

Comments on ‘Modelling social segregation’ by Goldstein, H. and Noden, P. (2003) in Oxford Review of Education, 29, 2, 225-237

Stephen Gorard
Cardiff University School of Social Sciences
Glamorgan Building
King Edward VII Avenue
CF10 3WT, UK
email: gorard@cardiff.ac.uk

Introduction

This brief paper is a commentary on a piece of work recently published in the Oxford Review of Education (Goldstein and Noden 2002) which purports to create a multi-level model of the social segregation between schools in England 1994-1999. Segregation, as a measure of the (un)evenness of the distribution of disadvantaged students between schools, is an important characteristic of the school system, related to social cohesion and school effects (Gorard, Taylor and Fitz 2003). Probably the first thing to note about the ‘new’ approach by G&N to measuring such segregation is that, if we accept the findings at face value then, it produces results that have already been published (e.g. Gorard and Fitz 2000). Their substantive findings are that segregation between schools in England rose slightly 1994 to 1999, and that areas with academic selection or a high proportion of schools with control over their own admissions tend to have higher and rising levels of social segregation. As is so often the case in practice (Gorard 2003), the use of multi-level modelling (MLM) in this instance tells us nothing new. MLM here appears to confirm our results obtained using the far simpler approach of analysing segregation at varying levels of aggregation. In fact, this appearance is deceptive because whatever it is that G&N believed they were doing they were not, in fact, measuring segregation at all.

What is the appropriate level of analysis?

Segregation between schools is a property of groups. This very simple fact means that it does not make sense to try and measure the extent to which an individual is, or is not, evenly distributed between administrative units such as schools or LEAs. We can, of course, attempt to measure the extent to which an individual is exposed to other individuals of the same type, and G&N have tried this before (e.g. Noden 2000). But such measures of exposure are very different, have some fundamental problems (Gorard and Taylor 2002), and were anyway misapplied by G&N. Noden (2000) used invalid arithmetic, when calculating a ‘national’ score for segregation by averaging the local scores for segregation irrespective of the size of each LEA, for example. This is a terrifying mistake by someone attempting the far more complex technique of MLM. They have now retracted this analysis (Noden 2002).

Segregation *between* schools also cannot be measured *within* schools. We cannot say by looking at any school in isolation whether it has, or has not, admitted its fair share of disadvantaged students. The lowest level of any MLM therefore cannot be the

school any more than it can be the individual. Segregation between schools is a comparative measure, meaningless at the school level. We could, as we have previously done, consider the social composition of a school in relation to its local residential population (Taylor and Gorard 2001). Or we could, as we have done with our segregation ratio, consider the social composition of a school in relation to its local school population (Gorard 2000). Or we could, as we prefer to do with our segregation index, consider the social composition of all schools in any analysis at the same time. In all of these approaches the school figures are, and must be, examined in comparison to something else. Of course, there is also a consideration of the most appropriate level of aggregation at which to conduct any analysis, but we have shown that this makes no difference in practice (Taylor, Gorard and Fitz. 2003).

Anyone attempting to create a multi-level model of segregation therefore faces several practical problems, which is why we rejected this approach. We sought, and used, what appeared to be the most appropriate tools for the job, rather than deciding in advance, as G&N clearly did, to use MLM and then try to make the empirical situation fit the demands of the model. If individual and school level data cannot tell us about segregation between schools then the lowest level of any MLM must be a group of schools. How, then, did G&N manage to create a two-level model based on schools within LEAs?

The answer is that they did not, for they did not model social segregation between schools at all. Their model uses as its dependent variable the proportion of students eligible for free school meals (FSM) in each school. Or as they put it, 'The response is, in this case, the proportion of disadvantaged pupils in a school measured as the proportion eligible for free school meals or, put another way, the probability that a randomly chosen pupil in school i in area j is eligible' (p.227-228). So their model is attempting to predict/explain the level of FSM, a commonly used indicator of social disadvantage, using a number of predictor variables, such as the type and number of schools in each area. They go on to explain that 'The term v_j denotes the residual (not explained by the predictor variables) at the area level and u_{ij} the school level residual. The area and school variances of these residuals σ_v^2 and σ_u^2 are measures of segregation... they measure segregation after adjusting for the fixed effects (X)' (p.228). But this last part is false, at least according to any standard definition of segregation.

According to the usage of G&N here, segregation is the variance in the residuals (i.e. in the extent of the failure to predict/explain) when explaining the proportion of FSM in each school after their predictors have been taken into account. This is ludicrous. What G&N are measuring over time, then, is the change in their ability to model the proportion of FSM accurately. Their Table I shows that it gets marginally harder to predict the proportion of FSM students in each school from 1994 to 1999. What does this mean? Does that mean that schools are more segregated by FSM, as the authors claim (but do not argue at all)? It seems to mean that the fixed effect predictors such as the type of school become less effective as predictors over time. This could be a good thing – by moving away from a situation where academic selection entails sorting by FSM.

To make the point clearer, consider the following perfectly plausible example. Larger LEAs may tend to have less segregated schools than small LEAs. If we use, as G&N

did, the number of schools in each LEA as a predictor variable, then our model may be able to predict the proportion of FSM students per school quite well. In this case the residual, and therefore its variation, will be low, but this bears no relation to the actual distribution of FSM students between the schools in each area, or between areas. In an extreme case, imagine a school system totally segregated so that all FSM students were evenly distributed in one type of school, and no FSM students were in any other type of school. The G&N model would be able to predict/explain the FSM level in each school perfectly, and the residual would be small or zero. They would conclude, against common sense, that this totally segregated school system had no social segregation. Here is another instance of what is also commonly observed with MLM, an attention to the mathematical detail of the model apparently at the expense of fuller consideration of what it is modelling (i.e. a case of good maths but poor science).

Other problems in modelling segregation

Superfluity of technical information

It has been observed before that researchers using MLM have a tendency towards providing a superfluity of information about their models (Gorard 2003). They can provide considerable details of the mathematics, and often tell us their computer-generated variable names, but then omit somewhat more crucial information about the science. Here is another example. G&N report that ‘In the first line of the model shown above, the predictor variables, denoted by X, *can* include time, school type, number of schools in an area etc.’... ‘Note that the size of school does not explicitly enter into this model (unless we *choose* to include it as a predictor)’ (p.228, emphasis added). Nowhere in the paper are we told whether or not they did choose to include size of school or indeed any of the listed predictor variables. The only one mentioned again is the selective LEA flag, used in a later model. And the use of time as a predictor is implied throughout. The paper is otherwise silent on this most important issue – which predictors were actually used? – perhaps a more important consideration for most readers than much of information actually given in the lengthy technical appendix.

Misappropriating ‘significance’

Authorities on segregation are agreed that its measurement ‘relies on having data on the full population under examination, since little or nothing is known about the sampling properties of segregation measures’ (Reardon and Firebaugh 2002, p.100). Yet G&N talk about a ‘significant change in segregation’ (p.227), and quote probabilities based on sampling theory even though it is clear that their dataset consists of population figures (and this not the first time they have demonstrated this confusion, Goldstein 2001). Their Table II shows null hypothesis significance tests being performed on their population data (a χ^2 test in this instance). They quote probabilities, such as ‘P<0.10’ (p.234), and use these to argue that there has been a *significant* change in ‘segregation’ from 1994 to 1999. Probabilities such as these are estimates of the likelihood that the observed differences between scores could have been generated solely by the random nature of the sample on which the tests are predicated (Siegel 1956). If there is no sampling process, or the sample is not random,

then the probabilities are meaningless. They certainly do not mean that the differences between scores are worthy of note – which is what G&N seem to imagine. Of course, even if the figures had come from a random sample a ‘significant’ probability would still mean very little, since we are looking at figures for every school in England giving us a very large N (3,560), for which an effect size would be safer (Coe 2003).

Other confusions

It should be noted that the model presented by G&N uses FSM, a limited but readily available indicator of poverty, which they have previously criticised us for using (Noden 2002). They use it because, in the absence of individual pupil records for all schools 1994-1999, this is the best proxy, as they discovered when they actually came to do some analysis themselves.

G&N claim (p.226) that one of the key advantages of MLM is that it ‘allows us to have random coefficients in the model’. Yet their model defines only fixed effects. What then of this *key* advantage?

They claim (p.226) that the value of a traditional index of segregation ‘will also depend intrinsically on school size’, but this claim is demonstrably false. Since the index we used (S) is proportionate it is unaffected by changes in school sizes that do not affect the distribution of disadvantaged students (Gorard and Taylor 2002).

G&N claim that ‘crucially’ their approach does not involve ‘aggregating data to the level of the LEA in such a way that information is lost’ (p.226). But this is also true of all other approaches. We have, for example, looked at individual school trajectories, competition areas, and school districts as well as LEAs. No data has been lost in this way, even though we did not use MLM. In a review of the methods used in a book that he now admits to not having read in full, Goldstein (2001) termed the use of data at one level of aggregation with conclusions appropriately drawn at the same level an ‘ecological fallacy’. But an ecological fallacy involves the analysis of data at one level of aggregation, and the drawing of unwarranted conclusions at *another* level (a very different matter indeed, see Robinson 1950). As with the distortion of the term segregation (see above) this abuse of terms leads us into a Humpty-Dumpty world where intelligent debate and the cumulation of knowledge is impossible.

Conclusion

In conclusion, and for the reasons given, the approach proposed by Goldstein and Noden (2003) is retrograde, erroneous, and of no clear practical value. In so far as it confirms our own previously published results their re-analysis is welcome. However, to use multi-level modelling in this way to work with a concept such as segregation which has to be defined in a specific way in terms of social groups within organisational units seems to be a case of the method cart pulling the empirical horse.

References

Coe, R. (2003) What is an effect size?, *Building Research Capacity*, 4, p.6

- Goldstein, H. (2001) Review of 'Education and Social Justice' by Stephen Gorard, *British Journal of Educational Studies*, 49, 3, 354-357
- Goldstein, H. and Noden, P. (2003) Modelling social segregation, *Oxford Review of Education*, 29, 2, 225-237
- Gorard, S. (2000) *Education and Social Justice*, Cardiff: University of Wales Press
- Gorard, S. (2003) What is multi-level modelling for?, *British Journal of Educational Studies*, 51, 1, 46-63
- Gorard, S. and Fitz, J. (2000) Investigating the determinants of segregation between schools, *Research Papers in Education*, 15, 2, 115-132
- Gorard, S. and Taylor, C. (2002) What is segregation? A comparison of measures in terms of strong and weak compositional invariance, *Sociology*, 36, 4, 875-895
- Gorard, S., Taylor, C. and Fitz, J. (2003) *Schools, Markets and Choice Policies*, London: Routledge
- Noden, P. (2000) Rediscovering the impact of marketisation: dimensions of social segregation in England's secondary schools, *British Journal of Sociology of Education*, 21, 3, 371-390
- Noden, P. (2002) Education markets and social polarization, *Research Papers in Education*, 17, 4, 410-412
- Reardon, S. and Firebaugh, G. (2002) Response: segregation and social distance – a generalised approach to segregation measurement, *Sociological Methodology*, 32, 85-101
- Robinson, W. (1950) Ecological Correlations and the Behavior of Individuals, *American Sociological Review*, 15, 351-357
- Siegel, S. (1956) *Nonparametric Statistics*, Tokyo: McGraw-Hill
- 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
- Taylor, C., Gorard, S. and Fitz, J. (2003) The modifiable areal unit problem: Segregation between school and levels of analysis, *International Journal of Social Research Methodology*, 16, 1, 41-60