to compare their progress with that of other ethnic minorities and with the majority white group (eg DfES 2002b). Were the different YCS cohorts to be pooled for analysis, however, a total sample of up to 200,000 individuals could be used in the analysis. This would allow for sufficient numbers for the main ethnic groups to be analysed. Further, comparisons across cohorts would be of considerable substantive interest in this pooled analysis.

The YCS does not have the breadth of age coverage of more general surveys, especially that of the BHPS among longitudinal surveys. However, the DfES also commissions specialist studies such as the National Adult Learning Survey 2002 (www.dfes.gov.uk – see Tender 123/2001) and *Post-16 transitions: a longitudinal study of young people with special needs (SEN) – Stage II* (www.dfes.gov.uk). These studies should be exploited to broaden the type of post-16 experience analysed.

The BHPS data has the advantage over the IoE Cohort Studies of having a more frequent data collection schedule – annual since 1991. This provides an opportunity to explore the subsequent employment and occupational trajectories of learners, using 'survival' and 'spell' analysis (see following subsection). The BHPS includes annual questions on whether an individual is currently in post-16 education or training, and on qualifications already obtained; also, there is a single year (1992) in which retrospective questions were asked on the month and year in which the respondent first left full-time education. The details of which qualifications were obtained are reasonably detailed – number of GNVQs, trade apprenticeship qualification, City and Guilds Part I, II and III, etc. Being a general survey, the BHPS sample sizes of any given age are relatively small, but the population coverage of persons of different ages is comprehensive. The unique advantage of the BHPS is its annual surveying of respondents of all ages. This allows for labour market trajectories to be compared between persons with and without post-16 education and training experiences.

Data for the IoE Cohort Studies (the 1958 NCDS and the 1970 BCS) is collected at less frequent intervals (an earlier and comparable cohort study is the 1946 Medical Research Council study: its data is held at University College, London), but covers a longer period of life (see www.cls.ioe.ac.uk/Cohort/mainncds.htm). They provide opportunities to study the influences of family background on participation and qualification in further education. The cohort studies also have the advantage of being specialised in the earlier part of life only. This gives both large sample sizes at the most common ages of participation, and more measures of early educational, behavioural and learning outcomes. The participation and qualification questions are very detailed, more so than in the BHPS. The Cohort Studies ask about the number and timing and the main subject of qualifications by source. These studies follow individuals through more stages of life than any other survey does, albeit with longer periods between surveys. Their geographical coverage is comprehensive, because they sample all births of one particular week.

Both the BHPS and the two current Cohort Studies (NCDS and BCS) are multistage cluster samples that do not oversample any particular population subgroup. A forthcoming additional study (The Cohort Study of Young People) of a cohort born in the mid-1980s will follow individuals from age 14 to 25. This new study will oversample ethnic minorities, allowing for statistically reliable comparisons between majority and minority groups.

Statistical issues and methods for longitudinal quantitative research

The design and implementation of longitudinal data collection are challenging. Analysis of this type of data likewise presents major challenges as well as rewards. Longitudinal sample surveys are not suitable for analyses at disaggregated geographical levels; their sample sizes are too small. Further, the national surveys provide little or no information on which institutions respondents attended, or in which work-based learning environments they participated. Surveys drawn from administrative sampling frames will continue to be necessary for analysis of institutional impacts on individual student outcomes from post-16 education.

With respect to analytical methods for exploiting longitudinal data, the main methods are 'survival' or 'spells' analysis (eg Jarvis and Jenkins 1997). These include the year-on-year student retention and final-year qualification achievement measures of the LSC analyses of ILR data. They could also include analogous staff retention measures. These analyses can be extended to consider multiple years of retention and multiple periods (spells) in full-time or part-time education.

A further statistical extension, potentially applicable to the data available from different FE databases, involves the combining of so-called 'stock' and 'flow' data. Studies of student and faculty flows into and out of post-16 education and training may make complementary use of data within and outside the educational system. Statistical methods provide for ways of combining flow data with 'participation' data to estimate probabilities of entry. The statistical basis and the objectives of these methods are closely related to those proposed in Section 2 to combine local and national data. In the case of flow and participation data, the implementation of the methods requires specialist statistical computer programs not found in statistical package software. See Lancaster and Imbens (1996) for a relevant example.

Research studies will typically benefit from a quantitative preface to the particular study, whether the study itself uses quantitative or qualitative methods. The National Statistics StatBase system is a recommended starting point. It has an excellent search facility (www.statistics.gov.uk/statbase/search.asp), covering data sources, published reports and analyses (including downloadable summary tabulations), and institutional contact information. I used 'further education' as input search text to arrive at the examples below.

Example 1**FE enrolment numbers compared with those in higher education**

- A tabulation of FE versus HE numbers decennially from 1970/71 and up to 1998/99 shows 4.04m students enrolled in further education in the year 1998/99, compared with 2.08m students enrolled in higher education that year.
- A ratio of approximately two FE students to one HE student was seen also in 1970/71, 1980/81 and 1990/91.
- The ratio of female to male students is higher in further education than in higher education. This is accounted for entirely by the higher numbers of female part-time FE students, who exceed male part-timers by more than one-third.

Example 2**Regional rates of FE enrolment among 16 and 17 year olds**

- Over the entire UK, fewer 16 and 17 year olds were at school (37.9% of 16 year olds and 28.7% of 17 year olds) than in further education (40.5% and 38.0%).
- Full-time enrolment constitutes about 80% of FE students at these ages.
- Regional patterns of FE versus secondary school enrolment offer some evidence that FE enrolment substitutes for secondary school enrolment (as opposed to substituting for non-enrolment).
- North-West England, for example, had England's lowest school participation at both ages 16 (24.3%) and 17 (19.4%) and its highest FE participation rates (40.8% at age 16 and 33.9% at age 17 in full-time further education).
- The East Midlands had the reverse situation: higher than average percentages of 16 and 17 year olds in school (36.7% and 29.1%) and below average proportions in further education (29.7% and 25.9% in full-time further education).
- Exceptionally, however, London saw above average participation rates in both school and further education, and the North-East combined low school participation rates with no more than average FE enrolment rates.

In these examples, StatBase acts as a clearing house for FE statistics produced by other agencies, notably the DfES and LSC. More recent and more comprehensive releases of these results may be found on the DfES (www.dfes.gov.uk) and LSC (www.lsc.gov.uk) websites. The StatBase system is also the clearing house for ONS-produced data of relevance to further education, including population estimates useful for participation rate studies and education and labour market results from the Labour Force Survey.

References

- Dale A (2000). Developments in census taking in the last 25 years. *Population Trends*, Vol. 100. The Stationery Office.
- Davies P and Rudden T (2000). *Differential achievement: what does the ISR profile tell us?* Learning and Skills Development Agency.
- DfES (2002a). *Participation in education and training by young people aged 16 and 17 in each local area and region, England, 1995/96 to 1999/00*. At www.dfes.gov.uk/statistics.
- DfES (2002b). *Youth Cohort Study sample questionnaire*. At www.dfes.gov.uk/statistics.
- DfES (2002c). *Youth Cohort Study: the activities and experiences of 19 year olds, England and Wales 2000*. At www.dfes.gov.uk/statistics.
- DfES (2002d). *Vocational qualifications in the UK: 2000/01*. At www.dfes.gov.uk/statistics.
- DfES (2003). *Learning and Training at Work 2002*. At www.dfes.gov.uk/statistics.
- Goldstein H (1995). *Multilevel statistical models*, 2nd edition. Edward Arnold.
- GRO Scotland (2001). *Mid-2000 population estimates: Scotland*. General Register Office for Scotland.
- Hastings D (2002). *Methodology for the 2001/02 annual local area Labour Force Survey data*. Office for National Statistics. At www.statistics.gov.uk.
- Heady P, Ruddock V and Goldstein H (1997). An investigation into the possible use of multilevel models based on survey data to update census estimates for small areas. In *Statistics Canada Symposium – new directions in surveys and censuses*. Statistics Canada.
- Hillage J, Pearson R, Anderson A and Tamkin P (1998). *Research into schooling*. DfEE Research Briefing RR7. Department for Education and Employment.
- Jarvis S and Jenkins S (1997). Low income dynamics in 1990s Britain. *Fiscal Studies*, 18(2), 1–20.
- Lancaster T and Imbens GW (1996). Case control studies with contaminated controls. *Journal of Econometrics*, 71, 145–160.
- LSC (2001a). *Work-based learning Individualised Learner Record, 2001–2002 Guidance Notes*. At www.lsc.gov.uk/data/studentstatistics.html.
- LSC (2001b). *Student numbers, in-year retention, achievements and destinations at colleges in the further education sector, external institutions and further education student numbers in higher education institutions in England in 1999/2000*. At www.dfes.gov.uk/statistics.
- LSC (2001c). *Individualised Student Record (ISR) 2001–2002: institution support manual*. At www.lsc.gov.uk/data/studentstatistics.html.
- LSC (2001d). *Staff Individualised Record (SIR) 2000–2001: college support manual*. At www.lsc.gov.uk/data/staffstatistics.html.
- LSC (2002a). *Student numbers at colleges in the further education sector and external institutions in England on 1 November 2001*. Statistical First Release 21. At www.lsc.gov.uk/documents.cfm.
- LSC (2002b). *Summary statistics for further education student institutions: England 1999/2000*. At www.dfes.gov.uk/statistics.
- McIntyre D and McIntyre A (1999). *Capacity for research into teaching and learning*. At www.tlrp.org (see under publications), 19 June 2002.
- Moss C (1999). Selection of topics and questions for the 2001 Census. *Population Trends*, Vol. 97. The Stationery Office.
- Nathan G (1999). *A review of sample attrition and representativeness in three longitudinal surveys (the British Household Panel Survey, the 1970 British Cohort Study and the National Child Development Study)*. Government Statistical Service Methodology Series No. 13. Office for National Statistics.
- NERF (2000). *Building research capacity*. At www.nerf-uk.org (see sub-group reports), 19 June 2002.
- ONS (1997). *Labour Force Survey user guide. Vol. 3 1997, Details of variables*. Office for National Statistics. Also at www.statistics.gov.uk for latest online version.
- ONS (2001). *Mid-2000 population estimates: England and Wales government office regions, unitary authorities, local government districts, health regional office areas and health authority areas*. Series PE No. 3. The Stationery Office.
- ONS (2002a). *Annual local area Labour Force Survey 2001–02*. At www.statistics.gov.uk.
- ONS (2002b). *Family spending: a report on the 2000–2001 Family Expenditure Survey*. The Stationery Office.
- Sutherland S and Blake A (2000). Nomis via the web. *Labour Market Trends*, 108(7), 349–352.
- TLRP (2002). *Research capacity-building network*. At www.cf.ac.uk/socsci/capacity/, 19 June 2002.
- Yang M, Goldstein H, Browne W and Woodhouse G (2002). *Multivariate analysis of examination results*. Institute of Education, University of London.

Abbreviations

BCS

British Cohort Study

BHPS

British Household Panel Survey

DfES

Department for Education and Skills

FE

Further education

GNVQ

General National Vocational Qualification

HE

Higher education

ILR

Individualised learner record

ISR

Individualised student record

LEA

Local Education Authority

LS

(ONS) Longitudinal Study

LSC

Learning and Skills Council

LSDA

Learning and Skills Development Agency

NCDS

National Child Development Study

ONS

Office for National Statistics

QDB

Qualifications Database

SARs

Samples of Anonymised Records

VO

Vocational qualification

YCS

Youth Cohort Study

How to contact the LSRC

The LSRC welcomes continuing interaction with researchers and research users. Please contact us with your questions, ideas and information.

Feedback should be sent to:

Andrew Morris

Learning and Skills Research Centre

Learning and Skills Development Agency

Regent Arcade House

19–25 Argyll Street

London W1F 7LS

Tel 020 7297 9104

Fax 020 7297 9190

amorris@LSDA.org.uk

