



PITFALLS AND PROMISES: THE USE OF SECONDARY DATA ANALYSIS IN EDUCATIONAL RESEARCH

Emma Smith

To cite this article: Emma Smith (2008) PITFALLS AND PROMISES: THE USE OF SECONDARY DATA ANALYSIS IN EDUCATIONAL RESEARCH, *British Journal of Educational Studies*, 56:3, 323-339, DOI: [10.1111/j.1467-8527.2008.00405.x](https://doi.org/10.1111/j.1467-8527.2008.00405.x)

To link to this article: <https://doi.org/10.1111/j.1467-8527.2008.00405.x>



Published online: 02 Jul 2010.



Submit your article to this journal [↗](#)



Article views: 3277



View related articles [↗](#)



Citing articles: 46 View citing articles [↗](#)

PITFALLS AND PROMISES: THE USE OF SECONDARY DATA ANALYSIS IN EDUCATIONAL RESEARCH

by EMMA SMITH, *University of Birmingham*

ABSTRACT: This paper considers the use of secondary data analysis in educational research. It addresses some of the promises and potential pitfalls that influence its use and explores a possible role for the secondary analysis of numeric data in the 'new' political arithmetic tradition of social research. Secondary data analysis is a relatively under-used technique in Education and in the social sciences more widely, and it is an approach that is not without its critics. Here we consider two main objections to the use of secondary data: that it is full of errors and that because of the socially constructed nature of social data, simply reducing it to a numeric form cannot fully encapsulate its complexity. However, secondary data also offers numerous methodological, theoretical and pedagogical benefits. Indeed by treating secondary data analysis with appropriate scepticism and respect for its limitations, by demanding that tacit assumptions about the unreliability of secondary data are applied equally to other research methods, and crucially by combining secondary data analysis with small-scale in-depth work, this paper argues for a return to prominence of secondary data analysis in its own right as well as becoming a central component of the new political arithmetic tradition of social research.

Keywords: secondary data analysis, political arithmetic, educational research, research methods

1. INTRODUCTION: WHAT IS SECONDARY DATA ANALYSIS?

Numerous definitions of secondary data analysis appear in the literature, many with subtle differences which together suggest a lack of consensus about what is meant by the term. Some definitions emphasise the usefulness of secondary data analysis for exploring new research questions: 'the study of specific problems through analysis of existing data which were originally collected for another

purpose' (Glaser, 1963, p. 11). However, such definitions appear to disregard the potential of secondary analysis in re-analysing existing data sets with novel statistical or theoretical approaches in such a way that: 'secondary analysis is the re-analysis of data for the purpose of answering the original research questions with better statistical techniques, or answering new research questions with old data' (Glass, 1976, p. 3). One apparent area of consensus among those looking for a definition of secondary analysis is that it should involve the analysis of someone else's data: 'a collection of data obtained by another researcher which is available for re-analysis' (Sobal, 1981, p. 149); but this can be disputed as: 'even re-analysis of one's own data is secondary data analysis if it has a new purpose or is in response to a methodological critique' (Schutt, 2007, p. 4127).

Given the differences in the definition and interpretation of secondary analysis that we see here, it seems likely that neat distinctions between primary and secondary data will not always be possible (Dale *et al.*, 1988). Such lack of consensus might leave one wishing to adopt a very general definition of secondary analysis such as that offered by Jary and Jary (2000): as 'any inquiry based on the re-analysis of previously analysed research data' (p. 540) or one such as Hakim's:

secondary data analysis is any further analysis of an existing dataset which presents interpretations, conclusions or knowledge additional to, or different from, those produced in the first report on the inquiry as a whole and its main results. (Hakim, 1982, p. 1)

Whichever definition one favours, secondary analysis should be 'an empirical exercise carried out on data that has already been gathered or compiled in some way' (Dale *et al.*, 1988, p. 3). This may involve using the original, or novel, research questions, statistical approaches and theoretical frameworks; and may be undertaken by the original researcher or by someone new.

Secondary data can embrace a whole spectrum of empirical forms; it can include the data generated through systematic reviews, through documentary analysis as well as the results from government sponsored surveys. It can be numeric or non-numeric. Non-numeric secondary data could include data retrieved second-hand from interviews, ethnographic accounts, documents, photographs or conversations. However, our interest here is with numeric secondary data: this might include data produced by the decennial national Census; by survey research, such as the Youth Cohort Study or the Social Values Survey; by national or international tests of student performance, such as the Programme for International Student

Assessment (PISA); or indeed by administrative data such as that collected for applications and acceptances to UK Higher Education programmes.

This paper will consider the role of numeric secondary data analysis in contemporary Educational research. It will begin by examining the place of secondary data analysis within current methodological debates, before considering some of the potential pitfalls and the promises that the method presents, and in particular its role in a 'new' political arithmetic tradition of social research.

2. THE USE OF SECONDARY DATA ANALYSIS IN EDUCATIONAL AND SOCIAL RESEARCH

The potential for the secondary analysis of numeric data is huge. From a nation's population census to snap-shot public opinion polls about the outcome of televised talent show competitions: 'nearly every important area of activity and attitude in the British population has now been the focus of a major national survey' (Thomas, 1996, p. 3). Secondary analysis also has a well-established pedigree. In 1790 the first national population Census was undertaken in the USA, followed in Great Britain in 1801. The potential of this data for secondary analysis and its contribution to the social sciences is exemplified by Booth's work on occupation patterns that were derived from secondary analysis of the 1801–1881 UK Censuses (Booth, 1886). By the end of the nineteenth century, the large-scale studies of urban poverty that were pioneered by Joseph Rowntree in York and Charles Booth in London marked the start of the social survey movement and the wealth of opportunities it afforded for secondary analysis. In the United States, secondary analysis as a research strategy coincided with a rapid increase in the number of attitudinal surveys before the Second World War, and where the 'first notable effort' at secondary analysis from a theoretical and methodological perspective was the four volume publication: *The American Soldier* (Glaser, 1963, p. 11). Arguably, however, the potential for secondary analysis as an important social science method has never fully been realised in many branches of the discipline, as many of the objections to its use, some of which will be considered below, attest.

In Education, recent methodological pre-occupations in both the UK and the USA have focused on the quality and relevance of research in the field. Educational research is widely viewed as having an 'awful' reputation (Kaestle, 1993) of being 'not very influential, useful or well funded' (Burkhardt and Schoenfeld, 2003, p. 3), of following fads (Slavin, 1989) and of being of indifferent quality

(Hargreaves, 1996; Tooley with Darby, 1998). While the call in both countries is for a greater unity between research and practice (Burkhardt and Schoenfeld, 2003; Hargreaves, 1996), there is some divergence in how this might actually be achieved. In the USA, legislation introduced in 2001 stipulates that all federally funded research must adopt scientifically-based research methods (Eisenhart and Towne, 2003; Olson and Viadero, 2002). For some this is seen as an opportunity to elevate educational research to the status of medicine and agriculture (Slavin, 2002) and for 'nurturing and reinforcing' a scientific research culture in the field (Feuer *et al.*, 2002, p. 4). For others it exemplifies the privileging of certain research methods: namely experiments and randomised control trials, a failure to understand the complexity of the field and a lack of commitment by the US federal government to promoting true evidence-based practice (Berliner, 2002).

In the UK, general methodological concerns centre on a perceived imbalance in the types of research methods adopted by educational and other social science researchers (ESRC, 2006; Gorard *et al.*, 2003). Much of this concern is centred on the 'dubious dichotomy' (Payne *et al.*, 2004, p. 153) that exists between 'quantitative' and 'qualitative' methods. For example, according to the Economic and Social Research Council (ESRC), 'the lack of quantitative skills is endemic in many areas of Social Science and ... there is an urgent need to enhance research quality' (ESRC, 2006, p. 12). In order to briefly explore these concerns about a lack of methodological pluralism in the use of quantitative methods in educational research and, more importantly for the purpose of this paper, the use of numeric secondary data analysis, a review of the published output of eight mainstream and well-regarded journals in the fields of Education, Sociology and Social Work over a seven-year period was undertaken. The findings are presented in Table 1. Although our primary concern here is with the field of Education, it is useful to examine the extent to which quantitative and secondary data analysis methods are adopted in other areas of the social sciences. Although such an analysis is not unproblematic, for example papers from 'Educational' researchers also appear in 'Sociology' and 'Social Work' journals (Smith, 2008), this is nevertheless a useful and established approach for estimating the frequency with which certain methods are used in research.

About one quarter of all the papers that were reviewed adopted some form of quantitative method (492/2016), of these around 41 per cent (202 out of 492) used secondary data analysis. Among the three disciplines, the use of quantitative methods ranged from

TABLE I: The number of papers using secondary data analysis and quantitative methods: selected Social Science journals

Journal	Secondary Data Analysis	Quantitative methods	Total Papers
British Educational Research Journal	34	85	274
Oxford Review of Education	30	56	220
Research Papers in Education	16	51	133
<i>Education total</i>	<i>80 (42 per cent)</i>	<i>192 (31 per cent)</i>	<i>627</i>
British Journal of Sociology	49	58	201
Sociology	26	37	294
Sociological Review	14	24	211
<i>Sociology total</i>	<i>89 (75 per cent)</i>	<i>119 (17 per cent)</i>	<i>706</i>
British Journal of Social Work	15	95	422
International Social Work	18	86	261
<i>Social Work total</i>	<i>33 (18 per cent)</i>	<i>181 (27 per cent)</i>	<i>683</i>
<i>All journals</i>	<i>202 (41 per cent)</i>	<i>492 (24 per cent)</i>	<i>2016</i>

around one third (192/627) of papers in the 'Education' journals to less than 20 per cent in 'Sociology' (119/706).

While less than 10 per cent (202/2,016) of all papers reviewed involved some analysis of secondary data, in the 'Sociology' journals the majority (75 per cent) of numeric papers did make use of secondary data, including the data from surveys such as the National Child Development Study, the British Family Resources Survey, the Labour Force Study and the European Values Survey. In 'Education' journals, less than half (42 per cent) of the papers which used numeric methods involved the analysis of secondary data. Among those that did, perhaps unsurprisingly the vast majority made use of school performance data, although others used studies such as the Youth Cohort Study, the 1958 British Birth Cohort Study and administrative data produced by the Higher Education Statistics Agency.

This quick analysis of the frequency of use of numeric and secondary analytic techniques in three areas of the Social Sciences reinforces the view of Gorard *et al.*'s (2003) stakeholders that quantitative methods are underused in social science research, although the finding that around one third of the papers in 'Education' journals used numeric methods perhaps gives some cause for optimism. However, in the 'Sociology' journals, where less than 20 per cent of papers adopted quantitative techniques, Payne *et al.*'s (2004) concerns about a lack of methodological pluralism appear to remain true. With regard to secondary data analysis, even though a relatively large proportion of numeric papers in the 'Sociology'

journals adopted this approach, the same cannot be said of the ‘Social Work’ and ‘Education’ journals. Indeed, in the field of Education, although the results of a recent survey of researchers as part of the TLRP Research Capacity Building Network (Gorard *et al.*, 2003) might suggest that large numbers of UK Education researchers *report* using secondary data analysis, very few *actually* appear to use the technique in their published research – further evidence perhaps for a lack of methodological pluralism in this part of the field.

One of the reasons why secondary data analysis is relatively under-used in Educational and Social research can perhaps be attributed to the criticisms and concerns that the field attracts. It is these potential pitfalls of secondary analysis that form the focus of the next section.

3. THE POTENTIAL PITFALLS OF 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; the technical expertise involved in developing good surveys and good datasets can lead to data that is of the highest quality; it can enable data to be analysed and replicated from different perspectives and in this way provides opportunities for the discovery of relationships not considered in the primary research. There are also many perceived pitfalls. And it is these perhaps that have contributed most to the under-use of secondary data analysis and the ‘exaggerated suspicion of social measurement ... and excessive distrust of officially-produced numeric data’ (Bulmer, 1980, p. 505) among the social science community. The very nature of secondary data leaves it particularly susceptible to criticism. For example, it often involves the analysis of data that has been collected with a very different purpose in mind, such as data from the PISA study being used to measure patterns of inequity across Europe (Gorard and Smith, 2004). In interview-based surveys, such as the British Social Attitudes Survey, the secondary data analyst is far removed from the source of the data and may be unaware of, or unconcerned with, the context in which the research took place and the nuanced relationship between the interviewer and respondent. However, the concerns and caveats we attach to the preparation, analysis and interpretation of secondary data are no different to those we should apply to any other type of data: numeric or qualitative, secondary or primary. Here we briefly consider two main objections to the use of secondary data in social research: that it is full of errors, and also that because of the socially constructed nature of social data, the act of reducing it to a simple numeric form cannot fully encapsulate its complexity.

Secondary Data is Full of Errors

‘When you are a bit older’, a judge in India once told an eager young British civil servant, ‘you will not quote Indian statistics with that assurance. The government are very keen on amassing statistics – they collect them, add them, raise them to the n th power, take the cube root and prepare wonderful diagrams. But what you must never forget is that every one of those figures comes in the first instance from the *chowty dar* (village watchman), who just puts down what he damn well pleases’. (In Huff, 1973, p. 84)

The notion that official data are so ‘vitiated with error’ (Bulmer, 1980, p. 508) as to render them unusable for research is widely held. However, this view also suggests that those who do use official statistics are unaware of the potential pitfalls: one cannot assume that ‘the world is made up just of knowledgeable sceptics and naive hard-line positivists’ (ibid.). In the UK, the ‘gold standard’ of survey research is probably the national Census which has taken place every decade since 1801. Huge amounts of resources and expertise are invested in developing Census questions and analytical techniques. Of course the Census is not without its limitations. There are conceptual problems involved in placing people into categories: social class is a good example of this, as is ethnicity. There are also practical problems in ensuring that everyone is counted: how does one account for tourists, for the homeless or transient, for people with multiple nationalities or those who simply refuse to be counted? (Jacob, 1984). In the 1991 Census it was estimated that around 1.2 million people, or 2 per cent of the total number of residents, were uncounted (Dorling and Simpson, 1993). Nevertheless, it is important not to forget that individuals are routinely excluded from research anyway. Consider the research into widening participation in post-compulsory education which focuses only on those actually in education, or the research into exclusion which surveys only those pupils who are present in school, or the research into IT usage which relies on Internet survey methods. To preclude the use of all official data simply on the basis that it may contain error is unrealistic; as Bulmer (1980) suggests, the patterns and trends revealed by interrogating this data can be so striking that if the data was so flawed then what could account for such regularity? Indeed, alongside concerns over the reliability of large scale data is the almost tacit assumption that other data is somehow error free, as with all data, numeric or otherwise, an awareness of its limitations and a ‘healthy scepticism’ (Bulmer, 1980, p. 508; Gorard, 2004; Newton, 2005) about its technical and conceptual basis is essential.

Secondary Data is Socially Constructed

That social data is socially constructed and cannot be reduced to numeric form is another fundamental concern of those who are sceptical about the use of numeric data in social research. The use of 'mere statistical exercises' in reducing the quality of life to numbers and then assuming that these numbers represent reality is, for some social researchers, an anathema. However, without secondary data, and the official data collected by governments and non-government organisations in particular, how would social scientists be able to describe the social world around them, posit theories and test them empirically?

Concern about the social nature of official data has been raised by Vulliamy and Webb (2001) in their research into the support offered to young people who have been excluded from school. In their view, a fundamental problem of using official statistics to measure school exclusions is that the data itself represents a 'considerable underestimate' of the actual numbers of fixed term and permanent exclusions. Reasons for this may include administrative errors in the data, the conflation of figures for exclusion and authorised or unauthorised absences, the 'voluntary' withdrawal of pupils by their parents, as well as schools' use of unofficial exclusions in order to avoid the bureaucratic procedures necessary to make a permanent exclusion. They also argue that school exclusion data is socially constructed and therefore has to reflect the variety of different meanings that can be accorded to it by the participants and the social context involved. Thus, for example, a black student who has been excluded from one school might be treated very differently had he or she attended a different school where teachers perhaps had less stereotypical views of ethnic minority attainment. Indeed, a similar issue arises with exclusions from schools involved in the new Academies programme, where there are fears that some students are more likely to be expelled from an Academy than from another local school (Gorard, 2005).

However, if the above two scenarios were indeed happening, then how would we know whether a disproportionate number of black students or those attending Academies were being excluded if it were not for the secondary analysis of official statistics? One reason that we do not in fact know for sure whether Academies have higher than expected exclusion rates is simply because national data on exclusions at the school level is hard to come by for independent academic researchers: it is not available through the DfES School Performance Tables and its reporting in the National Pupil Database

is not unproblematic. In our conjecture that Academies are excluding higher than expected numbers of students, we have to rely on anecdote and local reports. Without the official data, we cannot produce the national picture to put alongside these localised reports and, if there are injustices, to enable social scientists to challenge policy and demand equity.

That the social world adds complexity to such data should not in itself be a reason not to use it. The argument that this data is socially constructed and can therefore serve no real purpose in helping understand the social world is simply untenable. Secondary data can provide a window to the social world, it can help identify trends and inequities which further inquiry, often using in-depth research methods, can explore. However, as with all forms of data, numeric or otherwise, secondary data, whether official statistics, survey data or in any other form, has to be treated with an appropriate amount of scepticism and respect for its limitations. Official statistics are far too important not to be included in social research. They cannot simply be left for politicians or even the media to do with what they please. It is surely the role of the social scientist to engage with the data, with full understanding of its limitations, and to help establish the link between the empirical data, its social context and the theoretical models that might help explain it.

In the context of school exclusions, secondary data analysis can tell us which groups of students are excluded or at risk of exclusion, from which types of schools and over which particular period of time. Similarly the data from social surveys and government administrative records can tell us which families are living in poverty, the extent of such poverty, whether it has been reproduced over generations, the effect that it has on wider society, for example through educational attainment and health status, and also the impact of government policies to reduce poverty. Although of course while secondary data analysis can tell us *what* is happening in society, it cannot tell us *why* these inequalities exist – that requires combined approaches of the sort described later.

4. THE PROMISES OF SECONDARY DATA ANALYSIS

Much has been made of the potentials of secondary analysis in terms of economies of money, time and personnel (for example, Dale *et al.*, 1988; Glaser, 1962; Gorard, 2002; Hyman, 1972). It is a method that is seemingly perfectly suited to ‘the research needs of persons with macro-interest and micro-resources’ (Glaser, 1963, p. 11). The huge range of topics covered by secondary datasets also adds to its appeal

and, indeed, the availability of large-scale high quality datasets, often free or at nominal cost, is one of the true opportunities presented by secondary analysis. But there is more to secondary analysis than easy access to other people's data for the lazy or impoverished researcher. As this section shows, secondary analysis has a range of applications in teaching as well as in its social, theoretical and methodological contributions.

Social Benefits

As well as methodological and theoretical opportunities, there are also social benefits to secondary analysis. Secondary analysis is an unobtrusive research method. It has the ethical benefit of not collecting additional data from individuals and protecting their privacy by respecting an individual's right to be left alone 'free from searching inquiries about oneself and one's activities' (Bulmer, 1979, p. 4). This, of course, is of particular benefit for research into sensitive issues and of vulnerable and hard to reach groups (Dale *et al.*, 1988; Rew *et al.*, 2000).

Secondary analysis is also a very democratic research method. The availability of low cost, high quality datasets means that secondary analysis can 'restrain oligarchy' (Hyman, 1972, p. 9) and ensure that 'all researchers have the opportunity for empirical research that has tended to be the privilege of the few' (Hakim, 1982, p. 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, p. 12), the very accessibility of the data enables novice and other researchers to retain and develop a degree of independence. Often when researchers are employed on busy projects, there is limited time and resources to apply for grants for other funding and, if successful, there are likely difficulties in securing opportunities for fieldwork. By circumventing the data collection process, secondary analysis can enable novice researchers to gain valuable experience in undertaking independent 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 (for example, Smith, 2005).

In a similar way, secondary analysis also has an important role in teaching, and in research methods teaching, in particular. When teaching the methodologies of survey design, questionnaires from large-scale surveys can be examined for good or indifferent practice

in question wording, scale construction, question ordering (Sobal, 1981) and approaches to data analysis. It is also a useful tool for teaching statistics; students can examine patterns and findings using real data so lending the exercise a degree of relevance. Sampling and issues of generalisability can be taught by comparing the findings from sample surveys with Census datasets (Sobal, 1981). Encouraging students to undertake their own secondary analysis allows them the opportunity to test their hypotheses on good quality, large-scale *real* data (Cutler, 1978; Dale *et al.*, 1988; Sobal, 1981). Additionally, by encouraging students to adopt secondary analysis for at least part of their dissertation research, ethical issues and concerns regarding access to the field and respecting the confidentiality of respondents are reduced or may be avoided entirely.

Moving the Field Forward: Methodological and Theoretical Benefits

Secondary data analysis provides limitless opportunities for the replication, re-analysis and re-interpretation of existing research. It provides researchers with the opportunities to undertake longitudinal analyses, to research and understand past events and to engage in exploratory work to test new ideas, theories and models of research design. Secondary analysis can also enable triangulation with data from other sources, for example, by comparing sample survey results with Census data or the findings of early studies with more contemporary research. Such analysis can also reveal serendipitous relationships in the data (Dale *et al.*, 1988). One outstanding advantage of adopting secondary analysis is that it enables the researcher to access data that is usually of the highest quality. For example, according to Harrop (1980), in general the surveys conducted by professional social scientists working in government organisations are more likely to be of better quality, larger scale and more representative of the general population than the 'local and frequently non-random samples that often form the basis of surveys carried out by academic social scientists, with students often used as interviewers' (p. 15).

Secondary data analysis can allow researchers to gain a second perspective on the data: 'they can ask research questions differently, construct indices differently, analyse the data differently ... or have different theoretical orientations' (Cook, 1974, p. 162), in such a way that it has the potential to uncover errors in the original analysis. As well as challenging the findings of previous research, secondary analysis also has the potential to reinforce the results of the original analysts, as happened with the re-analysis of the Coleman report in the early 1970s. Although, while this emphasis on re-analysis may be

particularly important for policy makers, it might be less useful for the career advancement of the secondary analyst, where pressure to stand out as an individual thinker may require one to do more than replicate (Cook, 1974).

Of course, replicating another's work does not mean that secondary data analysis is necessarily atheoretical or merely descriptive. Away from the methodological potential of secondary data analysis, there is also its contribution to theory development, where according to Hakim (1982), it can 'allow for greater interaction between theory and empirical data because the transition from theory development to theory testing is more immediate' (p. 170). In removing the lag between research design and analysis, secondary analysis can enable researchers to 'think more closely about the theoretical aims and substantive issues of the study rather than the practical and methodological problems of collecting new data' (Hakim, 1982, p. 16). It is also important to emphasise the role of secondary data in descriptive research. Descriptive studies often have lower status in academic circles than research that tests a model or tries to substantiate a prediction, and often they are viewed as less scientific or as not leading to useful generalisations. This tendency to rush to explain phenomena before determining whether or not they actually exist means that 'many important ... phenomena are under-described and poorly measured ... there is too quick a tendency to jump to theory testing and prediction' (Campbell *et al.*, 1982, p. 78). In the field of Education, two examples come to mind: concerns over a shortage of teachers and the apparent underachievement of boys (Gorard, 1999; Gorard *et al.*, 2006; Smith, 2005). In both, the rush to theory testing has occurred before the phenomenon itself had been adequately measured and described and in both, it was secondary data analysis that was used to question the misperception and suggest possible alternative accounts.

New Political Arithmetic: a Role for Secondary Data Analysis?

The role of secondary data analysis in describing the inequalities in society shows it to be a fundamental component of the political arithmetic tradition of social research. There is a long established tradition of using a 'political arithmetic' approach to social research in the UK. The term was first applied to social research in the seventeenth century with William Petty's treatise on economic and social measurement (Hogben, 1939). It is 'an approach which seeks to describe the current state of society with a view to exposing its inequalities' (Power and Rees, 2006, p. 2). The tradition was named 'political' as

it sought to influence government, a fore-runner perhaps of evidence-based policy making; and 'arithmetic' because it utilised numbers, usually through the interrogation of large-scale datasets. In particular, the tradition sought to reinforce the place of social measurement in government policy and planning as well as in wider public debate (Heath, 2000). At its heart was a concern for equity, and issues of social justice were made implicit (Power and Rees, 2006).

However, this promise of a political arithmetic tradition in which social research was integral to social policy proved to be short lived. One reason for this was the advent, in the late 1960s, of phenomenological and ethnomethodological approaches to social research which challenged the orthodoxy of social statistics – the heart of political arithmetic approaches – and questioned the place of numerical data as 'objective fact about social reality' (Miles and Irvine, 1979, p. 116). Recently there has been a renewed interest among some social scientists in conducting social research in the 'political arithmetic' tradition. In a debate in the 2004 volume of the *British Journal of Sociology*, Lauder, Brown and Halsey set out their vision for a policy-oriented Sociology which links sociological theory with empirical data to inform, challenge and hold government accountable (Lauder *et al.*, 2004). This notion is set out explicitly when applied to combining methods in Educational research by Gorard (2002). Here he argues for a 'new' political arithmetic model, which in its simplest form would allow research findings to be combined in a two stage process:

In the first stage, a problem (trend, pattern, or situation) is defined by a relatively large-scale analysis of relevant numeric data. In the second stage this problem (trend, pattern, or situation) is examined in more depth using recognised 'qualitative' techniques with a subset of cases selected from the first stage. (Gorard, 2002, p. 351)

In the spirit of the political arithmetic approach, the numeric techniques adopted in this model would be relatively straightforward and largely descriptive, the combination and integration of smaller scale in-depth work would encourage inter-disciplinarity and the exchange of ideas, theories and perspectives between researchers of different methodological and substantive persuasions. It is an approach which seeks to transcend the traditional 'qualitative' and 'quantitative' paradigms (Gorard, 2002) and one which is sympathetic to the central role of quality and application in Educational and Social research. It is also one that holds secondary data analysis at its centre. While the combination of 'quantitative' and 'qualitative' approaches advocated by the new political arithmetic approach may be being made more

explicit in this round of the debate, as we have seen here it is still the case that secondary data analysis, which is arguably core to the political arithmetic tradition, is much under-used in social research.

5. CONCLUSION

Secondary data analysis can help save time, money, career, degrees, research interests, vitality and talent, self images and myriads of data from untimely, unnecessary and unfortunate loss. (Glaser, 1963, p. 14)

Secondary data analysis offers social, methodological and theoretical benefits. However, it remains a relatively under-used methodological technique in the Social Sciences (see also Pappasolomontos and Christie, 1998). It is also a technique that is open to much criticism. As we have seen in this paper, numeric secondary data has been censured for reducing the complexity of social experiences to mere quantities and for being 'vitiated' (Bulmer, 1980, p. 508) with errors. In defence of secondary data, we have suggested that without it how can social scientists describe the social world around them, posit theories and test them empirically? Official data, as reported by governments, should be used to ensure accountability, within this there is a role for social scientists in informing the collection of social data, for developing social indicators and providing theoretical justifications for the use, or exclusion, of different social categories. Numeric social data can never be error free, neither can the producers of the data prevent it from being used to make un-warranted comparisons, but neither should it be disregarded on this account. Rather it should be treated with the appropriate scepticism and attention to its limitations that we should apply to any social data: primary or secondary, numeric or qualitative.

Arguably secondary data analysis is most effective when combined with other approaches, most notably in the 'new' political arithmetic tradition of research. Indeed, it is the perfect compliment to a 'new' political arithmetic tradition of conducting social research: its scale aids generalisability, the numeric techniques needed for its analysis can be relatively straightforward and accessible to most social scientists – not just statisticians. It can be readily combined with in-depth approaches and the very nature of many large-scale datasets can reinforce the desire for social equity which is at the heart of the political arithmetic tradition. But none of this is particularly new. The call for combining large and small-scale approaches to social research predates much of the research and methodological debates considered in this paper, yet secondary data analysis is still an under-used technique.

It remains to be seen whether there is a role for secondary data analysis in this new round of the political arithmetic debate.

6. REFERENCES

- BERLINER, D.C. (2002) Educational research: the hardest science of all, *Educational Researcher*, 31 (8), 18–20.
- BOOTH, C. (1886) Occupations of the people of the United Kingdom 1801–1881, *Journal of the Statistical Society of London*, 49 (2), 314–444.
- BULMER, M. (1979) Introduction. In M. BULMER (Ed.) *Censuses, Surveys and Privacy* (London, Macmillan).
- BULMER, M. (1980) Why don't sociologists make more use of Official Statistics? *Sociology*, 14 (4), 505–524.
- BURKHARDT, H. and SCHOENFELD, A.H. (2003) Improving educational research: toward a more useful, more influential and better-funded enterprise, *Educational Researcher*, 32 (9), 3–14.
- CAMPBELL, J.P., DAFT, R.L. and HULIN, C.L. (1982) *What to Study: Generating and Developing Research Questions* (Beverly Hills, Sage).
- COOK, T.D. (1974) The potential and limitations of secondary evaluations. In M.W. APPLE, M.J. SUBKOVIAK and H.S. LUFLE (Eds) *Educational Evaluation: Analysis and Responsibility* (California, McCutchan).
- CUTLER, S.J. (1978) Instructional uses of the general social surveys, *Contemporary Sociology*, 7 (5), 541–545.
- DALE, A., ARBER, S. and PROCTER, M. (1988) *Doing Secondary Analysis* (London, Unwin Hyman).
- DORLING, D. and SIMPSON, S. (1993) Those missing millions: implications for social statistics of undercount in the 1991 Census, *Radical Statistics*, 55, 14–35.
- EISENHART, M. and TOWN, L. (2003) Contestation and change in national policy on 'scientifically based' education research, *Educational Researcher*, 32 (7), 31–38.
- ESRC (2006) *Economic and Social Research Council Delivery Plan 2006*. Available at www.esrc.ac.uk/ESRCInfoCentre/Images/ESRC_Delivery_Plan_06_tcm6-15443.pdf (accessed April 2007).
- FEUER, M.J., TOWNE, L. and SHAVELSON, R.J. (2002) Scientific culture and educational research, *Educational Researcher*, 31 (8), 4–14.
- GLASER, B.G. (1962) Secondary analysis: a strategy for the use of knowledge from research elsewhere, *Social Problems*, 10, 70–74.
- GLASER, B.G. (1963) Retreading research materials: the use of secondary analysis by the independent researcher, *The American Behavioural Scientist*, 6 (10), 11–14.
- GLASS, D.V. (1950) The application of social research, *British Journal of Sociology*, 1 (1), 17–30.
- GLASS, G.V. (1976) Primary, secondary and meta-analysis of research, *Educational Researcher*, 5 (10), 3–8.
- GOLDTHORPE, J.H., LLEWELLYN, C. and PAYNE, C. (1980) *Social Mobility and Class Structure in Modern Britain* (Oxford, Clarendon Press).
- GORARD, S. (2002) The role of secondary data in combining methodological approaches, *Educational Review*, 54 (3), 231–237.
- GORARD, S. (2005) Academies as the 'future of schooling': is this an evidence-based policy? *Journal of Education Policy*, 20 (3), 369–377.
- GORARD, S. and SMITH, E. (2004) An international comparison of equity in education systems, *Comparative Education*, 40 (1), 15–28.

- GORARD, S. with TAYLOR, C. (2004) *Combining Methods in Educational and Social Research* (Maidenhead, Open University Press).
- GORARD, S. REES, G. and SALISBURY, J. (1999) Reappraising the apparent underachievement of boys at school, *Gender and Education*, 11 (4), 441–454.
- GORARD, S., TAYLOR, C., RUSHFORTH, K. and SMITH, E. (2003) *What is the Research Capacity of the UK Education Research Community? Reconsidering the Shortage of Quantitative Skills Phenomenon*, Occasional Paper Series, Paper 56 (Cardiff, Cardiff University School of Social Sciences).
- GORARD, S., SEE, B.H., SMITH, E. and WHITE, P. (2006) *Teacher Supply: the Key Issues Workforce* (London, Continuum).
- HAKIM, C. (1982) Secondary analysis and the relationship between official and academic social research, *Sociology*, 16 (1), 12–28.
- HARGREAVES, D.H. (1996) *Teaching as a Research-based Profession: Possibilities and Prospects* (London, Teacher Training Agency).
- HARROP, M. (1980) Social research and market research: a critique of a critique, *Sociology*, 14, 277–281.
- HEATH, A. (2000) The political arithmetic tradition in the sociology of education, *Oxford Review of Education*, 26 (3/4), 313–331.
- HOGBEN, L. (1939) Sir William Petty and political arithmetic. In Lancelot Hogben *Dangerous Thoughts* (London, Allen and Unwin).
- HUFF, D. (1973) *How to Lie with Statistics* (London, Pelican Books).
- HYMAN, H.H. (1972) *Secondary Analysis of Sample Surveys: Principles, Procedures and Potentialities*, London: John Wiley & Sons.
- JACOB, H. (1984) *Using Published Data: Errors and Remedies* (Beverly Hills, Sage).
- JARY, D. and JARY, J. (2000) *Collins Dictionary of Sociology*, 3rd edition (Glasgow, Harper Collins).
- KAESTLE, C.F. (1993) The awful reputation of education research, *Educational Researcher*, 22 (1), 23, 26–31.
- LAUDER, H., BROWN, P. and HALSEY, A.H. (2004) Sociology and political arithmetic: some principles of a new policy science, *British Journal of Sociology*, 55 (1), 3–22.
- MILES, I. and IRVINE, J. (1979) The critique of official statistics. In J. IRVINE, I. MILES and J. EVANS (Eds) *Demystifying Social Statistics* (London, Pluto Press).
- OLSON, L. and VIADERO, D. (2002) Law mandates scientific base for research, *Education Week*, 21 (1), 14–15.
- NEWTON, P.E. (2005) The public understanding of measurement inaccuracy, *British Educational Research Journal*, 31 (4), 419–442.
- PAPASOLOMONTOS, C. and CHRISTIE, T. (1998) Using national surveys: a review of secondary analysis with special reference to education, *Educational Research*, 40 (3), 295–310.
- PAYNE, G., WILLIAMS, M. and CHAMBERLAIN, S. (2004) Methodological pluralism in British sociology, *Sociology*, 38 (1), 153–163.
- POWER, S. and REES, G. (2006) Making sense of changing times and changing places: the challenges of the new political arithmetic of education. Paper presented at the British Educational Research Association Annual Conference, University of Warwick, 6–9 September 2006.
- REW, L., KONIAK-GRIFFIN, D., LEWIS, M.A., MILES, M. and O’SULLIVAN, A. (2000) Secondary data analysis: new perspectives for adolescent research, *Nursing Outlook*, 48, 223–229.
- SCHUTT, R.K. (2007) *The Blackwell Encyclopaedia of Sociology*, Volume III (Ed. G. RITZER) (Oxford, Blackwell).
- SLAVIN, R.E. (1989) PET and the pendulum: faddism in education and how to stop it, *Phi Delta Kappan*, 70, 752–758.

- SLAVIN, R. (2002) Evidence-based educational policies: transforming educational practice and research, *Educational Researcher*, 31 (7), 15–21.
- SMITH, E. (2005) *Analysing Underachievement in Schools* (London, Continuum).
- SMITH, E. (2008) *Using Secondary Data in Educational and Social Research* (Milton Keynes, Open University Press).
- SOBAL, J. (1981) Teaching with secondary data, *Teaching Sociology*, 8 (2), 149–170.
- THOMAS, R. (1996) Statistics as organizational products, *Sociological Review Online*, 1 (3), 1–13. Available online at www.socresonline.org.uk/socresonline/1/3/5.html (accessed June 2007).
- TOOLEY, J. with DARBY, D. (1998) *Educational Research: A Critique* (London, Office for Standards in Education).
- VULLIAMY, G. and WEBB, R. (2001) The social construction of school exclusion rates: implications for evaluation methodology, *Educational Studies*, 27 (3), 358–370.
- WILES, P. (2004) Policy and sociology, *British Journal of Sociology*, 55 (1), 31–34.

Correspondence

Dr Emma Smith
School of Education
University of Birmingham
Birmingham B15 2TT
E-mail: e.smith@bham.ac.uk