

Education and Social Justice

By Stephen Gorard

242pp Cardiff, Univ. of Wales Press, 2000. £35.00. ISBN 0-7983-1619-0

Reviewed by Harvey Goldstein

Institute of Education

The main aim of this book, based partly upon the author's own research, is to question the received wisdom of the 1990s that educational participation and performance have become increasingly polarised as a result of market reforms pursued during that period; the author seeks to disprove and even reverse some of these views.

In the first chapter the author sets his scene by quoting 'evidence' from various sources about how English and Welsh students perform poorly on international comparisons and how performance has deteriorated over time. He then claims that existing interpretations of data are flawed on the basis of three alternative interpretations. The first involves essentially a presentation of numerical comparisons of percentages on a multiplicative rather than arithmetic basis, what he calls 'proportionate analysis'. He seems totally unaware that debates about how to present such comparisons have a long history and include a much wider discussion than the superficial justification given in this book, and there is no universally best method. The second point he makes is that comparisons, especially those over time, must take into account any underlying population changes; sensible but hardly new. The third is that raw comparisons of performance are misleading and that 'value added' analyses are needed. Again, this is hardly new. The rest of the book applies these interpretations to data, and I shall consider each one in turn and whether Gorard's reanalysis does take the debate forward.

Social segregation

Chapters 2 & 3 address the issue of whether social segregation between schools has increased. Gorard rightly points out the limitations of generalising from small-scale unrepresentative studies and introduces data that include information on social background such as free school meals entitlement. Interestingly Gorard never seems to address the much more important segregation issue with which many have been concerned, namely the issue of increasing *achievement segregation*. Irrespective of social background, the ability of schools to choose children can be expected to lead to increasing differences in achievement, given the emphasis on producing 'good' league table results. The evidence from school effectiveness research is that initial achievement is a far better predictor of final performance than social background, so that a 'rational' school would wish to select on that basis, even resisting efforts to enter 'desirable' schools by the socially advantaged who cannot also demonstrate high achievement. If such mechanisms are at work they may in fact lead to less social segregation, but that would not necessarily make them desirable. What, then, of his analyses?

The data used in these analyses come from administrative school level records and provide information such as the proportion eligible for free meals (a crude measure of social disadvantage), ethnic minorities and special educational needs, supplemented by area level social data. The real problem with this analysis, however, is a technical one. Gorard defines a 'segregation index' as follows.

Within a given area, such as an LEA, if there is no underlying difference between schools in the proportion of disadvantaged children, then the *expected* number of disadvantaged pupils in a school is simply obtained by applying the overall proportion for the area to the number of children in the school. The difference between the *expected* and *observed* numbers in each school is then calculated, the absolute value taken and then these are added together across the area and divided by (twice) the total number of disadvantaged pupils in the area. This index is used to compare LEAs and, crucially, to make comparisons over time; Gorard finds that across England between 1989 and 1997 there is actually a decrease in the value of this index and this is used to argue against the thesis of increasing social segregation. Unfortunately, the value of this index is a function of the size of school, and without controlling for this, differences in the index cannot be interpreted as measuring changes in segregation. This dependence on size becomes negligible as the underlying segregation increases, but in the case where the true segregation is small, as a result of sampling fluctuations this index will be (approximately) inversely proportional to the square root of the size of school. Even where there is no real underlying variation among schools, sampling variations generally will ensure that the index will be greater than zero (contrary to Gorard's claim) by an amount depending on the school sizes (for a set of primary schools of size 200, the expected value of the index for an average of 20% eligible for free meals and where there is no real underlying difference, is about 6%). Thus, for example, consider an LEA with no underlying differences between schools, apart from random fluctuations, measured at two time points. If the number of pupils remains constant but some schools close so that the average size of schools increases by 25%, then the segregation index will decrease by about 10%. In addition, even where the sizes of schools remains unchanged and there is no real difference among schools, an increase in the percentage of those eligible for free school meals will also tend to decrease the index, again approximately inversely proportional to the square root of the percentage. For at least some of the LEAs considered by Gorard, the actual observed values of the index may in fact be consistent (on the basis of a significance test) with no underlying differences among schools.

Over the period he considers, the average size of secondary schools increased by 18% which would tend to reduce the segregation index. Furthermore, amalgamations would also tend to reduce the index; Gorard does mention this possibility but dismisses it without empirical justification. This is not to say, of course, that in some sense segregation has not changed, simply that Gorard's failure to recognise and account for these measurement problems means that his results must remain an unsound basis for drawing conclusions.

International comparisons

Chapter 4 looks at comparative international studies of student achievement. Gorard makes some useful criticisms of some of the wilder claims made on behalf of these studies. He points out the problems of translation and curriculum coverage, although he fails to comment on the fact that these studies are almost all cross sectional which makes it extremely difficult to make any kind of causal inference about educational systems. He also fails to point out that in Mathematics and Science the overall scores mask some quite different sub test comparisons and that reliance on overall scores means that one accepts the actual balance of items (e.g. predominantly arithmetic in Maths) making up the total. Thus, although England does not appear to do well overall it does do very well in problem solving. Gorard's criticisms of school effectiveness research suffer from the same superficiality and apparent unfamiliarity with much of the research literature. Nevertheless, he does repeat some useful criticisms of attempts to make absolute comparisons over time.

In Chapter 5 Gorard compares examination results in England and Wales. He uses data for LEAs and adjusts for LEA level social factors including free school meals. He concludes that there is no evidence for Wales performing worse than England. Unfortunately, again there are technical problems with this analysis. First of all it uses aggregate level data and thus is open to what is known as the 'ecological fallacy' whereby quite different relationships can occur for data fitted to aggregate level units and data fitted to individual pupils. Secondly, the school effectiveness literature has demonstrated quite clearly that the most important variable to adjust for when comparing 'output' performance is intake achievement, social factors simply do not make an adequate adjustment. Gorard seems not to recognise these difficulties so that once more his conclusions are unsound.

Chapter 5 reverts to the 'proportionate analysis' comparisons over time and seeks to reverse existing interpretations, claiming that in fact there is convergence between schools in terms of examination results. For the reasons explained earlier all of this is debateable and it is difficult to rely upon such analyses. An appropriate analysis of 'convergence' would require, at the very least, longitudinal 'value added' data on cohorts of pupils in schools.

Chapter 7 suffers from the same drawbacks as previous chapters in using the segregation ratio, failing to take account of prior achievement when comparing schools and carrying out an aggregate level analysis using schools as units. Chapter 8 uses 'proportionate analysis' to revise conclusions of others. Here, as elsewhere, however, Gorard does make some useful criticisms of other people's simplistic interpretations. In Chapter 9 Gorard makes the important point in comparing boys and girls over time, that entry changes need to be taken into account. He again uses a 'proportionate analysis' and simply claims that this is superior to other methods of presentation and on this basis comes to different conclusions from other researchers.

In his final Chapter Gorard makes several claims on the basis of his analyses; that social segregation is decreasing, that international differences are exaggerated, that there are no clear differences between types or sectors of schooling and that differences between different types of pupils are decreasing over time. Unfortunately, none of these conclusions can be relied upon to follow from Gorard's analyses.

Interestingly, he makes an attack on 'qualitative' researchers who are dismissive of quantitative research, and I have some sympathy with that. But he then goes on to dismiss 'complex methods of analysis' in favour of his own 'middle way'. The trouble is that Gorard's analyses are over-simple, use insecure statistical techniques and actually demonstrate the real dangers of over-simplification. The processes being studied are genuinely complex and only by developing techniques that seek to match that level of complexity can we hope to advance quantitative contributions to knowledge (Goldstein, 1998). Far from providing a step in the right direction this book provides a sad example of how not to approach the analysis of complex educational data.

Reference

Goldstein, H. (1998). *Models for reality*. London, Institute of Education.

A summary response to the review, by Harvey Goldstein, of 'Education and Social Justice'

242pp Cardiff, University of Wales Press, 2000. £35. ISBN 0-7983-1619-0

by Stephen Gorard

In his review Goldstein makes a number of criticisms of the various methodological approaches described in my book. None of these criticism holds up under scrutiny, and I have responded fully to each of these points. For interested readers, the review, the responses and the ensuing discussion are available on my web site (www.cf.ac.uk/socsi/markets) or via email attachment (gorard@cardiff.ac.uk). Here, I shall discuss four examples which are typical of the defects in the review. Goldstein was apparently unfamiliar with the complete contents of the book he was reviewing; used non-standard meanings for crucial methodological terms; ignored the rationale of significance testing; and based some criticisms on one key underlying assumption that is not standard practice, not made in the book, and not made explicit in his review.

In paragraph two of his review Goldstein discusses my argument for greater use of proportions rather than simple differences when making comparisons over time or place. He states that Gorard is 'totally unaware that debates about how to present such comparisons have a long history'. This point is false and can be shown to be false by reference to the pages of the book he is reviewing. The earliest citation to this issue is Wright, J. (1937) 'Some measures of distribution', *Annals of the Association of American Geographers*, 27, 177-211. My book contains over twenty references to the 'index wars' since the Wright paper and the Duncan & Duncan papers of the 1950s (see pages 203-213 for example).

In paragraph eight, Goldstein claims that my analysis in Chapter Five 'is open to what is known as the ecological fallacy'. This fallacy occurs when events are examined at one level of aggregation, while the conclusions are drawn at a lower level. For example, it can not be deduced that an individual student, from a school with good examination results, will necessarily have good examination results. However, in the book I quite properly present conclusions at one level of aggregation based on analyses at the *same* level. Chapter Five is about LEAs, and does not involve an ecological fallacy - or indeed a fallacy of any sort. Goldstein's argument, as subsequently explained, is that it might have been better for me to have conducted the analysis at a lower level of aggregation since there may be interesting variations to examine. This is what he terms the 'ecological fallacy', claiming that it is part of the 'traditional interpretation' of that term. Goldstein is almost certainly right about the sub-LEA variations (which are anyway addressed in the book by Chapter Seven on schools and Chapter Nine on pupils), although he ignores the practical consideration that the required data are not always available at a more detailed level. But our debate is clearly about selecting an appropriate level for analysis, which is a matter of judgement and not an 'ecological fallacy' on any interpretation. Like Humpty-Dumpty, Goldstein is using this philosophical term imprecisely. He is therefore wrong to suggest to review readers that my work contains the fallacy that they will understand from the term used.

In paragraph five Goldstein suggests that some of the differences I report between/within LEAs might be spurious 'on the basis of a significance test' which I had failed to do. Null hypothesis significance testing

(NHST) is used by statisticians to try and separate any variation due to random sampling from variation due to design bias or 'real' effect. It is, therefore, only appropriate for use when a random sample is used. My analysis is for a population. Goldstein either does not realise that the data refer to populations, or is clinging to a, happily near-obsolete, practice where NHSTs are used routinely without regard to the nature of the problem at hand.

In paragraph five Goldstein also criticises the segregation index that I use in my work, claiming that 'the value of this index is a function of the size of school', and suggests that without controlling for changes in school size the results are therefore unreliable. While more interesting than those above, this criticism also has no basis in fact, revealing a somewhat limited consideration of the measurement of segregation. My segregation index is mathematically equivalent to the Hoover coefficient of income inequality, the Equal Opportunity Commission's achievement gap, Cortese's 'exchange proportion', and other variations. All of these have the same property of strong compositional-invariance, such that scaling up or down in any analysis leaves the resulting coefficient unchanged. Of course, where only a sub-group of schools change in size this *can* change the value of the segregation index. When this happens it is because the nature of segregation has changed in fact - by altering the proportion of students who would have to change schools for there to be no segregation. This is what the index measures, and it is made plain in the book that a partial explanation for my findings is the variation in number of schools over time (see page 45 for example). The segregation index, all of those listed above, and all other popularly employed approaches including the Gini coefficient and the dissimilarity index have the same property (although unlike my index the latter is not strongly invariant). Interested readers might like to note my forthcoming journal article in 'Sociology' entitled 'What is segregation? A comparison of indices in terms of compositional invariance'. Some commentators in education have confused my index with the dissimilarity index, incorrectly attributing characteristics of the latter to the former. Perhaps that has happened again here.

In summary, Goldstein makes a number of criticisms of the work described in my book - although his 'review' makes no mention of whether he considers the book well-written, easy to follow, interesting, transparent in method, or important in topic. All of his criticisms are unfounded, and most can be shown to be so in relatively simple ways. The first type can be checked by reference to the pages of the book itself, the second by reference to a dictionary of terms, and the third by reference to a textbook of statistics. The fourth issue is more complex, and there is insufficient space here to take the reader through the comparisons between indices which are made in the book. Suffice to say that rather than simply stating that the segregation index has a problem, Goldstein needs to convince the reader of this by detailed calculation, or reference to other published work. He will also wish to explain why his approach is so different from that of Marshall, Taeuber, Duncan, Lieberman, Goldthorpe, Gilbert, Blackburn, Hakim, and many others who have been fighting the index wars since 1937. Given that the work described in this book has all been published in peer-reviewed international journals of education and social science, perhaps these would also provide a more appropriate arena in which for Goldstein to present such an argument, rather than a book review.

Send reply to: <h.goldstein@ioe.ac.uk>
From: "Harvey Goldstein" <h.goldstein@ioe.ac.uk>
To: "Stephen Gorard" <Gorard@Cardiff.ac.uk>
Subject: RE: BJES
Date sent: Wed, 20 Jun 2001 11:18:42 +0100

I do accept, as you point out, that you are aware of the issues, and I was wrong to claim that you were completely unaware, but that isn't quite the point. As I said, you can't have it both ways - claiming fairly unequivocally that your method is better and also that there is considerable debate about a choice and in the appendix expressing at least hesitation about which one to use. If our debate is to continue, as I think it should, it would be useful to air this issue.

Harvey

Harvey Goldstein
Institute of Education
20 Bedford Way
London, WC1H 0AL, UK

-----Original Message-----

From: Stephen Gorard [<mailto:Gorard@Cardiff.ac.uk>]
Sent: 20 June 2001 12:02 PM
To: h.goldstein@ioe.ac.uk
Subject: RE: BJES

I am somewhat disappointed by this response, but keen to keep the debate going. Why not start with your first substantive criticism? Can you now agree that it is not true that I am 'unaware' of the long-term debate over presentation of differences/changes? This seems to me to be a simple matter of fact. I am aware, and the evidence for this is in the book. We may still disagree, or you may think I have misunderstood but that is healthy and an entirely different matter altogether. To say that I am unaware must be wrong. Or do you still stick to your first criticism in the review?

Best wishes,

Stephen

Send reply to: <h.goldstein@ioe.ac.uk>
From: "Harvey Goldstein" <h.goldstein@ioe.ac.uk>
To: "Stephen Gorard" <Gorard@Cardiff.ac.uk>
Subject: RE: BJES
Date sent: Fri, 22 Jun 2001 11:58:37 +0100

Stephen

It is one aspect of the ecological fallacy as traditionally understood. The 'fallacy' arises because there is an implied level (whether at that of the school or the pupil) which is where people are usually most interested in an interpretation. In the example I quoted of your table, a fuller story exists at the level of the school which is highly relevant to our understanding. Quoting only the aggregate level result as if it is the only relevant comparison is to commit the ecological fallacy. If you had pointed all of this out and, ideally, carried out an appropriate lower level set of analyses I would have little quarrel. This would be a good issue to discuss in a seminar.

Harvey

Harvey Goldstein
Institute of Education
20 Bedford Way
London, WC1H 0AL, UK

-----Original Message-----

From: Stephen Gorard [<mailto:Gorard@Cardiff.ac.uk>]
Sent: 22 June 2001 12:41 PM
To: h.goldstein@ioe.ac.uk
Subject: RE: BJES

Ah! But this is rather different both from the traditionally understood 'ecological fallacy' you invoke in the review - which will be interpreted by readers to mean analysing data at one level and drawing conclusions at another (references to the meaning of the term available if required).

Clearly at whatever level we do study phenomena there could be, and usually is, considerable variation at levels below (and since you are so widely read you will know that I have argued this on many occasions). To accept this is not a fallacy, since there is no theoretical lowest level of analysis. You could say the same thing at whatever level someone works - schools are misleading, and should be pupils, or pupil:teacher combinations, or pupil:teacher:environment, or pupil:teacher:environment:subject:day-of-week:time-of-day etc.

An argument that any level hides potentially interesting lower level variation is valid. However, this does not imply a fallacy. It is merely a plea for further data/work. In the real world of research we work with what we can get.

How do you respond?

Stephen

