CLUSTERING EFFECTS IN TIMSS

Dougal Hutchison
National Foundation for Educational Research in England and Wales, UK

Abstract

Analyses of TIMSS data have, naturally enough, largely concentrated on comparisons of overall attainment between and within countries. There is however a huge range of data collected, around an educational framework, and there is substantial scope for secondary analyses on other themes. This presentation describes the result of one such investigation, on the degree of within-school clustering of variables, both outcomes and background factors.

The degree of similarity among members of a group, such as a school or classroom, is a very important factor in the design and analysis of studies in education. It is measured by two statistics, the intracluster correlation and the design effect. These help determine the total size and structure of the sample of respondents used. They also determine the extent to which it is possible to discover school or class effects on behaviours such as learning.

This paper, using data from TIMSS, classifies outcomes into a number of categories and estimates intracluster correlation and design effect. The relevance of the results to the design and analysis of international studies is discussed. We also discuss briefly what these findings tell us about a shared youth culture.

INTRODUCTION

Human beings are innately social animals, and a large proportion of human activities take place within the context of groups- families, employment, clubs, schools, classes, spectators at a rugby match and so on. Some of these groupings are long lasting and important, while others are relatively ephemeral and unimportant, but in general the members are in some relevant way more similar to each other than are the population taken as a whole. This is particularly so in education and education research. Education and information about education is highly structured: individuals are grouped into classes, which are grouped into schools, which are grouped into local authorities, which are grouped into countries. If there is a degree of similarity among members of a group, then there are methods, known variously as multilevel modelling or hierarchical linear modelling (Snijders & Bosker, 1999; Raudenbush & Bryk, 2002; Goldstein, 2003), which have been devised to take account of this similarity, and it is widely recommended [1] that these are used for analysing such data. Snijders and Bosker (1999) state that 'If the macro units have any meaningful relationship with the phenomenon under study, analysing only aggregated or only disaggregated data is apt to lead to misleading conclusions'.

Failing to take account of the multilevel structure means that results can be seriously biased, especially levels of statistical significance, and by taking aggregated effects as indicators of individual behaviour (Goldstein, 2003; Robinson, 1950).

This grouping affects not only the analysis of data, but also its collection. Where data is collected by sampling clusters of respondents, rather than by simple random sampling from the entire data set, then this means that there is less independent information to be obtained from the sample than there would be from a simple random sample: it is as if the data were collected from a smaller, and in some cases, considerably smaller, sample. It will obviously be important to those designing international studies to have some indication of these effects in designing surveys of schools. Ways of estimating this effect are discussed below. This is not simply an arid statistical exercise. Efficient design of samples enable the best collection of information from schools with minimum dislocation, and once the data is collected; interpretation of design effects enables the research to look at such topics as inequality and shared cultures.

To quantify the degree of clustering and its effect on design and analysis, it will be necessary to introduce a small amount of statistical terminology. I introduce two statistics, the intracluster correlation ρ, and the design effect *Deft.* These are explained in the next section.

Intra-class correlation

If we are looking at a measurement on a particular outcome, such as test score, attitude to school, etc, then we label it by Y. Schools would be numbered from 1 to *J*, and within each school pupils would be numbered from 1 to *I(j)*. To indicate the i^{th} person in the j^{th} group we use the term Y_{ij}.

The most basic model for the hierarchical situation is given by

$$Y_{ij} = \beta_0 + u_i + e_{ij},$$

β_0 is the overall mean and e_{ij} measures how far that person is from the mean performance in the group, while u_i measures how far the school is from the overall mean. The degree of clustering in the data is indexed by coefficient ρ (Kish, 1965), and

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$, for a two-level hierarchy. [2]

Some examples of the interpretation of particular values of this coefficient may make the meaning clearer. If $\rho = 1$, there is complete homogeneity within the clusters, and all items in a cluster are equal. $\rho = 0$ means random sorting of elements into clusters. $\rho = -\frac{1}{b-1}$ where *b* is the size of the cluster means extreme heterogeneity within clusters.

In this paper, we will consider a two-level hierarchy, pupils within schools. ρ is generally referred to as intra-class correlation. It gives a measure of the proportion of total variation in a quantity that is variation between the clusters. An example could be the proportion of variation in number of GCSEs that lies between schools.

Uses of ρ

While ρ is a relatively simple statistic, it is also one with a number of uses, in designing and analysing surveys and experiments.

First, it can be used in the design of surveys. In their review of design issues in multilevel studies, Reise & Duan (2002) emphasise that intracluster correlation is an important factor in deciding whether to take (for example) more schools and fewer pupils in each, or fewer schools with more pupils. The basic statistic ρ is used as the basis of a more applied statistic *Deft* or Design Effect (Kish, 1965), which is frequently used to take account of the structure of a sample. It is defined by

$$Deft = sqrt(1 + (b-1)\rho),$$

where b is the average number of units in the cluster. It represents the factor by which the simple random sampling standard error of a statistic such as a mean or total is multiplied to give the standard error of a clustered sample. Where both ρ and b are sizeable, then $Deft$ can be quite substantial. These relationships can be applied to the design of experiments as well, and to meta-analysis (Morebeeke, 2000; Raudenbush and Liu, 2000).

In addition to its use in survey design, estimating ρ can also be a useful first step in analysing data, and in deciding what kind of methods to use. If one wishes only to estimate a mean and a standard error, then estimating ρ can help to decide whether it is necessary to take account of the structure. In this case, if ρ is sufficiently small, there is little point in using anything more complex than using simple random sampling formulae. Even if ρ, and with it $Deft$, are sizeable, it may be simpler to estimate the simple random sampling formula and multiply it by $Deft$ estimated from previous studies, if there are a number of similar ones already carried out (Hox, 2002).

However, if the value of ρ is very small, this does not necessarily mean that one should not use multilevel methods. It could be that there is a degree of variation in the slopes of the outcome on covariates. If, however there is little sign of either, analysing data using multilevel methods is not necessary (Snijders & Bosker, 1999).

ρ can also be of interest in its own right, to compare how much of the variation in a characteristic, such as an attitude to classical music, is between schools, and how much within. Thus, if we found that relatively little of the variation in attitudes, interests and out-of-school activities was between schools we might consider that this supported the notion of a common youth culture. Conversely, if we found that ρ was large, and thus most of the variation was between schools, then this might suggest that each school had its own, relatively distinct, culture.

Size of ρ

Hox (2002) suggests .05, .10 and .15 respectively as indicating small, medium and large values of ρ in most situations, though where larger values are the norm, such as within families, he suggests 0.1, 0.2 and 0.3 respectively. The former set of recommended values would be relevant to the education situation. ‘Small’ should not necessarily be taken as equivalent to 'negligible': a value of ρ of 0.05 with an average cluster size of 30, a fairly usual cluster size in education research, gives a quite substantial value of $Deft$ of approximately 1.6

Statistical significance of ρ could be taken as another guide, though this can depend on the size of the sample. In large samples, if ρ is not statistically significant at the .05 level, then ρ can safely be discounted. If it is statistically significant on a sample of (say) 2500, then it could still be insignificant for practical purposes.

THE RESEARCH

The research described here looks at the values of ρ found in the TIMSS survey in England and Wales. Results are analyses for both primary (nine-year-olds) and secondary (thirteen-year-olds). The information collected covers a wide range of topics. Table 1 shows the categories used to group the variables. The variables themselves are shown in Table 2, sorted by category.

Table 1

Category	N of items
Demographic	21
Leisure activities	15
Attitudes to school	62
Education attainment	16
Educationally relevant home characteristics	43
Ethnic and linguistic groups	8

This grouping was carried out by inspection and in collaboration with colleagues with extensive experience in educational surveys. Confirmatory factor analyses of this grouping were carried out using the BIFACTOR facility in the program TESTFACT (Bock *et al.,* 2002).

Table 2: Intracluster correlation and design effects (Deft)

Type	Characteristic measured	N of Values	Survey	Rho	Chi-sq	Prob	Deft	N in Schl
Demographic	Student lives with mother	2	P	0.00	1.5	0.2	1.1	45
	Student lives with father	2	P	0.03	52.6	0.0	1.5	45
	Student lives with brother(s)	2	P	0.01	11.5	0.0	1.2	45
	Student lives with sister(s)	2	P	0.02	19.7	0.0	1.3	45
	Student lives with stepmother	2	P	0.01	3.2	0.1	1.1	45
	Student lives with stepfather	2	P	0.01	16.6	0.0	1.3	45
	Student lives with grandparent(s)	2	P	0.02	21.2	0.0	1.3	45
	Student lives with relative(s)	2	P	0.02	38.8	0.0	1.4	45
	Student lives with other(s)	2	P	0.01	6.0	0.0	1.2	45
	N of people living in home	8	P	0.06	174.3	0.0	1.9	45
	Gender	2	P	0.01	17.1	0.0	1.3	46
	Student lives with mother	2	S	0.00	0.0	1.0	1.0	28
	Student lives with father	2	S	0.01	7.4	0.0	1.2	28
	Student lives with brother(s)	2	S	0.01	2.6	0.1	1.1	28
	Student lives with sister(s)	2	S	0.01	2.4	0.1	1.1	28
	Student lives with stepmother	2	S	0.00	0.3	0.6	1.0	28
	Student lives with stepfather	2	S	0.01	2.0	0.2	1.1	28
	Student lives with grandparent(s)	2	S	0.01	7.6	0.0	1.2	28
	Student lives with relative(s)	2	S	0.02	14.2	0.0	1.2	28
	Student lives with other(s)	2	S	0.00	-0.2	1.0	1.0	28
	Gender	2	S	0.16	407.4	0.0	2.3	28
N				21			29	29
Mean				0.02			1.3	37.0
StdDev				0.03			0.3	8.7
Leisure activities	Outside school clubs participation	4	P	0,05	87,2	0,00	1,7	39
	Outside school watch TV or videos	4	P	0,02	42,5	0,00	1,4	42
	Outside school play computer games	4	P	0,03	67,3	0,00	1,5	41
	Outside school play with friends	4	P	0,01	15,5	0,00	1,2	40
	Outside school do jobs at home	4	P	0,02	29,2	0,00	1,3	40
	Outside school play sports	4	P	0,03	53,2	0,00	1,5	41
	Outside school read a book	4	P	0,02	23,9	0,00	1,3	41
	Outside school clubs participation	4	S	0,04	32,2	0,00	1,4	26
	Outside school watch TV or videos	4	S	0,07	108,1	0,00	1,7	27
	Outside school play computer games	4	S	0,07	113,2	0,00	1,7	27
	Outside school play with friends	4	S	0,11	232,5	0,00	2	27
	Outside school do jobs at home	4	S	0,04	47,8	0,00	1,4	27
	Outside school play sports	4	S	0,02	9,5	0,00	1,2	27
	Outside school read a book	4	S	0,04	39,0	0,00	1,4	27
	Outside school work at paid job	4	S	0,03	23,2	0,00	1,3	26
N				15			15	15
Mean				0,04			1,5	33,2
StdDev				0,03			0,2	7,2

Attitudes to school	Important to friends good in sports	4 P	0,01	5,5	0,02	1,1	Attitude school
	Important to friends have time for fun	4 P	0,01	12,7	0,00	1,2	
	Self important good in sports	4 P	0,01	18,7	0,00	1,3	
	Self important have time for fun	4 P	0,02	22,3	0,00	1,3	
	I think maths is boring	4 P	0,03	62,6	0,00	1,5	
	I like computers maths class	4 P	0,03	78,1	0,00	1,6	
	I think maths is an easy subject	4 P	0,03	73,1	0,00	1,6	
	I think enjoy learning maths	4 P	0,03	66,1	0,00	1,5	
	Important to friends do well in maths	4 P	0,03	63,3	0,00	1,5	
	I usually do well in maths	2 P	0,01	12,0	0,00	1,2	
	I like mathematics	4 P	0,04	97,7	0,00	1,6	
	Self important do well in maths	4 P	0,01	17,8	0,00	1,3	
	I think science is boring	4 P	0,02	38,2	0,00	1,4	
	I like computers science class	4 P	0,05	123,3	0,00	1,7	
	I think science is an easy subject	4 P	0,04	82,4	0,00	1,6	
	I think enjoy learning science	4 P	0,04	93,7	0,00	1,6	
	Important to friends do well in science	4 P	0,04	91,8	0,00	1,6	
	I usually do well in science	2 P	0,03	66,2	0,00	1,5	
	I like science	4 P	0,05	117,3	0,00	1,7	
	Self important do well in science	4 P	0,02	27,5	0,00	1,4	
	Important to friends to do well in language	4 S	0,03	25,2	0,00	1,3	
	Important to friends to have time for fun	4 S	0,01	2,8	0,09	1,1	
	Important to friends to good in sports	4 S	0,05	68,9	0,00	1,5	
	Important to friends to high achieve in class	4 S	0,05	52,9	0,00	1,5	
	Important to me to do well in language	4 S	0,01	6,1	0,01	1,2	
	Important to me to have time for fun	4 S	0,02	10,9	0,00	1,2	
	Important to me to good in sports	4 S	0,03	36,1	0,00	1,4	
	Important to me to high achieve class	4 S	0,07	114,9	0,00	1,7	
	I think maths is boring	4 S	0,04	37,1	0,00	1,4	
	I like computers in maths class	4 S	0,23	667,9	0,00	2,7	
	Do well in maths because natural talent	4 S	0,03	20,2	0,00	1,3	
	Do well in maths because good luck	4 S	0,06	76,8	0,00	1,6	
	Do well in maths because hard work study	4 S	0,03	35,7	0,00	1,4	
	Do well in maths because memorise notes	4 S	0,06	76,0	0,00	1,6	
	I think maths is an easy subject	4 S	0,02	13,8	0,00	1,2	
	I think enjoy learning maths	4 S	0,05	53,7	0,00	1,5	
	Important to friends to do well in maths	4 S	0,03	30,1	0,00	1,4	
	I do well in maths to get desired job	4 S	0,03	31,9	0,00	1,4	
	I think maths is important in life	4 S	0,01	6,3	0,01	1,2	
	I like mathematics	4 S	0,04	37,4	0,00	1,4	
	I do well in maths to please parents	4 S	0,03	28,7	0,00	1,3	
	I do well in maths to please myself	4 S	0,02	9,2	0,00	1,2	
	Important to me to do well in maths	4 S	0,02	19,1	0,00	1,3	
	I usually do well in mathematics	2 S	0,02	10,2	0,00	1,2	
	I think like job involving maths	4 S	0,03	20,1	0,00	1,3	
	I think science is boring	4 S	0,03	27,2	0,00	1,3	
	I like computers in science class	4 S	0,11	219,4	0,00	2	
	Do well in science because natural talent	4 S	0,02	17,6	0,00	1,3	
	Do well in science because good luck	4 S	0,05	59,0	0,00	1,5	
	Do well in science because hard work study	4 S	0,03	31,0	0,00	1,4	
	Do well in science because memorise note	4 S	0,04	51,1	0,00	1,5	
	I think science is an easy subject	4 S	0,04	48,6	0,00	1,5	
	I enjoy learning science	4 S	0,04	48,7	0,00	1,5	
	Important to friends to do well in science	4 S	0,03	21,2	0,00	1,3	
	I do well in science to get desired job	4 S	0,01	4,3	0,04	1,1	
	I think science is important in life	4 S	0	0,9	0,35	1,1	
	I like science	4 S	0,04	52,5	0,00	1,5	
	I do well in science to please parents	4 S	0,03	25,4	0,00	1,3	
	I do well in science to please myself	4 S	0,02	9,1	0,00	1,2	
	Important to me to do well in science	4 S	0,03	22,5	0,00	1,3	
	I usually do well in science	2 S	0,04	36,0	0,00	1,4	
	I think I'd like job involving science	4 S	0,03	23,3	0,00	1,3	
N			62			62	N
Mean			0,03			1,4	Mean
S.D.			0,03			0,2	S.D.

Blank in n categories means treated as continuous

Ed attainment	International mathematics score	.	P	0,19	871,2	0,0	3,1	46
	International science score	.	P	0,17	773,5	0,0	3,0	46
	Maths raw score	.	P	0,09	297,6	0,0	2,2	46
	Maths rasch score	.	P	0,15	613,1	0,0	2,8	46
	Standardized maths score	.	P	0,15	648,3	0,0	2,8	46
	Science raw score	.	P	0,07	210,6	0,0	2,0	46
	Science rasch score	.	P	0,12	507,1	0,0	2,6	46
	Standardized science score	.	P	0,14	581,4	0,0	2,7	46
	International mathematics score (EAP)	.	S	0,21	593,3	0,0	2,6	28
	International science score (EAP)	.	S	0,16	379,4	0,0	2,3	28
	Maths raw score		S	0,24	718,8	0,0	2,8	8
	Maths Rasch score	.	S	0,25	768,2	0,0	2,8	28
	Standardised maths score	.	S	0,25	761,9	0,0	2,8	28
	Science raw score	.	S	0,2	511,6	0,0	2,5	28
	Science Rasch score	.	S	0,21	583,7	0,0	2,6	28
	Standardised science score	.	S	0,22	592,8	0,0	2,6	28
N				16			16	16
Mean				0,18			2,6	35,8
StdDev				3,84			3,3	12,2

Ed-relevant	N of books in student's home	5 P	0,1	347,1	0,00	2,3	45
home chars	Important to mother good in sports	4 P	0,02	24,0	0,00	1,3	43
	Important to mother to have time for fun	4 P	0,04	86,9	0,00	1,6	43
	Home possesses calculator	2 P	0,04	85,4	0,00	1,6	46
	Home possesses computer	2 P	0,02	35,3	0,00	1,4	46
	Home possesses study desk	2 P	0,02	49,1	0,00	1,5	46
	Home possesses dictionary	2 P	0,03	70,2	0,00	1,5	46
	Home possesses encyclopedia	2 P	0,09	311,7	0,00	2,2	46
	Home possesses car	2 P	0,09	363,5	0,00	2,3	46
	Home possesses tumble dryer	2 P	0,03	75,2	0,00	1,6	46
	Home possesses dishwasher	2 P	0,08	242,6	0,00	2,1	46
	Home possesses microwave	2 P	0,02	38,4	0,00	1,4	46
	Home possesses video recorder	2 P	0,01	6,0	0,01	1,2	46
	Maths outside school extra lessons	4 P	0,06	173,7	0,00	1,9	42
	Important to mother do well in maths	4 P	0,01	7,9	0,00	1,2	44
	Sci outside school extra lessons	4 P	0,04	64,6	0,00	1,6	40
	Important to mother do well in science	4 P	0,03	45,9	0,00	1,4	43
	N of books in student's home	5 S	0,17	411,5	0,00	2,4	28
	Important to mother do well in language	4 S	0,01	3,2	0,07	1,1	27
	Important to mother good in sports	4 S	0,03	30,7	0,00	1,3	27
	Important to mother have time for fun	4 S	0,02	12,4	0,00	1,2	27
	Important to mother high achieve class	4 S	0,07	107,8	0,00	1,7	27
	Home possesses calculator	2 S	0,01	2,4	0,12	1,1	28
	Home possesses computer	2 S	0,02	11,3	0,00	1,2	28
	Home possesses study desk	2 S	0,02	17,2	0,00	1,3	28
	Home possesses dictionary	2 S	0,01	1,3	0,26	1,1	28
	Maths outside school extra lessons	4 S	0,05	57,2	0,00	1,5	26
	Important to mother do well in maths	4 S	0,02	9,8	0,00	1,2	28
	Science outside school extra lessons	4 S	0,02	13,7	0,00	1,2	26
	Important to mother do well in science	4 S	0,03	20,9	0,00	1,3	27
	Time spent maths homework per week	4 S	0,05	70,5	0,00	1,5	27
	Time spent science homework per week	4 S	0,07	119,4	0,00	1,7	27
	Time spent doing homework per day	4 S	0,09	171,3	0,00	1,8	27
	Mother went to college etc	3 S	0,03	31,0	0,00	1,4	27
	Father went to college etc	3 S	0,05	76,6	0,00	1,6	27
N			43			43	43
Mean			0,04			1,5	35,7
SD			0,03			0,4	9,0

Blank in n categories means treated as continuous

Ethnic groups	Born in country	2 P	0,04	80,2	0,00	1,6	46
	Father born in country	2 P	0,18	606,9	0,00	2,8	39
	Mother born in country	2 P	0,17	572,4	0,00	2,8	41
	Speak language of test at home	3 P	0,08	255,3	0,00	2,1	43
	Born in country	2 S	0,04	50,4	0,00	1,5	28
	Father born in country	2 S	0,27	735,5	0,00	2,8	27
	Mother born in country	2 S	0,26	729,4	0,00	2,8	27
	Speak language of test at home	3 S	0,11	196,0	0,00	2,0	27
N			8		1	8	8
Mean			0,14			2,3	34,8
StdDev			0,09			0,6	8,3

Blank in n categories means treated as continuous

Table 2 shows the values of the Intracluster Correlation Coefficients ρ grouped by category, together with some other related information. In each case we are dealing with a two-level hierarchy, pupils within schools. Under Leisure Activities, for primary, the value of ρ for Outside School Clubs is .05. The value of chi-sq given is the change in the likelihood ratio after the inclusion of a second (school) level and may be taken as a measure of the p-value of ρ. In this example it is 87 with 1 degree of freedom, giving a p-value of .00. The design effect (Deft) is 2.78 in this example. This means that in this example, the precision is slightly better than one-third of that of a simple random sample of comparable size.

RESULTS BY CATEGORY

Demographic

The Demographic group mainly comprises details of family structure as it relates to the survey pupil plus the gender balance. For Pupil Gender, ρ is very small (.01) for primary, but rather larger (.16) for secondary[3]. With the exception of Number of People Living at home, the ρs for the latter group seem to be relatively small, of the order of .01 or .02. The preponderance of the family structure ρs have a P-value of <.05, that is, they would be described as statistically significant at the .05 level, though this is scarcely surprising, since TIMSS studies contain very large numbers of pupils.

Leisure Activities

This is a relatively small category (15 items), asking what pupils do, or like to do, when they are not in school. Again the level 2 variation, though small, is statistically significant because of the large number of cases, with the average proportion of variation that is between schools being .04. For secondary school age pupils, though not primary, the two highest between school effects are playing with friends, and computer games, and watching TV and videos. This may represent a polarisation in the attitude to study between primary and secondary, with pupils at some schools increasingly concentrating on study, while the rest are less likely to occupy their time with school work after hours.

Attitudes towards school

Attitudes towards school contain a range of measures, but mainly study related, and maths-or science-related. With only one exception, (I like computers in my maths class, Secondary), the ρs are less than 0.20, averaging out around 0.03, at a comparable level to the leisure activities, or slightly lower. It could be that this exception, rather than being an attitude, represents a description of school provision or school policy.

Educational attainment

Educational attainment shows a much higher degree of clustering within schools, ranging from 0.07 to 0.25 with a mean ρ of 0.18. The mean ρ for primary schools (0.14) is lower than that for secondary schools (0.22|). This suggests that there is a relatively high degree of within-school clustering in academic attainment, and that this increases as pupils get older. These values are comparable to those found in a meta-analysis of school effectiveness studies (Bosker and Witziers, 1996).

Educationally relevant home characteristics

This category is a mixture of attitudinal, behavioural and material aspects of the student's home. Thus attitudinal questions include whether the parents consider it important to do well at school, while behavioural questions investigate such topics as time spent on homework, widely considered to be a very important factor in encouraging learning progress. Material aspects include obviously educationally relevant possessions such as a dictionary and a study desk, as well as less obviously relevant possessions such as a tumble drier. It is likely that the latter, while less likely to influence education progress directly, are indicative of a more comfortable background which will have the resources to promote the educational progress of its young people. Surprisingly, the degree of clustering of these characteristics is quite low (mean 0.04), suggesting that in a relatively affluent society these material characteristics no longer discriminate particularly well. The highest is number of books in the home, apparently still an indicator of attitudes towards education.

Ethnic and linguistic factors

The degree of clustering here is medium, with a mean of 0.14. The degree of clustering is lower for children (not) born in the country than for their parents, suggesting that clustering in these surveys is greater for second-generation immigrants than new immigrants. This of course is to some extent because numbers of new immigrants are relatively small.

CONCLUSIONS

In this presentation we have looked at the degree of within-school clustering of a range of variables in data from one country (England) in TIMSS. The variables have been classified into a number of broad topic areas, and the results compared between topics and between primary and secondary surveys.

The first strand of conclusions relates to the values of ρ.[4] The value of ρ for educational attainment is large, with a mean for the items considered of 0.18, compared with Hox's recommendation of 0.15 for a large ρ, and is larger for older pupils. This confirms the well-known observation that between-school differences in attainment are substantial. Similarly there is a medium intra-class correlation for ethnic and language groupings, confirming that minority groups tend to cluster in particular schools.

At the opposite end of the size of ρ, demographic (family formation), leisure activities, attitudes towards school and educationally relevant home characteristics have quite low values of ρ (mean < 0.5), defined as small. These are small, but not negligible, and taking account of the structure will mean a small but valuable improvement in the accuracy of the estimate of standard errors.

The second strand of conclusions relates to the implications of these findings for academics and policy makers. While it is perhaps unsurprising that despite all attempts at equalisation, there is still a substantial degree of segregation between schools on academic success, it is of slightly more concern that there is a comparable degree of segregation based on ethnic or linguistic characteristics. However, it seems that if academic attainments are segregated, factors considered to affect attainment, attitudes towards education and educationally relevant home characteristics, are very much less so.

If the sizeable value of ρ for academic attainment is no surprise, then it is more surprising that there is not a correspondingly high value of ρ for educationally relevant home characteristics. There are high values of ρ for pre-school attendance, but this is likely to be related to local provision of pre-school facilities. Among the remainder of the characteristics in this category, the size of ρ is related to the type of variable. It has always taken some thought to relate possession of a microwave to educational progress, but the rationale was that while not directly relevant such goods indicated a home that had sufficient resources to be able to ensure adequate educational provision for its children. It seems that increasing affluence has meant that such possessions are near enough to saturation in the population that they no longer distinguish sufficiently. Perhaps, at least in more developed countries, instead of asking whether a household possesses a car, we should ask what make and how old!

Within the more obviously educationally relevant features, such characteristics as the child having his own desk to study at, and access to a dictionary, seem not to be differentiated much among schools. It seems that increasing education awareness has meant that a much larger proportion of pupils now have the benefit of such important basic features. Amount of homework and number of books in the house have higher values of ρ.

It may be that some stratifying indicators are no longer important in distinguishing between schools in the United Kingdom. In an increasingly affluent society, characteristics that were once the mark of better-off circumstances are now so widespread that they do not distinguish among population groups. It is also remarkable that there appears to be little difference between schools on such factors as leisure activities, attitudes towards education and perception of school policies. Recent headlines have suggested the existence of a nation-wide shared informal culture, which is evidenced by pupils who can spell the name of a famous footballer, but not Shakespeare. Does such a shared youth culture actually exist, or is it more that attitudes are less differentiated because pupils can only compare themselves directly with other schools on assessed achievement, not on attitudes? If this were the sole explanation, one would expect that the relatively factual variables, such as what the student does out of school, should be more differentiated than attitudinal questions. In fact this does not seem to occur.

To sum up, one could say that on the crucial instrumental variables to do with education attainment, there still is a substantial difference between schools, but, off duty, the country's young people are much more similar.

TIMSS and other international studies of education attainment have generated a tremendous amount of interest of late. National comparisons have been headline news in the media and the results have been closely studied by policy makers (see, for example, Royal Statistical Society, 2003.) While the spread of attainment has been reported (see Beaton et al., 1996), the main focus of interest is, perhaps naturally, the mean or average level, though there has been some research looking at inequality. While the main focus of TIMSS has been the level of attainment, it has collected an outstanding set of data on a very wide range of factors, dealing with schools, home, families, opinions and attitudes. Naturally this has been mainly about factors considered to affect education attainment, but there is no reason why secondary analysis of this data should confine itself in this way. The current presentation, for example, has got input to the topic of attitudes of young people and youth culture.

It is intended to extend this work to compare results in countries with different degrees of development, social cohesion, etc.

References

BEATON, A. E., MULLIS, I.V.S., MARTIN, M.O., GONSALEZ, E.J., KELLY, D.L. & SMITH, T.A. (1966) *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study* Boston College: TIMSS International Study Centre.

BOCK, R. D., GIBBONS, R., SCHILLING, S. G., MURAKI, E., WILSON, D. T. & WOOD, R. (2002) TESTFACT4 (Computer Program). Chicago: SSI.

BOSKER, R.L. and WITZIERS, B. (1996). *The magnitude of school effects, or: Does it really matter which school a student attends?* Paper presented at the annual meeting of the American Educational Research Association, New York: NY.

GOLDSTEIN, H. (2003). *Multilevel Statistical Models, third edition.* London: Arnold.

GORARD, S. (2002). 'Let's keep it simple: the multilevel modelling debate', *Research Intelligence*, 81, 24–5.

HOX, J. (2002). *Multilevel Analysis: Techniques and Applications*. London: Lawrence Erlbaum.

JESSON, D. (2002). 'Response to Schagen and Goldstein', *Research Intelligence*, 80, 16–17.

KISH, L. (1965). *Survey Sampling*. London: John Wiley.

MOERBEEK, M., VAN BREUKELEN, G.J.P. and BERGER, M.P.F. (2000). 'Design issues for experiments in multilevel populations', *Journal of Educational and Behavioural Statistics*, 25, 3, 271–84.

RAUDENBUSH, S.W. and BRYK, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Second ed (Advanced Quantitative Techniques in the Social Sciences Series, 1). Thousand Oaks, CA: Sage.

RAUDENBUSH, S.W. and LIU, X. (2000). 'Statistical power and optimal design for multisite randomized trials', *Psychological Methods*, **5**, 2, 199–213.

REISE, S. and DUAN, N. (2002). 'Design issues in multilevel studies.' In: REISE, S. and DUAN, N. *Multilevel Modelling: Methodological Advances, Issues and Applications.* London: Lawrence Erlbaum Associates.

ROBINSON, W. S. (1950). 'Ecological correlations and the behaviour of individuals.' *American Sociological Review,* 15, 351-357.

ROYAL STATISTICAL SOCIETY (2003) *Conference on TIMSS and PISA*. London: Royal Statistical Society, 3rd December 2003.

SCHAGEN, I. and GOLDSTEIN, H. (2002). 'Do specialist schools add value? Some methodological problems', *Research Intelligence*, 80, 12–15.

SCHAGEN, I. and GOLDSTEIN, H. (2002). 'Serious reservations about the interpretation of what we said…', *Research Intelligence*, 81, 27.

SNIJDERS, T. and BOSKER, R. (1999). *Multilevel Analysis: an Introduction to Basic and Advanced Multilevel Modelling.* London: Sage.

NOTES

[1] Widely but not universally: for a contradictory point of view and discussion see Jesson (2002), Gorard (2002), Schagen and Goldstein, (2002), Goldstein and Schagen (2002),

[2] There is more than one definition of ρ in a three-level hierarchy, and a fortiori for more extended hierarchies.

[3] None of the schools samples here were designated single sex, so the degree of variation represents solely natural fluctuation.

[4] Many of these quantities are, strictly, ordinal rather than continuous. We follow common practice here by treating them as continuous.

ACKNOWLEDGMENTS

I am grateful to my NFER colleagues, Liz Cleaver, Lesley Kendall, Ian Schagen and Chris Whetton for taking the time to read this and earlier versions and making suggestions that have improved the structure of this paper.