Using numeric secondary data in education research

To cite this reference:

Smith, E. (2011) Using numeric secondary data in education research, British Educational Research Association online resource. Available online at: [INSERT WEB PAGE ADDRESS HERE] Last accessed [insert date here].

Professor Emma Smith
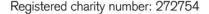School of Education, University of Birmingham
July 2011

Contents

Summary

This resource provides a brief introduction to using numeric secondary data in educational research. Secondary data analysis is a relatively under-used technique in the field of education yet the potential for its use, among novice researchers in particular, is huge. From the academic attainment of everyone in school in England to young people's views on friendship and social networks, there are not many aspects of the educational and social world that aren't covered by some type of secondary data. Indeed, it is very likely that whatever your research topic, the answers to at least some of your research questions can be found by analysing secondary data.

What is secondary data analysis?

Secondary data analysis is 'an empirical exercise carried out on data that has already been gathered or compiled in some way' (Dale et al. 1988:3). In other words it is an approach where the researcher analyses data that has already been collected, usually by someone else. The analysis may involve using the original, or novel, research questions, statistical approaches and theoretical frameworks. Secondary data comes in many forms. It can include the data generated from systematic reviews, through documentary analysis as well as the results from large-scale surveys such as the National Census or the Programme for International Student Achievement (PISA).

Secondary data can be numeric or non-numeric. However, our interest here is with numeric secondary data. In the field of education, examples of this might include the following:

- Data on every school pupil in England gathered as part of the Pupil Census

- Cohort studies, such as the Youth Cohort Study or the Millennium Cohort Study

- Large surveys such as the Young People's Social Attitudes survey or the Longitudinal Study of Young People in England

- National or international survey data such as the Programme for International Student Achievement (PISA), the World Values Survey, the International Adult Literacy Survey.

- Administrative data such as that collected by the Department for Education and the Higher Education Statistical Agency.

We will look at some of these sources in a little more detail later on, but first we consider some of the reasons why one might wish to use secondary data in their research.

What are the advantages to using secondary data analysis?

The promises of secondary data analysis are many. It can allow researchers to access data on a scale that they could not hope to replicate first hand; and the technical expertise involved in developing good surveys and good datasets can lead to data that is of the highest quality. Secondary data can be analysed from different empirical or theoretical perspectives and in this way provides opportunities for the discovery of relationships that may not have been considered in the primary research. In addition, secondary analysis is also a very democratic research method. The availability of low cost, high quality datasets means that secondary analysis can ensure that 'all researchers have the opportunity for empirical research that has tended to be the privilege of the few' (Hakim, 1982:4). As 'it is the costs of data collection that are beyond the scope of the independent researcher, not the costs of data analysis' (Glaser, 1963:12), the very accessibility of the data enables novice and other researchers to retain and develop a degree of independence. Often when early career researchers are employed on busy projects (or undertaking Masters or Doctoral level study) there is limited time and resources to apply for grants for other funding. By circumventing the data collection process, secondary analysis can enable novice researchers, including students, to gain valuable experience in undertaking research in an area of their own interest, as well as presenting opportunities to publish and present their findings as independent researchers. In this sense secondary data analysis has a valuable role in the capacity building of research skills as well as in developing an early career researcher's theoretical and substantive interests (Smith 2008).

One of the key attractions of secondary data analysis is that you do not have to be a statistician to access and engage with this type of data. Some of the most powerful and useful analyses can be done by using relatively straightforward arithmetic techniques which are well within the capabilities of even the most stats-phobic researcher. The next section will give two short examples of how secondary data might be used in educational research.

Using secondary data in small scale research

One excellent application of secondary data analysis is in small scale research and especially in work that involves a mixture of different research methods. Two examples illustrate this point.
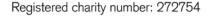
Example 1: Researching life in a Sure Start children's centre

This example shows how secondary data from the national Census can be used to characterise the local population of an area which is the focus of a case study. This study is located in a Sure Start children's centre and involves a longitudinal examination of the experiences of practitioners and users. In order to situate the study in the wider context of who might access the service in terms of their social and economic characteristics, a useful place to start is with data from the UK National Census.

Using a tool called CASWEB (see below) it is possible to access aggregate data from the 2001 Census (the most recent at the time of writing). CASWEB provides demographic data on the population living in small geographical areas such as electoral wards, as well as data on health, housing and occupation. So, for example, we can examine the composition of households in the Sure Start area in terms of how many children are present and the employment status of all adults. This will enable the researcher to understand some of the characteristics of families that might be using the Sure Start centre and would be useful in identifying the characteristics of non-participants. Data like this would be very useful for providing a context to the more in-depth study of life at a Sure Start centre.

Example 2: Researching student experiences of taking a gap year

In recent years taking a gap year appears to have become the aspiration of choice for many young people before, or sometimes after, embarking on higher education courses. But who actually takes a gap year, what do they do, and what are the reasons young people give for deciding to take a year out? The Youth Cohort Study (YCS) is an ideal source of information to begin answering some of these questions. The YCS is a series of surveys of young people aged 16 and upwards. Its aim is to identify and explain the factors that influence transitions after compulsory education, such as educational attainment, training opportunities and experiences at school. It also asks young people about their intentions, motivations and experiences of taking a gap year. The YCS is

a useful source for exploring the gap year phenomenon: the data are large scale, of high quality and contain a vast amount of additional information on each individual. Additionally, because those who take a gap year are still a relatively small proportion of the age cohort, collecting such information on a small group of people by other means would, at best, be time consuming and expensive. Accessing the YCS datasets is relatively straightforward, often requiring just institutional access and log in through the UK Data Archive (see below). The data can then be downloaded in a format of your choice (for example in SPSS). Some familiarity with using SPSS can help you manage the data and identify the questions that are most relevant to your study. Analysis of this data can be as complex as you wish: from sophisticated statistical modelling to the more straightforward calculation of frequencies and percentages. As with the Sure Start example above, this analysis would complement in-depth work about the gap year experience.

Secondary data resources

There are a huge number of secondary data sources, both national and international. Below is a very small selection of sources that might be of interest to researchers working in the UK context.

Department for Education (DfE)

This provides access to huge range of administrative data on examination results, class sizes, school buildings, participation and the labour market. The DfE website is at www.education.gov.uk. Follow the links from the Research & Statistics gateway.

Economic and Social Data Service (ESDS)

ESDS provides an access and support gateway to a wide range of economic and social data. It also links to the UK Data Archive. Access to the ESDS datasets requires registration but is free to academic users. The Nesstar analysis tool is also worth looking at, this enables researchers to analyse data without needing to download whole datasets. Data accessible through ESDS and the UK Data Archive include:

- British Social Attitudes Survey,

- Labour Force Survey,

- ONS Opinions Survey

- Millennium Cohort Study,

- Youth Cohort Study

- Longitudinal Study of Young People in England

UK Data Archive

The UK Data Archive curates the largest collection of digital data in the social sciences in the UK. It is part-funded by the ESRC and is a lead partner of the Economic and Social Data Service (ESDS). Its data catalogue provides access to over 5000 datasets for research and teaching purposes across a range of disciplines. It is well worth accessing the UK Data Archive and browsing through its list of datasets, just to appreciate the variety and range of resources that are available. The homepage for the Data Archive is at www.data-archive.ac.uk/.

Neighbourhood Statistics

The Neighbourhood Statistics website draws on social statistical data from a vast range of different sources with a particular focus on deprivation. In addition to providing access to complete datasets, it also generates summary tables that describe the characteristics of the population of a particular geographical area. The database is straightforward to use and requires the user simply to enter the postcode or name of the neighbourhood. It then generates local information on areas such as education, health, crime, employment and housing. Further information on the Neighbourhood Statistics resource is at http://neighbourhood.statistics.gov.uk/dissemination/

Census Area Statistics on the Web

Census Area Statistics on the Web (CASWEB) is a useful interface for accessing data from the UK National Census. It provides coverage of every national census since 1971. Data are presented in tabular form for the range of census topics. Data can be retrieved from the country level all the way down to the smallest level of aggregation – the output area level which, in England and Wales, consists of around 125 households. CASWEB is available at http://www.census.ac.uk/casweb/.

Organisation for Economic Cooperation and Development

The Organisation for Economic Cooperation and Development (OECD) is well known for its publications and statistics covering economic and social issues including trade, health and education. There is a huge amount of data available through the OECD website (www.oecd.org) including e-books, annual compendia of data, aggregate. Selected original datasets can also be downloaded, including those linked to the Programme for International Student Assessment (PISA). A good place to start is with the Education at a Glance publication which is a useful comparative resource on areas such as financial and human investment in education, access to education, learning conditions and educational outcomes.

www.secondarydataanalysis.com

This website provides links to a large range of national and international sources of secondary data in the field of education.

References

Dale, A., Arber, S., Procter, M., (1988), Doing Secondary Analysis, London: Unwin Hyman.

Glaser, B.G., (1963), Retreading Research Materials: The Use of Secondary Analysis by the Independent Researcher, The American Behavioural Scientist, 6(10), p11-14.

Hakim, C., (1982), Secondary analysis in social research: a guide to data sources and methods with examples, London: Allen and Unwin.

Smith, E., (2008), Using Secondary Data in Educational and Social Research, Buckingham: Open University Press.