

Chapter 11

A regrettable oversight or a significant omission?

Ethical considerations in quantitative research in education

Keith Jones

Policy decisions about the teaching of school subjects such as literacy and mathematics are increasingly becoming intensely political. At the centre of these debates are often the findings of quantitative research in education, particularly the results of national and international surveys of pupil achievement. For example, McQuillan (1998) has attempted to debunk what he identifies as common myths about levels of literacy in the United States of America, such as the claim that students in the United States are among the worst readers in the world. In terms of mathematics teaching, Brown (1998) has revealed the manipulation of international comparisons for political ends, particularly the data from the Third International Mathematics and Science Study (TIMSS), a survey which claims to be the largest and most ambitious international comparative study of student achievement yet undertaken. Indeed, the debate about mathematics teaching in the United States has come to be dubbed the 'Math Wars'. Such has been the ferocity of these 'Math Wars' that the US Secretary of Education has requested a 'cease-fire' saying that there is a need

to bring an end to the shortsighted, politicized, and harmful bickering over the teaching and learning of mathematics. I will tell you that if we continue down this road of infighting, we will only negate the gains we have already made -- and the real losers will be the students of America.

(Riley, 1998)

These are strong words and convey a clear message. It seems that some of the fallout from the TIMSS study, and similarly, perhaps, from studies of literacy standards, could be harming the educational chances of some of the very students the studies sought to benefit. This goes to the heart of ethical issues in educational research whether quantitative or qualitative. Yet in examining, for instance, the TIMSS technical reports (Martin and Kelly, 1996; Martin and Kelly, 1997), the TIMSS Quality Assurance in Data Collection report (Martin and Mullis, 1996), and the

monograph detailing the research questions and the study design (Robitaille and Garden, 1996), there appears to be no explicit consideration of ethical issues. The question this chapter addresses is how the treatment of ethical concerns within the field of quantitative research is constructed and practised and whether this apparent lack of attention to research ethics is a regrettable oversight or a significant omission.

The chapter begins by reviewing a range of standard texts on quantitative research in education, to show the consideration such texts give to ethical issues in the research process. This is followed by a review of the key ethical considerations inherent in quantitative approaches to educational research. Finally, this range of issues is used to examine a number of reports of quantitative-based research in education, including research on gender differences in achievement, teaching methods, intelligence testing, and school effectiveness. The overall aim of the chapter is to reveal the situated nature of research ethics and how they are practised in quantitative research in the field of education.

ETHICAL CONSIDERATIONS IN RESEARCH TEXTS IN EDUCATION AND THE SOCIAL SCIENCES

There is no doubt that an awareness of ethical issues within educational research has grown in recent years and this is reflected in the prominence given to such concerns within standard texts on educational research methods (for a review see Dockrell, 1988). For example, in her book *Research Methods in Education and Psychology*, Mertens places ethical considerations in chapter 1, suggesting that such issues are ‘an integral part of the research planning and implementation process, not viewed as an afterthought or a burden’ (Mertens, 1998:23). While such contemporary general texts address ethical issues in the research process, the situation is quite different if we turn to texts that contain a reference to statistics in their title. For example, none of the following popular texts have any explicit mention of ethics even in their indexes:

Agresti and Finlay, *Statistical Methods for the Social Sciences* (Agresti and Finlay, 1997, 3rd edition)

Glass and Hopkins, *Statistical Methods in Education and Psychology* (Glass and Hopkins, 1996, 3rd edition)

Healey, *Statistics: a tool for social research* (Healey, 1996, 4th edition)

Howell, *Statistical Methods for Psychology* (Howell, 1997, 4th edition)

Pagano, *Understanding Statistics in the Behavioral Sciences* (Pagano, 1994, 4th edition)

Of course, these are revised editions of earlier texts, and so ethical issues may not yet have emerged as a major concern for such texts. Yet even newly published texts on research methods in the social sciences that contain a reference to statistics in their title have no explicit mention of ethics even in their indexes. Some examples are:

Bartholomew, *The Statistical Approach to Social Measurement*

(Bartholomew, 1996)

Hinton, *Statistics Explained: a guide for social science students* (Hinton, 1995)

Peers, *Statistical Analysis for Education and Psychology* (Peers, 1996)

Plewis, *Statistics in Education* (Plewis, 1997)

The only book on research methods in the social sciences, as revealed by the small-scale survey carried out in preparation for this chapter, that has statistics in its title and that explicitly mentions ethical issues seems to be Coolican's *Research Methods and Statistics in Psychology* (Coolican, 1994). Even then the coverage of ethical issues is within the last chapter and is more concerned with general issues, such as confidentiality, deception, stress and discomfort, rather than the particular consideration of statistical techniques.

Where texts are designed to promote particular quantitative approaches to educational research, such as Neuman and McCormick's *Single-subject Experimental Research: applications for literacy* (Neuman and McCormick 1995), explicit ethical considerations are again notable by their absence. Furthermore, in Creswell's general text on qualitative and quantitative approaches to research design, the only mention of research ethics in the index is for 'ethical issues in qualitative study' (Creswell, 1994:148).

In contrast to this situation in educational research, practice in other fields employing quantitative research methods is gradually being informed by ethical considerations given to statistical approaches. An example from medicine is Kadane's *Bayesian Methods and Ethics in Clinical Trial Design*, which presents a methodology for clinical trials that claims to accommodate both ethical and scientific imperatives (Kadane, 1996).

The above review, though only based on a small-scale survey, appears to reveal that while contemporary general texts on educational research methods explicitly address ethical issues, such considerations do not extend to texts that cover statistical procedures. Of course, no text can cover everything, yet the contrast in coverage revealed here is quite striking. It also accords with the apparent omission of explicit ethical considerations within the TIMSS study. This evidence suggests that while texts on statistical techniques employed in quantitative research in education usually

emphasise the importance of using such approaches appropriately, they do not devote much space to discussing what ‘appropriate’ means. Such a discussion would entail a consideration of appropriate for whom, who gets to benefit, what is the potential utility and consequences of the findings or of the results. The implication is that the use of such quantitative techniques is possibly seen as ethically neutral and so an ethical consideration of them can be omitted.

The purpose of this section is not to repeat the debate about the appropriateness of quantitative approaches to educational research at the general level. Such issues are more than adequately dealt with elsewhere (for example in chapter 9 of Hammersley, 1992; or, in relation to research ethics, in Sammons, 1989). Rather, the aim here is to reveal how, in the practice of quantitative research in education, ethical issues are located in the practice of such research. After all, quantitative approaches account for a sizeable proportion of published research in education. For example, in an analysis of 2882 published research articles a third of the articles could be classified as quantitative (Weiner, 1998).

Such articles of the type identified by Weiner provide the empirical site for an examination of what situated ethics means within quantitative research in education and how ethics work in practice. What the articles do not necessarily do is provide the framework for revealing their ethical practices.

ETHICAL CONSIDERATIONS IN CONDUCTING AND REPORTING QUANTITATIVE RESEARCH IN EDUCATION

Sackett (1979) identified 56 possible ‘biases’ that may arise in what he refers to as ‘analytic research’. Over two-thirds of these biases relate to aspects of study design and execution.

Table 11.1: Sites of possible bias in the research process

<i>Stage in research process</i>	<i>Number of possible biases</i>
planning	5
design	22
execution/data collection	18
data processing	-
data analysis	5
presentation	-
interpretation	6
publication	-

Source: Data from Sackett (1979) as displayed by Altman (1980), reproduced by kind permission of the *British Medical Journal*

Altman reveals, in a series of articles published in the early 1980s in the *British Medical Journal* (Altman 1980 and 1981), the distribution of 'biases' suggested by Sackett reflects, to a close degree, the relative seriousness of statistical errors at each stage of the research process, and indicates where there is greatest need for statistical expertise - in the design of quantitative research. Errors in analysis or interpretation can usually be rectified, if detected in time (for example, before publication), but deficiencies in design are nearly always irremediable. Altman reports that various analyses of quantitative-based articles published in medical journals during the 1970s found that between 44% and 72% of such articles had statistical errors or important errors of omission (Altman, 1981: 45). Such errors impact on people, both those who are subjects and those effected as a outcome of the research. No wonder that Altman demonstrates, and the medical profession concludes, that the misuse of statistics is unethical.

The review that follows of the ethical issues pertinent at each stage of the quantitative research process is informed by the 'Declaration on Professional Ethics' of the International Statistical Institute (1986), which, amongst other things, advises that quantitative researchers and statisticians:

- should consider the likely consequences of collecting and disseminating various types of data and should guard against predictable misinterpretations or misuse.
- should not exaggerate the accuracy or explanatory power of their data
- should alert potential users of their data to the limits of their reliability and applicability

This review is also informed by the Ethical Guidelines for Statistical Practice' of the American Statistical Association (1998), which states that that quantitative researchers and statisticians should:

- present their findings and interpretations honestly and objectively
- avoid untrue, deceptive, or undocumented statements
- collect only the data needed for the purpose of their inquiry
- be prepared to document data sources used in an inquiry; known inaccuracies in the data; and steps taken to correct or to refine the data, statistical procedures applied to the data, and the assumptions required for their application

Further issues of ethical concern are taken from articles by Altman (1980; 1981), Herrera (1996), Krenz and Sax (1986), Kromrey (1993), Sammons (1989), and Sax (1986). The following sections cover the main categories of ethical consideration in the process of conducting and reporting quantitative research in education.

Research design and sampling

Research design encompasses all the structural aspects of a study. In quantitative research in particular, this can include the nature of the study sample, the sample size, the type of statistical design, the form of quantification employed, and so on. The stage is quite critical to the success of a research study as no amount of sophisticated analysis will later be able to compensate for major design flaws. The major areas of ethical concern at this stage are defining those who are eligible to be studied, and ensuring randomisation. In the latter case, the main ethical problem arises when attempting to balance the welfare of those being studied against potential benefit for them or future groups.

A crucial aspect of quantitative research design is the process of quantifying (or measuring) phenomena. The complexities involved in quantification are considerable and can easily be under-estimated. Bradley and Schaefer (1998), for example, document the limitations of the process of quantification, including the problems associated with defining variables and with methods of measurement.

In terms of sampling procedure, Altman (1980:1336) suggests that ethical issues arise in two ways. First, if the sample is too large, this entails the unnecessary involvement of additional people, wasting people's time and causing possible harm through, for instance, unnecessary testing, not to mention incurring additional costs. Secondly, if too small a sample is chosen the study may well be unable to detect results of practical importance, and so be a complete waste of time, resources, and possibly good-will. A particular fault in some quantitative research, Altman suggests, is the omission of proper consideration of the statistical power of the test being used, and the lack of use of this in calculating sample size. Such an omission can mean that studies looking for differences can miss effects of practical importance. This may well have different consequences for different groups of people.

Data collection, processing, and analysis

When processing data it is necessary to check that all records of numerical data are within reasonable limits. Particular consideration needs to be given to questions of how to deal with outliers (data values that are very different from virtually all other values), particularly if these are to be excluded from analysis, and when dealing with missing values. Picking and choosing data just to make the results look better is unethical. The removal of any records from the collection of data requires careful thought and justification.

The incorrect analysis of data, Altman (1980) claims, is probably the best known misuse of statistical methods. Such misuse can apply to most of the standard techniques. For example, with the t-test, problems usually relate to the fact that the data being analysed fails to comply with the assumptions underlying the test; that the two sets of data come from populations that are statistically normal and that have the same variance. With the Chi-squared test, the problem can be too few observations. Correlations can be spurious, while with tests of significance, which are based on the assumption of joint normality of two variables, a decision has to be made about which variable is independent. In regression, problems have several sources. For instance, the underlying assumption is that the dependent variable is normally distributed with the same variance for each value of the independent variable. Problems with regression also come from making predictions outside the known range (this is referred to as extrapolation), with linear regression applied to non-linear data, and with the use of simple regression rather than analysis of variance. The ethical impact of any of these misuses of quantitative procedures can be to exaggerate the accuracy or explanatory power of the data, or even to derive conclusions that are deceptive or untrue.

Ethical issues also arise when selecting which data to analyse. The basic principle should be to analyse according to the original design. Other results that look like they may be interesting should really only be used as pointers to further research. This is a different situation to some qualitative research where the data collected may be rich enough to allow unexpected outcomes to be investigated with confidence and integrity. The nature of quantitative research is that the measures employed may have been designed for quite specific purposes with the consequence that it is unlikely to be appropriate to use them for purposes other than those for which they were designed.

Data Presentation and Interpretation

The presentation of data and results is open to unethical practice in at least two ways. For example, well-known instances when care is needed include the choice of scale for graphs, and adding a regression line to scatter graphs for no legitimate reason. Quantitative indicators also need to be used in an ethical manner. An example might be choosing to quote the standard error of the mean as a measure of variability, rather than the variance, just because the standard error is always much smaller.

It is ethically (and vitally) important that valid interpretation is presented of the results of any study, since misleading conclusions can falsely influence practice and further research. For example, assuming causal relationships can be tempting, even in the absence of any supporting evidence. Yet many associations are not causal; and

misleading associations can occur when each variable is actually correlated with a third, 'hidden' variable.

As is the case with a number of terms used in quantitative research, the term 'significant' is perhaps an ill-chosen one. By definition, one in 20 comparisons of equally effective treatments will, in the long run, be 'significant' at the 5% level. So, accepting all statistically significant results as 'real' is quite unwise. While the issues of ethical concern raised above may appear to be of a technical nature, quantitative-based research in education has had, and continues to have, a considerable impact on both policy and practice.

ETHICAL ISSUES IN REPORTS OF QUANTITATIVE-BASED RESEARCH IN EDUCATION

The ethical guidelines of research associations usually include statements of the form 'educational researchers should aim to avoid ... misrepresentation of evidence, data, findings, or conclusions' (British Educational Research Association, 1992). In terms of quantitative research, the declaration on professional ethics by the International Statistical Institute (ISI, 1986) states that:

one of the most important but difficult responsibilities of the statistician is that of alerting potential users of their data to the limits of their reliability and applicability. The twin dangers of either overstating or understating the validity or generalisability of data are nearly always present. No general guidelines can be drawn except for a counsel of caution. Confidence in statistical findings depends critically on their faithful representation. Attempts by statisticians to cover up errors (see Ryten, 1981), or to invite over-interpretation, may not only rebound on the statisticians concerned but also on the reputation of statistics in general.

In this section, the range of issues identified in the previous section concerning quantitative research is used to examine a number of reports of quantitative-based research in education. These examples comprise research on intelligence and schooling, gender differences in achievement, teaching methods, and school effectiveness. A return will also be made to research on international comparisons. These examples have been chosen because of their impact on educational policy and practice. Through such examples, the intention is to reveal the situated nature of ethical principles in the practice of quantitative research in education.

Intelligence and schooling

No study of the ethics of quantitative research in education could be complete without some mention of the mountain of research carried out into the study of

intelligence and schooling. Indeed, the study of intelligence has been intimately involved with the development of statistical techniques. Unfortunately, one of the most infamous cases of fraud accusation also occurs within research on intelligence. This is the case of the work of the eminent educational psychologist Sir Cyril Burt, who, after his death, was claimed to have 'faked' at least some of his data. Such accusations have been refuted by others. For a full discussion of the case see, for example, MacKintosh (1995).

Research in intelligence and schooling continues to court controversy. When it was first published in 1994, Herrnstein and Murray's book 'The Bell Curve' sparked intense public debate. The book claims that it is intelligence levels, not lack of education, environmental circumstances, or poverty, that are at the root of many social problems, and that race-ethnicity and social class differences in intelligence are for the most part determined hereditarily. Such claims of genetic inferiority of certain ethnic and social groups set off what has been called a 'firestorm' of controversy about the assertions, methodology and conclusions of the book. It became a bestseller, selling out its first print run of 125,000 copies.

In their book, Herrnstein and Murray make extensive use of statistical analyses to support their position. A number of texts (for example, Devlin, 1997; Fischer, 1996; Fraser, 1995; Kincheloe, 1996) provide critical examination of many of the statistical and measurement aspects of 'The Bell Curve', raising concerns about, for instance, the appropriateness of the causal inferences, aspects of model specification (most notably the absence of measures of education from the models), model fit, and the validity of IQ and socio-economic status measures.

Particular problems have been identified with Herrnstein and Murray's repeated misuse the concept of heritability as if it were an absolute number (for instance, when they 'estimate' that the heritability of IQ is about 60%). It is, in fact, meaningless to talk about heritability as if it were a number, heritability is defined as a ratio (the ratio of genetically determined variance in a population to the total variance). On top of that Herrnstein and Murray then use measures of heritability *within* groups to make assertions regarding the differences *between* groups. Herrnstein and Murray also present regression curves but do not show the scatter of variation around the curves, an essential element of proper statistical presentation. This means that the graphs do not show anything about the strength of the relationships between variables.

Most education researchers would consider education to be an intervening variable in the chain that runs from child background to adult outcomes. Yet Herrnstein and Murray omit any measure of education as a conditioning variable in their basic analysis. On pages 124-5, for instance, they write that 'the role of education versus IQ as calculated by a regression equation is tricky to interpret'. Yet considerable research (documented in Devlin, 1997; Kincheloe, 1996) points to that fact that

education makes a profound difference in attainment such that when students have equal access to high-quality curriculum, teachers, and school resources, disparities in achievement narrow sharply.

In medical research and in the professional practice of statisticians, consideration of research ethics extends to the design and analysis of research studies. In 'research' such as that by Herrnstein and Murray, such considerations are notable by their absence. The emphasis that Herrnstein and Murray place on differences, and their attempt to argue that such differences are due to hereditary factors, has played a part in education research into gender differences in education outcomes. It is to this topic that I turn next.

Research on gender differences in achievement

There is a substantial body of educational research into gender differences in achievement. While some early research (around the turn of the twentieth century) now appears to have been designed solely to demonstrate an apparent intellectual inferiority of women compared to men, the consensus from some current research (for a recent review see Arnot, 1998) seems to be that females have a slight advantage on average in verbal and literary abilities and males have a slight advantage on average in mathematics, and perhaps science. This has led to recent interest in the apparent relative low achievement of boys in language and literacy. The apparent relative low achievement of girls in mathematics continues to be a focus for research (see, for example, the special edition of *Educational Researcher*, June-July 1998), with the differential effects of socialisation being seen as the causal factor (for example, see Campbell and Beaudry, 1998). Yet such research, in that it frequently uses so-called standardised tests to measure achievement, may often have what Walden and Walkerdine (1985: 23) call 'hidden problems'. Indeed, MacIntyre (1997) refers to this area of work as 'a minefield of research issues' because of its 'politicisation, the proliferation of studies without adequate methodologies, and a number of biases'.

Recent analyses reveal that not only are gender differences in the means and variances of achievement scores very small, but they are of only minor practical significance (Nowell and Hedges 1998, Tate 1997). As Walden and Walkerdine (ibid) explain, the focus on gender differences, usually differences that can be quantified and then stated to be real using statistical techniques, has two important consequences. First, similarities are treated as failures to show differences and hence become non-results. Yet similarities may be of more practical importance than differences. Secondly, such research can tend to emphasise even trivial differences. In large-scale surveys, for example, the close link between sample size and statistical power (the probability that a statistical test will detect a difference) means that

trivially small differences can emerge as statistically highly significant. The practical significance of such differences may as a result be overemphasised.

While it might suit some to continue to emphasise the apparent low achievement of girls in mathematics, or to demand more resources for boys in literacy, better theoretically informed and more methodologically sound research is needed, particularly if it can inform policy decisions. As noted above by MacIntyre, the ‘proliferation of studies without adequate methodologies’ appears to indicate the omission of any proper consideration of the main categories of ethical consideration in the process of conducting and reporting quantitative research in education.

Teaching methods

The ways in which the actions of the teacher impact on the learning of the students in their class is reasonably well-documented, at least in general terms (see Brophy 1986 or Sylva 1994 for reviews). One study that made a substantial impact at the time is that by Bennett and colleagues into the relationship between teaching styles and pupil progress (Bennett 1976). This study claimed to show that pupils taught in a ‘traditional way’ as a whole class made significantly more progress than pupils taught using ‘progressive’ teaching methods. It may be that the study continues to influence policy in education given that an expectation of whole-class teaching has recently emerged in the UK in advice from Government and school inspection officials (and see Francis and Grindle 1998). Indeed, this re-emphasis on ‘traditional’ teaching may well be another consequence of the UKs apparently poor performance in international comparisons such as TIMSS, mentioned at the start of this chapter. Yet Bennett and colleagues subsequently reanalysed their data (following critiques by, for instance, Gray and Satterly 1976) and completely changing their conclusions, indicating, instead that the ‘significance of the differences was reduced’ (Aitken, Bennett and Hesketh 1981). This re-evaluation the findings is summarised by Gray and Satterly (1981) as follows:

- differences between teachers within teaching styles was far greater than differences between styles
- differences between teaching styles were so small as to be overwhelmed by differences between other systematic effects
- the direction of differences between styles did not consistently favour one style over another.

The original publication by Bennett caused a media storm and is said to have sold more than 10,000 copies on the first day of UK sale alone (Shipman 1997:31), something relatively unusual for a book reporting education research. However, it was not primarily a research report. It was written in a non-technical style and

presented as if in an attempt to resolve a popular controversy (McIntyre 1976). As a piece of quantitative research, the criticism of its use of statistical techniques was wide-ranging, although probably the most telling was for its use of analysis of covariance rather than, for instance, multiple regression, a technique which, by comparison, leads to more ambiguous results. The riposte from Bennett (Bennett and Entwistle 1977) was that at the time of the research there was no consensus concerning the most appropriate statistical analysis, and that they had received conflicting advice from statisticians. Bennett's later reanalysis (Aitken, Bennett and Hesketh 1981), which appeared in a scholarly journal without the publicity that surrounded the original publication, was prompted by the 'rapid developments in the statistical methods available for the analysis of complex data'. As Shipman (1997:31) remarks, the most significant findings of the re-analysis were 'the recognition of the difficulties in defining teaching styles and establishing how they were linked to pupil progress'.

While Bennett's work was carried out over twenty years ago, research into school (and teacher) effectiveness has been a characteristic of much recent education research, usually taking a quantitative approach. It is to that research I turn next.

Research into 'school effectiveness'

In recent years research into 'school effectiveness' has come to have a profound influence on education policy in many countries. Yet Slee and Weiner (1998:2) argue that, while the aim that all schools should be as effective as possible seems to be common-sense, 'school effectiveness' is, in their view, 'epistemologically problematic and politically promiscuous and malleable'. They claim that 'while purporting to be inclusive and comprehensive, school effectiveness research is riddled with errors' (p. 5). They suggest that school effectiveness research is discriminatory in excluding 'children with special needs, black boys, so-called clever girls', normative and regulatory by operating mainly within narrow sets of (quantifiable) performance indicators, and is bureaucratic and dis-empowering.

As Goldstein (1997) explains, 'school effectiveness' has come to describe educational research concerned with exploring differences within and between schools. Much effort in 'school effectiveness' research has gone into developing 'performance indicators' yet difficulties associated with 'performance indicators' are well known. Goldstein suggests such difficulties are two-fold. First, their use tends to be very narrowly focused on the task of *ranking* schools rather than on that of establishing factors which could *explain* school differences. Secondly, as Goldstein relates, a number of studies have now demonstrated that there are serious and *inherent* limitations to the usefulness of such performance indicators for providing reliable judgements about institutions. Goldstein maintains that much of the existing

school effectiveness research is methodologically weak, so that strong conclusions about why schools become 'effective' or why that 'fail' are difficult to sustain.

A number of the methodological weaknesses with school effectiveness research are described by Rowe and Hill (1995). They point to the failure of many studies to account for the inherently hierarchical or multi-level sampling structure of the data. Clustering poses special problems such as aggregation bias, undetected heterogeneity of regression, mis-estimated parameter estimates, and so on. Such problems lead to an increased probability of making Type 1 errors, falsely rejecting the null hypothesis in tests of statistical significance. Rowe and Hill point to a crucial requirement in quantitative research, that of ensuring the statistical models used are commensurate with the sampling structure of the data to which the models are applied.

The work of Goldstein and other suggests that, with even with the use of the best currently available statistical techniques, most classes and schools cannot be reliably separated for purposes of comparison. Yet the publication of lists of 'effective' schools and 'failing schools' continues, with such publication being either seriously misleading or potentially harmful. As a result Goldstein and Myers (1996), have proposed the development of a code of ethics for performance indicators. The aims of such a code would be to combine avoidance of unwarranted harm, a core ethical issue, with the right to information.

International comparisons of educational achievement

In this section I return, briefly, to the issue of international comparisons of educational achievement with which we began this chapter. As in the case of school effectiveness research, Brown (1998:33) argues that 'the information in international league tables is often too technically flawed to serve as an accurate measure of national effectiveness'. Brown also illustrates how recent events have demonstrated how international comparisons have been manipulated for political ends. For example, the results of the TIMSS assessment of the performance of 13 years olds in mathematics and science was known in May 1996, but a delay in the publication until November 1996 was agreed internationally so that it did not effect the outcome of the American Presidential elections due to take place in the first week of November that year. One country broke the internationally agreed embargo, the UK. The reason, according to Brown, was that UK Government ministers wanted to announce, in June 1996, the inclusion of mental arithmetic and no-calculator papers in UK national tests at eleven and fourteen. Unfavourable international comparisons appeared to provide a legitimate reason for doing so.

The methodological weaknesses in international comparisons include recurring problems with the definition of the sample, curriculum match, the development of suitable tests, and the forms of reporting. As Brown concludes, international comparisons 'have been seized on as providing a perfect justification of whatever moves government, the media, or others have decided need immediate implementation'. In a similar vein, Keitel and Kilpatrick (1999: 254) argue that 'TIMSS threatens to poison for some time the waters of educational policy, as politicians and researchers scramble to take advantage of what TIMSS allegedly says about teaching and learning mathematics in their country'. They suggest that unwarranted inferences are being made concerning the link between selected views of teaching and fallible indicators of performance. They conclude that studies such as the Third International Mathematics and Science Study (TIMSS), 'rest on the shakiest of foundations ...' and '... assume that the mantle of science can cover all the weaknesses in design, incongruous data and errors of interpretation'. As noted earlier, methodological weaknesses can appear to be solely a technical issue, yet following TIMSS there has been intense policy activity in a number of countries which could be harming different groups of pupils and their life chances.

FINAL COMMENTS

What all the above studies in quantitative research in education demonstrate is the complex and subtle nature of ethical issues encountered in educational research (see Cohen and Manion 1994: 347ff). The strength of quantitative approaches to research in education lies in the clarity and rigour of the standard procedures, the clear definitions and control of variables, the opportunity for replications and comparison across different studies and sites. However, one of its weaknesses is that it cannot reveal deep meanings nor, as yet, cope with the full complexity of teaching and learning in all its settings. Even in qualitative research, the concept of arriving at external 'truths' has come under intense scrutiny as interest has grown in the situated nature of meanings and interpretations of events that are given by the participants in the research context. The consideration given to the particular studies chosen in the previous section is not meant to demonstrate that ethical guidelines were deliberately flouted in carrying out the research. Rather, as Vaughn and Lyon (1994) suggest, 'the overriding question ... is at what point does the researcher go astray to such a marked degree that the ethics of his/her research procedures and findings are in question?'

All information, whether systematically collected or not, is subject to misuse. No information is devoid of possible harm to one interest or another. Researchers are not, in general, in a position to prevent action based on interpretations of their data. Indeed, to guard against the use of their findings would be to disparage the very purpose of much educational inquiry which is to improve public discourse and understanding. Quantitative research in education is seductive in its claim to

objectivity, clarity and rigour of established procedures, and techniques to manipulate large samples and a wide range of variables. With such a methodology it is perhaps easy to see how ethical issues may be buried. Yet in some fields, in medical research and in the work of professional statisticians for example, what appear to be technical issues of design, analysis and presentation are considered to have important ethical dimensions. The lack of an explicit consideration of ethical issues in quantitative research in education may once have seemed like a regrettable oversight (Cohen and Manion 1994: xxi). Currently, it looks more like a significant omission.

References

- Agresti, A. and Finlay, B. (1997), *Statistical Methods for the Social Sciences*. Upper Saddle River, NJ: Prentice Hall. 3rd edition.
- Aitken, M., Bennett, S. N. and Hesketh, J. (1981), Teaching Styles and Pupil Progress: a re-analysis. *British Journal of Educational Psychology*, **51**, 171-186.
- Altman, D. G. (1980), Statistics and Ethics in Medical Research: Misuse of Statistics is Unethical; I. Misuse of Statistics is Unethical; II. Study Design; III. How Large a Sample; IV. Collection and Screening Data; V. Analyzing Data; VI. Presentation of Results; VII. Interpreting Results, *British Medical Journal*, **281**; 1182-1184, 1267-1269, 1336-1338, 1399-1401, 1473-1475, 1542-1544, 1612-1614.
- Altman, D. G. (1981), Statistics and Ethics in Medical Research, VIII: Improving the Quality of Statistics in Medical Journals, *British Medical Journal*, **282**, 44-46.
- American Statistical Association (1998), *Ethical Guidelines for Statistical Practice*. This is a revised edition based on the 1989 ASA *Ethical Guidelines for Statistical Practice* and the 1985 *Declaration on Professional Ethics of the International Statistics Institute*.
- Arnot, M. (1998), *Recent Research on Gender and Educational Performance*. London: Stationery Office
- Bartholomew, D. J. (1996), *The Statistical Approach to Social Measurement*. San Diego: Academic Press.
- Bennett, N. (1976), *Teaching Styles and Pupil Progress*. London: Open Books
- Bennett, N. and Entwistle, N. (1977), Rite and Wrong: a reply to 'A Chapter of Errors'. *Educational Research*, **19**(3), 217-22.
- Bradley, W. J. and Schaefer, K. C. (1998), *The Uses and Misuses of Data and Models : the mathematization of the human sciences*. Thousand Oaks: Sage.
- British Educational Research Association (1992), *British Educational Research Association Ethical Guidelines*. London: BERA.
- Brophy, J. (1986), Teacher Influences on Student Achievement. *American Psychologist*, **41**(10), 1069-1077.
- Brown, M. (1998) The Tyranny of the International Horse Race. In R.Slee and G. Weiner, with S. Tomlinson (Eds) *School Effectiveness for Whom? challenges to the school effectiveness and school improvement movements*. London: Falmer Press.
- Campbell, J. R. and Beaudry, J. S. (1998), Gender Gap Linked to Differential Socialization for High- achieving Senior Mathematics Students, *Journal of Educational Research*, **91**(3), 140-147.
- Cohen, L and Manion, L (1994), *Research Methods in Education*. London: Routledge. 4th edition.
- Coolican, H. (1994), *Research Methods and Statistics in Psychology*. London: Hodder and Stoughton. 2nd edition.
- Creswell, J. W. (1994), *Research Design: qualitative and quantitative approaches*. Thousand Oaks: Sage.

- Devlin, B. (Ed) (1997), *Intelligence, Genes, and Success: scientists respond to the bell curve*. New York: Springer Verlag.
- Dockrell, W. B. (1988), Ethical Considerations in Research. In John P. Keeves (Ed), *Educational Research, Methodology, and Measurement: an international handbook*. Oxford: Pergamon.
- Fischer, C. S. (1996), *Inequality by Design: cracking the bell curve myth*. Princeton, NJ: Princeton University Press.
- Francis, L. J. and Grindle, Z. (1998), Whatever Happened to Progressive Education? A comparison of primary school teachers' attitudes in 1982 and 1996. *Educational Studies*, **24**(3), 269-279.
- Fraser, St. (Ed) (1995), *The Bell Curve Wars: race, intelligence, and the future of America*. New York: BasicBooks.
- Glass, G. V. and Hopkins, K. D. (1996), *Statistical Methods in Education and Psychology*. Boston: Allyn and Bacon. 3rd edition.
- Goldstein, H. (1997), Methods in School Effectiveness Research. *School Effectiveness and School Improvement*, **8**(4), 369-395.
- Goldstein, H. and Myers, K. (1996), Freedom of Information: towards a code of ethics for performance indicators. *BERA Research Intelligence*, **57**, 12-16.
- Gray, J. and Satterly, D. (1981), Formal or Informal? A re-assessment of the British evidence. *British Journal of Educational Psychology*, **51**, 187-196.
- Hammersley, M. (1992), *What is wrong with ethnography? methodological explorations*. London: Routledge.
- Healey, J. F. (1996), *Statistics: a tool for social research*. Belmont: Wadsworth. 4th edition.
- Herrnstein, R. J. and Murray, C. (1994), *The Bell Curve: intelligence and class structure in American life*: New York: Free Press.
- Herrera, C. D. (1996), An Ethical Argument against leaving Psychologists to their Statistical Devices. *Journal of Psychology*, **130**(3), 125-130.
- Hinton, P. R. (1995), *Statistics Explained: a guide for social science students*. London: Routledge.
- Howell, D. C. (1997), *Statistical Methods for Psychology*. London: Wadsworth. 4th edition.
- International Statistical Institute (1986), Declaration on Professional Ethics. *International Statistical Review*, **54**, 227-242.
- Kadane, J. B. (1996), *Bayesian Methods and Ethics in Clinical Trial Design*. New York: John Wiley.
- Keitel, C. and Kilpatrick, J. (1999), The Rationality and Irrationality of International Comparative Studies. In G Kaiser, E Luna and I Huntley (Eds), *International Comparisons in Mathematics Education*. London: Falmer.
- Kincheloe, J. L. (Ed) (1996), *Measured Lies: the bell curve examined*. St Martins Press.

- Krenz, C. and Sax, G. (1986), What Quantitative Research Is and Why It Doesn't Work. *American Behavioral Scientist*, **30**(1), 58-69.
- Kromrey, J. D. (1993), Ethics and Data Analysis. *Educational Researcher*, **22**(4), 24-27.
- McIntyre, D. (1976), Review of "Teaching Styles and Pupil Progress". *British Journal of Teacher Education*, **2**(3), 291-297.
- MacIntyre, T. (1997), Gender differences in cognition: A minefield of research issues. *Irish Journal of Psychology*, **18**(4), 386-396.
- McQuillan, J. (1998), *The Literacy Crisis: False Claims, Real Solutions*. Portsmouth, NH: Heinemann.
- MacKintosh, N. J. (Ed) (1995), *Cyril Burt: fraud or framed?* Oxford: OUP.
- Martin, M. O. and Kelly, D. L. (Eds) (1996), *TIMSS Technical Report Volume 1: Design and Development*. Chestnut Hill, Ma: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Martin, M. O. and Kelly, D. L. (Eds) (1997), *TIMSS Technical Report Volume 2: Implementation and Analysis*. Chestnut Hill, Ma: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Martin, M. O. and Mullis, I. V. S. (Eds) (1996), *TIMSS Quality Assurance in Data Collection*. Chestnut Hill, Ma: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Mertens, D. M. (1998), *Research Methods in Education and Psychology: integrating diversity with quantitative and qualitative approaches*. Thousand Oaks: Sage.
- Neuman, S. B and McCormick, S. (1995), *Single-subject Experimental Research: applications for literacy*. Newark: International Reading Association.
- Nowell, A. and Hedges, L. V. (1998), Trends in gender Differences in Academic Achievement from 1960 to 1994: an analysis of differences in mean, variance, and extreme scores. *Sex Roles*, **39**(1-2), 21-43 .
- Peers, I. S (1996), *Statistical Analysis for Education and Psychology*. London: Falmer.
- Plewis, I. (1997), *Statistics in Education*. London: Arnold.
- Riley, R. W. (1998), *The State of Mathematics Education: building a strong foundation for the 21st century*. Speech to the Conference of American Mathematical Society and Mathematical Association of America, Baltimore, Maryland, Thursday, January 8, 1998. Text available in *Notices of the American Mathematical Society*, **45**, 487-490 (1998).
- Robitaille, D. F. and Garden, R. A. (Eds) (1996), *Research Questions and Study Design*. Vancouver: Pacific Educational Press.
- Rowe, K. J and Hill, P. W (1995), Methodological Issues in Educational Performance and School Effectiveness Research: a discussion with worked examples. *Australian Journal of Education*, **39**(3), 217-248.

- Ryten, J. (1981), Some general considerations as a result of an Ecuadorian case study. *Bulletin of the International Statistical Institute* (Proceedings of the 43rd Session), **49**(2), 639-665.
- Sackett, D. L. (1979), Bias in Analytical Research. *Journal of Chronic Diseases*, **32**, 57-63.
- Sammons, P. (1989), Ethical Issues and Statistical Work. In R G Burgess (Ed), *The Ethics of Educational Research*. Lewes: Falmer.
- Sax, G. (1986), *The Ethics of Quantification and Why It Doesn't Work; or Life among the Numerologists, Innumerates, and Qualitatives*. Paper presented at the 70th Annual Meeting of the American Educational Research Association (San Francisco, CA, April 16-20, 1986).
- Shipman, M. D. (1997), *The Limitations of Social Research*. London: Longman. 4th edition.
- Slee, R. and Weiner, G. (Eds) (1998), *School effectiveness for whom? Challenges to the school effectiveness and school improvement movements*. London: Falmer.
- Sylva, K. (1994), School Influences on Children's Development. *Journal of Child Psychology and Psychiatry*. **35**(1), 135-170.
- Tate, W. F. (1997), Race-ethnicity, SES, Gender, and Language Proficiency Trends in Mathematics Achievement: an update. *Journal for Research in Mathematics Education*, **28**(6), 652-679.
- Vaughn, S. and Lyon, G. R. (1994), Ethical Considerations When Conducting Research with Students with Learning Disabilities. In Sharon Vaughn and C Bos (Eds), *Research Issues in Learning Disabilities: theory, methodology, assessment, and ethics*. Berlin: Springer-Verlag.
- Walden, R. and Walkerdine, V. (1985), *Girls and Mathematics: from primary to secondary schooling*. London: Institute of Education. Bedford Way Paper 24.
- Weiner, G. (1998), *Getting Published: an account of writing, refereeing and editing practices*. Final Report to the UK Economic and Social Research Council.