Misunderstanding and misrepresentation: a reply to Hutchison and Schagen

Stephen Gorard
The School of Education
The University of Birmingham
s.gorard@bham.ac.uk

Introduction

In this journal, I previously published a paper discussing how to conduct an analysis based on a cluster sample. In that paper, I outlined several widely-adopted alternative approaches, and pointed out that such approaches are anyway not needed for population figures, and not possible for non-probability samples (Gorard 2007). Thus, I queried the prevalence and relevance of complex approaches such as multi-level modelling (MLM) for most real-life analyses. I illustrated the lack of substantive difference in the results obtained via MLM and other approaches in a number of recent studies. And I repeated my challenge (from Gorard 2003a) for those advocating the use of MLM to find and publish a real study in which the policy or practice implications of the research would differ depending on whether MLM or a sensible alternative was used.

Instead of taking up this challenge, Hutchison and Schagen (2008) responded with a paper that misrepresents (and even misquotes) what I said, in an attempt to defend their favoured technique of MLM. In my reply to them here, I hope to establish their misunderstanding of this important topic. This might lead readers to imagine that this will be a highly technical paper about a statistical issue. It is not. I also hope to establish their misrepresentation of my paper. The issue is one of wider and greater concern than the surface argument about whether multi-level modelling is of any great value. It is about the kind of intellectual processes we are prepared to endorse as a field.[i]

This paper deals first with areas in which we agree, and then areas in which Hutchison and Schagen (2008) simply restate their position, before turning to the more substantive consideration of why it is not necessary or appropriate to use multi-level modelling in most real analytical situations.

Where we agree

The first few pages of Hutchison and Schagen (2008), after the summary, are largely uncontentious – to me at least since I have written all of it before. Hutchison and Schagen should know from my substantive work that much of what I do involves stressing and illustrating the differences between models and reality (see Gorard 2008a, 2008b for current examples). In the past I have had to do this in conflict with authors like Hutchison and Schagen who believe that only multi-level modelling, for example, can reveal the 'truth' of a matter (see below). They should also know this from attendance at my BERA sessions on the role of judgement in statistical analysis (which they disagreed with at the time, although they now repeat it in substance), and

from repeated and welcome attendance at presentations as part of the ESRC RDI project that I recently led (www.trials-pp.co.uk). I have used the wind tunnel analogy, that they now use, previously in many books and papers, including Gorard with Taylor (2004, see chapter seven). I have many times stressed the importance of judgement, and that analysis of any kind is an art, not a direct representation of reality (e.g. Gorard 2006a). I have often referred to methods as a toolkit, and through my own research I regularly use a much wider range of tools from that kit than either Schagen or Hutchison. Even the quotation from George Box with which they preface their paper has been used by me several times previously. See, for example, Sloane and Gorard (2003) where the quotation appears as the frontispiece of my paper, just as it does five years later in the paper by Hutchison and Schagen (2008). It is hard to see, knowing as I do that they have both heard me discuss these ideas for years, how they might believe that the first part of their paper, which simply echoes my own approach, could be a criticism of my purportedly 'misleading or misconceived' reasoning (p.11).

Re-statement as argumentative technique

For much of the rest of their paper, it becomes clear that, despite their apparent newfound agreement with my prior writings, Hutchison and Schagen (2008) *do* intend their preamble to set the scene somehow for contradicting my 2007 paper. But they do not explain how or why. To a large extent, Hutchison and Schagen (2008) contradict Gorard (2007) without offering any evidence, argument or even sources to back up their assertions. They simply state that I am wrong. In doing so they make the mistake of assuming as fact that which they are trying to establish in the first place, so replacing logic with dogmatism. If we all did this it would be the end of intellectual endeavour. Recall that my paper was questioning the value of widespread use of multilevel modelling, and that their paper was intended to show that 'there is much in his [i.e. my] line of reasoning which is misleading or misconceived' (Hutchison and Schagen 2008, p.11). Their paper is littered with brief quotations from my paper followed by brief contradictions such as 'this is wholly inaccurate' and 'this is incorrect' (both p.17) without any attempt to justify their claim. In their minds, readers should, apparently, just take their word for it.

Sometimes this invalid form of argumentation appears in slightly more disguised form. For example Hutchison and Schagen (2008, p.16) write in response to my request for a comparison of the value of multilevel modelling with other techniques:

> But the Hutchison article [Hutchison 2003] relates to the influences of one factor at one level on another at a different level, a situation for which multilevel modelling is specifically designed. What would be the point of using an inferior technique?

Here they are merely assuming from the outset that which they, presumably, seek to establish - namely that other techniques *are* 'inferior' to MLM. The paper is awash with this approach. See also, on the topic of class sizes (Hutchison and Schagen 2008, p.15, italics added):

> Only with a more sophisticated, value-added model do we start to be able to clear away the distracting factors and begin to see the *true* effect of reduced class sizes.

There are at least four problems here. First, they are simply restating their position. We must do it their way to get the 'true' answer, is their stance. Second, their claim does not relate specifically to MLM. Value-added models do not require MLM (see the official DCSF model, in Gorard 2006b). Third, value-added (with or without MLM) is not the 'only' way to sort out the apparently perplexing finding that pupils in larger classes have better results. It could be done in a host of ways including contextually (Gorard 2000) or via a randomised controlled trial (Cook and Gorard 2007). Fourth, and most hypocritically, Hutchison and Schagen (2008) use the term 'true' to describe the findings as they advocate them. But elsewhere (p.17) they lambast me for using the term 'incorrect' since, according to them, that 'illustrates a fundamental misunderstanding of the purpose of modelling and statistical analysis'.

A further example of their circular argumentative approach says:

> Only with the appropriate analysis tools and models do we have a chance of making progress – but as we start to do analysis we begin to find that simple models do not do justice to complex data. And recognition of the multilevel structure helps us to organize our analysis. [Hutchison and Schagen 2008, p.15]

Again this passage merely restates their initial position. They offer no justification for these substantial and unlikely claims – either here or elsewhere in the paper. How could they know that we 'only' have any chance to progress using what they consider to be appropriate tools and models? Was there no progress before 1986 and the advent of hierarchical linear modelling? Was there no progress after 1986 except via MLM? The assumption that complex data requires complex forms of analysis may be seductive to some readers but has no basis in either logic or evidence. Helpful insights might come equally from simple consideration of complex situations and not *only* from complex considerations. Of course it is true that looking at clustered datasets in terms of multiple levels can be helpful. I said as much in the paper they are purportedly arguing with. But it is not necessary to use the specific technique of multilevel modelling to recognise that structure. And it is inappropriate for Hutchison and Schagen to suggest here that it is, because Gorard (2007) outlines and references a variety of widely used and accepted alternatives. More importantly, there may be important clusters within any complex dataset that cannot be handled by multilevel modelling (which is merely hierarchical in nature). Sex, social class, and ethnicity, for example, are important theoretical clusters within many social science datasets. But they are not handled as clusters by multilevel modelling. So, in my opinion, not using multilevel modelling can be liberating and more insightful for an analyst because no analytical preference is given to hierarchical clusters like schools and teaching units, over equally relevant but non-hierarchical ones like sex and ethnicity.[ii]

Evading the comparison with multi-level modelling

In my paper I put forward arguments for and then against the use of multi-level modelling. I ended it by asking for clear evidence of where MLM helps us, because

'if MLM has the benefits described at the start of this paper then analysts should be able to devise a way of illustrating them' (Gorard 2007, p.234). In their paper, Hutchison and Schagen (2008, p.16, italics added) say:

> Gorard argues (230) that he [Hutchison] should have used both OLS and multilevel modelling, and compared the results, as *apparently* suggested in an article published at or about the same time (Gorard 2003a).

This is an interesting passage because of what it reveals about their cavalier use of evidence. I have many times asked that both MLM and simpler analyses are published together for a period so that we can judge empirically what we gain and lose through the use of each. In the paper cited by Hutchison and Schagen one passage was as follows:

> What Plewis and Fielding have failed to do in promoting what they term the '*potential*' of MLM is to show it leading to any practical discovery in education that could not have been obtained otherwise. This is a key empirical point (and note that this is very different from simply showing that MLM has been used in many useful pieces of work). Perhaps we should ask journals to print both the MLM and alternate analyses in all relevant papers for some time so that we can gain a clearer picture of the practical difference it makes? [Gorard 2003a, pp.420-421]

Where then does the word 'apparently' come from in the quotation from Hutchison and Schagen (2008, p.16)? They have read my 2003 paper and so must know that the suggestion is there.

Hutchison and Schagen (2008, p.8) also say:

> On page 230, he [Gorard] criticizes Schagen and Schagen (2005) for not finding anything new using multilevel modelling…. The fact that careful analysis of a more comprehensive dataset gave similar results to those obtained with a previous dataset should not be a cause for criticism.

I will quote my original passage at length to be clear here. What I said was:

> In a study of the impact of school diversity, Schagen and Schagen (2005) state that their paper 'acts as a case study of a piece of research which illustrates the *power* of combining multilevel modelling techniques with the rich linked national data sets now becoming available' (p.309). In the paper, the authors explain the great advantages of using MLM rather than simpler forms of regression, and also explain that the quality of datasets available to education research of this type is improving year-on-year. Thus, the reader is led to expect that their 'illustration' of the 'power' of MLM will lead to new and even unexpected findings. In fact, nothing of the sort happens. The authors conclude that 'the research described above replicated research undertaken earlier, but was based on a different dataset (2001 GCSE outcomes linked with 1996 KS2 levels) and used a wider range of outcomes and an improved methodology. The findings were broadly the same…'. [Gorard 2007, p.230]

Nowhere am I critical that their findings are similar to those previously. This similarity across studies is a very likely occurrence, and presumably outside the control of researchers anyway (if research evidence means anything at all). My point was simply that the paper is presented by them very forcefully as an illustration of powerful data allied with powerful techniques. Schagen and Schagen (2005, p.309) said, for example:

> The advent of large-scale matched data sets, linking pupils' attainment across key stages, give new opportunities to explore the effects of school organisational factors on pupil performance. Combined with currently available sophisticated and efficient software for multilevel analysis, it offers educational researchers the chance to develop objective evidence about issues both old and new.

and

> it acts as a case study of a piece of research which illustrates the power of combining multilevel modelling techniques with the rich linked national data sets now available. [p.309]

So what is the impact of this alliance of better data with sophisticated contemporary software for a refined multi-level analysis? The authors themselves say:

> the findings from the latest analysis were broadly consistent with those from the earlier research. [Schagen and Schagen 2005, p.325]

I raised this example in the section of my 2007 paper where I ask again what it is that multi-level modelling (MLM) has found out for us that we could not have found out by other, and simpler, means. None of the examples I found and quoted had found anything substantially different using multi-level modelling. Hutchison and Schagen (2008) try to evade the challenge of showing that MLM is of any practical value by claiming that I was being critical of their replicated findings. The issue for those who stand to gain professionally and financially from the use of MLM is quite simple. All they have to do is point to examples where the use of MLM has led to a clear analytic gain. The paper by Schagen and Schagen (2005) is, patently and in their own words, *not* such an example. And an artificial example will not do. Even an example of a different result gained from using MLM is not sufficient, by itself. They have to show that it is substantially (pragmatically) different *and* better than results obtained any other way. I have never seen such an example. If there were one, Hutchison and Schagen (2008) could have ended the matter by presenting and explaining it in their paper. I assume, therefore, that they do not have one, and I recommend that readers assume the same, until Hutchison and Schagen produce one.

In fact, Schagen must be rather confused here anyway since he has already published his agreement that using MLM is not always necessary. To remind readers, in my 2007 paper I quoted the following passage from his prior paper which says:

> In principle, this model should be fitted using multilevel modelling techniques, to allow for the clustering of pupils within school and to estimate more accurate standard errors. However, for this paper we have used ordinary least squares (OLS) regression – this is likely to give very similar results for the modelling coefficients and the predicted pupil values. [Schagen 2006, p.126]

Yet in addition to the quotation above about *only* being able to make progress with MLM, Hutchison and Schagen (2008, p.13) say:

> The choice of OLS or MLM is one of many that a modeller has to make… If there is any chance of non-zero variance at higher levels either in overall level of outcome or in slopes, then it makes sense to run a multilevel analysis – if these variances are in fact zero, the model will indicate this and the results will be essentially identical to those from OLS, but no harm will have been done. On the other hand, if the variances are non-zero, then different results are likely to be produced, which in general are likely to be more realistic.

So, Schagen's advice to you and I is that we should run MLM because it may be important if there is *any chance* of non-zero variance at higher levels. But in practice, as he admits in his 2006 paper, he himself does not do this. He even admits that for all practical purposes MLM and OLS will give pretty much the same results. Which is precisely the main point I was making in my 2007 paper. Why bother with the anti-democratic complexity of MLM when a simpler analysis will lead to the same practical findings?

Hutchison and Schagen ignore most of my suggestions for legitimate alternatives to multi-level modelling which is, after all, only one of the widespread ways for dealing with a cluster randomised sample, and they simply deny the rest. These alternatives come from a variety of informed and highly-regarded sources – such as my colleague Martin Bland, Professor of Health Statistics at the University of York, writing in the BMJ (cited in my 2007 paper), and the BMA Consort Guidelines for the analysis of data from cluster randomised samples. Of course, these and many other sources like them that allow both MLM *and* alternatives might be wrong. But if Hutchison and Schagen dispute these authorities then they need to make clear to their readers that this is what they are doing, and then explain why.

Instead what they do is repeat the bold part of the following section of my paper:

> There is no necessity to use MLM in any analytic situation. Perhaps the simplest way to respect the clustering in data is to analyse only at the cluster level. In fact, according to established statistical procedures this is what should be done anyway. **If pupils are clustered within schools, for example, and schools are selected to take part in a survey, and the survey instrument is then completed by a number of pupils in each of these schools, then the sample is of schools (not pupils) and can quite properly be analysed as a sample of schools**. Better, where the analyst has access to both individual and grouped data, is to analyse the data at each level separately (see, for example, Gorard et al. 2003). This is a very effective and perfectly safe procedure. Perhaps the simplest ways to avoid the supposedly terrible fate awaiting analysts who under-estimate their standard errors or confidence intervals through not using MLM include increasing the sample size (especially the number of clusters), using a proper random sample instead, using population figures instead, multiplying the standard errors by a weighting to reflect the estimated scale of the bias (or design effect), or – most simply – increasing the threshold for significance from the usual 5%. [Gorard 2007, p.232-233]

They then say (of the bold part)

> This is wholly inaccurate. It is possible to envisage some situations in which this might be true, but in general this would be a sample of pupils. [Hutchison and Schagen 2008, p.17]

It is hard to imagine how they came to write this, and how it was allowed to be published. First and most obviously, their statement contains an internal contradiction. What I said cannot be both 'wholly inaccurate' and 'might be true'. Second, of course, if it might be true (even if only in some situations) then such an approach 'can' be used – and that is all I said (it '*can* quite properly be analysed as a sample of schools'). Third, if pupils within schools had been selected randomly then a significance-based statistical analysis of pupils could, indeed, be undertaken. But this is not the situation. Where schools are selected as the cases (as stated clearly in the quotation from my paper above), and measurements are taken from a number of pupils per school to generate estimates of school-level values then the sample (and any 'uncertainty' as Hutchison and Schagen term it) can only be at the school level.

As an analogy, consider a random sample of individual pupils selected to measure their heights. If we take several estimates of each pupil's height using different personnel and different measuring devices at different times of day we may then have a better averaged estimate of the underlying hypothetical true heights for each individual. But our sample is still of pupils, and any significance-based analysis need only be done at that level. We have, it is true, a kind of sample of estimated heights for each pupil as well, but we have no reason to believe that variation between personnel, devices or contexts is random rather than biased in nature. One researcher might always measure to the top of any threshold line on a ruler, and another to the bottom or middle. One device might intrinsically tend to measure shorter than another. Pupils might tend to shrink slightly during the day, and so on. These are *not* random differences. Therefore, we need not and should not use a significance-based analysis at the within-pupil estimate level. Instead, we can quite properly, and as outlined in my original paper, analyse data at both levels using appropriate techniques for each. To treat the estimates as random would be unscientific in throwing away what we might know about them (that the shortest estimate for each pupil was regularly from the same instrument, or whatever). Multi-level modelling is neither relevant nor useful for this situation.

Hutchison and Schagen (2008, p.13) claim that two simple regression models at both individual and school level 'would be much more complex' than the equivalent multi-level model. In my original paper I directed readers to areas of work (mine and others) where both approaches were used. In my opinion, an MLM analysis and explanation is considerably more complex than the two ordinary regressions. I think that others, including academics like Carol FitzGibbon and Lindsay Patterson (both cited in my 2007 paper) and all research users and practitioners I have ever consulted, agree with me here. If Hutchison and Schagen really believe that MLM is simpler to do and explain then they could have illustrated this in their paper, or referenced somewhere that they have seen it done. As with their un-tested and un-illustrated claim that MLM gives better results, they do not provide either an example or a reference. What should we conclude from this absence?

A reprise of basic statistical ideas

In Gorard (2007) I question those, like Hutchison (2003), who quite properly use population data but then look for significance, standard errors, confidence intervals and the like as though trying to generalise from a random sample to their population. Schagen and Schagen (2005, p.316) make the same kind of mistake. Although this error of imagining that sampling theory and its creations (like significance) can be applied to non-random non-samples is widespread, it is still wrong. Perhaps the easiest way to see this error is to imagine, when Hutchison (2003, p.38) talks about 'significant', what he could actually mean by this in the context of a full cohort of data from one LEA. He would not, presumably, imagine that he could generalise statistically from this one LEA to all others – given that he has only one case and it was selected non-randomly anyway. He clearly could not want to generalise to other pupils of the same age within the LEA, since there are no others in the LEA to generalise to. Statistical analysis does not allow us to generalise from actual cases to imaginary cases (such as future cohorts who might be very different). So, statistical significance is irrelevant to the kind of analysis conducted by Hutchison (2003).

The reaction of Hutchison and Schagen (2008) to my reasoned position – that significance has a place in the toolbox but not for use with populations – is:

> Clearly the practice of significance testing has been grossly misused…. However, the strong reaction to this which would sweep all use away is excessive. [p.14]

They seem unable to distinguish between some and none here (something of a key problem for a numeric analyst). Saying, as I and many other commentators do, that significance tests are of no value with population data is not to 'sweep all use away' but to retain their widespread use with random samples, whether for modelling or not.

A good random sample is an excellent heuristic for a population, and the creations of sampling theory such as significance are there to help ensure that we are not misled by this resource-efficient process of sampling (by what Hutchison and Schagen refer to as 'uncertainty' which, of course, does not apply to population data). Therefore, it is obviously possible to model, statistically or otherwise, with population data, and using a population is analytically preferable to having a random sample. Anything that can be done with a model from a sample can also be done, better, with the population that the sample would otherwise seek to generalise to. Hutchison and Schagen (2008), among others, worry about how we would analyse data from populations if not using derivatives from sampling theory. This misses the point that we use sampling theory only because we have a sample. If we have a population we do not need, and could make no sense of, things like confidence intervals. We can use judgement to retain or eliminate variables from our model or decide which models to pursue in exactly the same way for populations as we must do when we have a sample (once we have disposed of the additional complication with a sample of assessing its generality). We could use a variety of approaches to help our judgement in selecting variables for models, such as 'effect' size (the amount of variance

explained) or predictions from theory, none of which depend on statistical significance.

Hutchison and Schagen (2008, p.15) use the recent emergence of the PLASC/NPD dataset to try and back up their point that:

> In fact, more and better data *needs* more sophisticated modelling approaches in order to interpret and understand it.

But nowhere do they explain how this follows. I regularly use the PLASC/NPD dataset. Sometimes I use complex models and sometimes simple arithmetic (Gorard et al. 2008). The latter is just as effective in its place. Because PLASC/NPD is for an entire population its analysis is simpler than other datasets in many ways because there is no issue of significance, confidence intervals or standard errors to deal with before looking for potential patterns and trends. If we find the mean GCSE points scores in maths for boys and girls, we do not need and cannot use the panoply of traditional statistics to help us decide whether one mean is larger than another. It either is or it is not. In that year, for that assessment, as recorded by these data, the scores are what they are. Making the analysis multivariate to include area, age, FSM eligibility, other subjects and so on does not change this basic simplicity.

Of course, we should not just assume that any difference, pattern or trend we see in PLASC/NPD data is of any scientific value or practical consequence. Like all such datasets, PLASC/NPD has limitations (see Gorard 2008c). For example, in the 2006 PLASC there were 3,371 cases with no ethnicity recorded from a total of 601,230 pupils at Key Stage 4. This does not sound like a large proportion until we consider the NPD which has 664,630 pupils at Key Stage 4. This means that a minimum of 66,771 cases (more than 10% of the total) appear to have no recorded ethnicity, and we have to add to this total those cases unmatched across PLASC and NPD. What all of these inevitable limitations of PLASC/NPD (and all other such datasets) have in common is that they are not random events. To propose, as Hutchison and Schagen do, that we can allocate them to a bucket of all unknown random factors is far-fetched and based on a false premise. They say of multi-level models (p.14, italics added):

> The random part of the model is not necessarily so because of random sampling. But just because we have constructed it to include *all those relationships and influences which we do not otherwise understand*.

This is their argument for using statistical testing on PLASC/NPD data, and part of their 'reasoning' for the need to use multi-level modelling. But I feel in contrast that we must use what we *do* know about missing data (variables, cases, and values) to help judge the importance of any pattern we find. One of the things that we, but not Hutchison and Schagen apparently, know from decades of social research is that we can have no reason to assume that missing data is random. Where this has been possible to test, data is more likely be missing from stratified subsets of society – those who cannot read and complete official forms, those without a fixed address, or those who just miss out on benefit thresholds, for example (see Gorard 2003b). None of this is amenable to fixing via significance tests, or indeed by MLM. I am afraid Hutchison and Schagen are just plain wrong.

References

Cook, T. and Gorard, S. (2007) What counts and what should count as evidence, pp.33-49 in OECD (Eds.) *Evidence in education: Linking research and policy*, Paris: OECD

Gorard, S. (2000) 'Underachievement' is still an ugly word: reconsidering the relative effectiveness of schools in England and Wales, *Journal of Education Policy*, 15, 5, 559-573

Gorard, S. (2003a) In defence of a middle way: a reply to Plewis and Fielding, *British Journal of Educational Studies*, 51, 4, 420-426

Gorard, S. (2003b) *Quantitative methods in social science: the role of numbers made easy*, London: Continuum

Gorard, S. (2006a) Towards a judgement-based statistical analysis, *British Journal of Sociology of Education*, 27, 1, 67-80

Gorard, S. (2006b) Value-added is of little value, *Journal of Educational Policy*, 21, 2, 233-241

Gorard, S. (2007) The dubious benefits of multi-level modelling, *International Journal of Research and Method in Education*, 30, 2, 221-236

Gorard, S. (2008a) Research impact is not always a good thing: a re-consideration of rates of 'social mobility' in Britain, *British Journal of Sociology of Education*, 29, 3, 317-324

Gorard, S. (2008b) The value-added of primary schools: what is it really measuring?, *Educational Review*, 60, 2, 179-185

Gorard, S. (2008c) Which students are missing from HE?, *Cambridge Journal of Education*, 38, 3

Gorard, S. and Fitz, J. (2006) What counts as evidence in the school choice debate?, *British Educational Research Journal*, 32, 6, 797-816

Gorard, S., See, BH. And Smith, E. (2008) *The impact of SES on participation and attainment in science – a review of available data*, London: Royal Society

Gorard, S., Taylor, C. and Fitz, J. (2003) *Schools, Markets and Choice Policies*, London: RoutledgeFalmer

Gorard, S., with Taylor, C. (2004) *Combining methods in educational and social research*, London: Open University Press

Hutchison, D. (2003) The effect of group-level influences in pupils' progress in reading, *British Educational Research Journal*, 29, 1, 25-40

Hutchison, D. and Schagen, I. (2008) Concorde and discord: the art of multilevel modelling, *International Journal of Research and Method in Education*, 31, 1, 11-18

Noden, P. and Goldstein, H. (2007) A brief response to Gorard and Fitz, *British Educational Research Journal*, 33, 2, 273-274

Plewis, I. and Fielding. A. (2003) What is multi-level modelling for? A critical response to Gorard (2003) , *The British Journal of Educational Studies*, 53, 4, 408-19

Schagen, I. (2006) The use of standardized residuals to derive value-added measures of school performance, *Educational Studies*, 32, 2, 119-132

Schagen, I. and Schagen. S. (2005) Combining multi-level analysis with national value-added data sets – a cases study to explore the effects of school diversity, *British Educational Research Journal*, 31, 3, 309-328

Sloane, F. and Gorard, S. (2003) Exploring methodological aspects of design experiments, *Educational Researcher*, 32, 1, 29-31 and 35-37

Taylor, C. and Gorard, S. (2001) The role of residence in school segregation: placing the impact of parental choice in perspective, *Environment and Planning A*, 30, 10, 1829-1852

---

[i] Whenever I have written previously about multi-level modelling, its advocates have resorted to amending what I wrote, quoting the incorrect text and then arguing with it in a way that relies for any efficacy on readers not caring enough about the truth of the matter to read the full accounts. Presumably this is easier for MLM advocates than actually discussing what I did write. I have described this process of distorting evidence, by writers such as Plewis and Fielding (2003) and Noden and Goldstein (2007), as 'the antithesis of intellectual endeavour' (Gorard and Fitz 2006). Noden and Goldstein (2007), for example, attempted to deflect criticism of their faulty arithmetic in a previous paper where they calculated a national average by adding together scores for each LEA without regard to the size of each LEA. They did so by claiming that I had made the same mistake in another paper (Taylor and Gorard 2001). I invite all interested readers to check the facts out for themselves. They will see that nowhere in the paper do I find a national average for anything. I do use unequal size enumeration districts and unequal size schools as organisational units but their size is automatically taken into account when aggregating to LEA-level. Therefore I did not, indeed could not, make the same serious arithmetic error as Noden and Goldstein.

Hutchison and Schagen (2008) appear to approve of such misquotations and distortions by citing a paper by Plewis and Fielding (2003) who used this dishonest and anti-intellectual technique of mis-quoting widely, despite the fact that Hutchison and Schagen also cite my response to Plewis and Fielding (Gorard 2003a), and so must have read about these inaccuracies. In case any readers are in any doubt about how blatant these examples are, please follow the references in Gorard and Fitz (2006) and read the original papers. I have to assume that Hutchison and Schagen either have not done so, or cannot see what Plewis and Fielding (2003) did wrong. Perhaps this is because, like them, Plewis and Fielding were seeking to defend the widespread use of MLM. If such apparent leaders in their field have to resort to misrepresentation to defend their position, and if peer-reviewers and editors cannot spot it when they do, then education research faces a poor future. This is partly why I say that the issue underlying this paper is important, and about more than whether MLM produces better results than OLS.

[ii] On the topic of clustering, Hutchison and Schagen (2008, p.16) say:

> He [Gorard] also states that: 'a girl in one class may be much more similar to a girl in another class than to a boy in another class. And so on'. Again this is a misunderstanding. As an analogy of why this is incorrect, for example it is known that boys are on average taller after puberty than girls. This statement of averages does not imply that all boys are taller than girls.

There are many problems with this passage. First, the quotation is wrong (see footnote i). What I actually wrote was 'A girl in one class may be more similar to a girl in another class than to a boy in *her own* class. And so on.' (Gorard 2007, pp.227-228). Second, presumably both versions – what I said and what they misquote me as saying – are actually perfectly reasonable. A girl might, in many respects, be more similar to another girl of the same age in a different teaching group than to a boy whether in the same or a different teaching group. How can Hutchison and Schagen call this 'incorrect'? Third, what is their 'analogy' intended to show? It reminds me that boys might tend to be taller than girls of the same age. This would increase the probability that, in terms of height at least, two girls might be more similar than a girl and boy - whatever teaching groups they were in, since height is not usually a criterion for allocating pupils to classrooms. This would seem to me to agree both with what I actually said and with what they thought I said. I cannot imagine how they think this any kind of argument for the use of MLM.