Working Paper 10/09

A Geodemographic Classification of Primary Schools in London

Anne Gibbs, John Stillwell and Linda See

School of Geography
University of Leeds
Leeds LSJ 9JT

Emails: aegibbs@hotmail.co.uk
j.c.h.stillwell@leeds.ac.uk
linda.see@hotmail.co.uk

December 2010

Contents

List of Tables

List of Figures

Abstract

Using the '*k means*' clustering method, fourteen categories of primary schools in London have been identified on the basis of their ethnic composition and certain socio-economic characteristics. This typology can provide a framework for examining the geography of primary schools in London, the relationship that each school has with its neighbourhood and school performance, which is the subject of further research.

Keywords: geodemographics; education; primary schools; ethnicity; London.

1. Introduction

Parents are becoming increasingly concerned about the 'quality' of primary schools. At one time, children would simply attend the local primary; nowadays, the potential to choose a school makes the decision-making process more complex for parents as they weigh up distance from the school against other characteristics such as the type of school governance, resources and facilities, the socio-demographic complexion of the intake and, crucially, the academic track record of school performance. In this paper, we are particularly concerned about the ethnic dimension of primary schools and we seek to establish whether primary schools in London can be classified into groups or clusters that have similar ethnic and socio-economic characteristics. In so doing, we are interested in the geography of primary schools in London. Since the relationships are complex, it is contended that geodemographic classification is a useful analytical approach since it enables the researcher to summarise, simplify and order a large data set, thus highlighting similarities and differences between elements contained within it (Everitt *et al.*, 2001). In the context of school improvement, classification is a potentially helpful tool since, by identifying schools that have similar characteristics, benchmarking is facilitated and collaboration between schools is encouraged. Moreover, in an academic context, classifications can be used to explore quantitative relationships with other secondary data sets, such as attainment levels, or as sampling frames for the selection of schools for qualitative research, thus contributing to mixed methods analyses.

The research reported in this paper outlines the development of a geodemographic classification of primary schools in London based on data from the School Census. The paper begins with a review of the Families of Schools (FoS) classification, which has been published annually by the DCSF (now the Department for Education) since 2006 and highlights the limitations of this product. Data sources and the research methodology are then outlined and the resultant classification of London primary schools is presented. In the concluding section, consideration is given to the potential use of the primary school classification as a benchmarking tool or sampling framework and the means by which it might be updated or used interactively online by policy makers or researchers.

2. The Families of Schools Classification

The only known classification of London primary schools is that published annually since 2006 by the DCSF as part of the London Challenge initiative: ‘Primary Families of Schools’ (FoS). This defines 97 groups or ‘families’ of London primary schools, each with between 10 and 25 members, on the basis of their school composition and performance. As with the parallel analysis of secondary schools, available since 2003, primary schools were initially grouped together on the basis of prior attainment (measured by average points score at KS1) and context data (a composite measure comprising an average of the proportion of pupils eligible for FSM and the IDACI index). In the most recent edition, EAL and mobility were added as factors in the grouping of schools. The four factors were then weighted as follows: prior attainment (66%); context data (20%); EAL (10%); and mobility - defined as the percentage of pupils in Year 5 or Year 6 who have joined the school within the last two years - (4%) (DCSF, 2009b). In this latest edition, schools which were outliers on prior attainment data were somewhat arbitrarily reclassified by assigning them to the smallest family with the closest overall score within the nearest geographical area.

It is contended that the classification presented here is potentially more useful than the FoS (DCSF, 2009b) for the following reasons:

- It focuses on pupil characteristics, rather than prior attainment, which has been deliberately excluded as a defining factor in order to focus on identifying contextually similar schools for comparative purposes.
- The area-based IDACI is not a defining factor in the classification, although it is in FoS. The classification is derived using data related specifically to the pupils within the schools only.

- It takes account of SEN in grouping schools whereas the FoS does not.
- The factors in the classification are not weighted although, due to the range of variables used, ethnicity effectively has the greatest weight.

The recently published schools White Paper (DfE, 2010) proposes that Families of Schools database continues to be annually updated by the Department for Education and, indeed, extended to all the English regions .

3. Data Sources and the Selection of Variables

This section details the data sources and variables used in the analysis. The data were compiled from extracts of the three core DfE databases which together make up the National Pupil Database (NPD) (Jones and Elias, 2006): the pupil-level School Census (PLASC), the Key Stage 2 results tables and the school-level Edubase. Each of these elements is now examined in turn, showing how a database of school-level variables was compiled from which 11 variables were selected to characterise the schools' pupil populations and from which the school classification was derived.

A database of school-level variables was compiled for all state maintained primary schools in London, excluding special needs schools. Pupil-level data from the School Census (PLASC) and Key Stage 2 results tables were aggregated and averaged to the school-level as appropriate and matched to the Edubase data. The Spring Census, taken in May 2007, was chosen to align as closely as possible with the extract on pupil performance in the Key Stage 2 Standard Assessment Tests (SATs) taken in the same month. Table 1 details the pupil-level variables included in the standard PLASC extract.

Table 1: Variables extracted from PLASC

Variable	Description	Value
ETHNICGROUP_ SPR0x	Pupil’s ethnic group based on ethnic code	18 groups defined
ETHNICITYSOURCE_SPR0x	Source of pupil’s ethnicity code	Child (C), Parent (P), current school (S), previous school (T), Other (O)
FSMELIGIBLE_SPR0x	Pupil eligible for free school meals	Binary variable
CONNEXIONS_SPR0x	Do parents consent to pupil data being shared with Connexions?	Binary variable
LANGUAGEGROUP_SPR0x	Pupil’s language group based on language code	ENG English ENB Not known but believed to be English OTH Other than English OTB Not known but believed to be other than English REF Refused NOT Information not obtained INV Invalid code
GANDTINDICATOR_SPR0x	Indicates if the pupil is in the Gifted and Talented cohort for the school	Binary variable
MODEOFTRAVEL_SPR0x	Indicates the usual mode of travel used by the pupil for the greater part (in distance) of the journey to school	WLK Walk CYC Cycle CAR Car/van CRS Car share (with a pupil from another household) PSB Public service bus DSB Dedicated school bus BNK Bus (type not known) TXI Taxi TRN Train LUL London Underground MTL Metro/tram/light rail BDR Boarder- not applicable OTH Other
ENROLSTATUS_SPR0x	Indicates enrolment status of pupil	C (Current, single registration at this school), G (Guest, pupil not registered at this school but attending some lessons), M (Main current, dual registration), S (Current subsidiary, dual registration)
ENTRYDATE_SPR0x	Date of entry to current school	Date
LEAVINGDATE_SPR0x	Date pupil left current school	Date
PARTTIME_SPR0x	Indicates if the pupil is part time or not	Binary variable
BOARDER_SPR0x	Indicates if the pupil is a boarder or not	=ANY(B,6,7,N) (see data tables)
NCYEARACTUAL_SPR0x	The year group in which the pupil is taught regardless of their chronological age	=ANY(N1,N2,R,1-14,X) (see data tables)
SENPROVISION_SPR0x	Provision types under the SEN Code of Practice	=ANY(N,A,P,S) (see data tables)
SENUNITINDICATOR_SPR0x	Indicates if a pupil with SEN in a mainstream school is a member of a SEN unit	Binary variable

RESOURCEDPROVISIONINDICATOR_SPR0x	Indicates if a pupil with SEN in a mainstream school is a member of a resourced provision	Binary variable
LLSOA_SPR0x	National Statistics Postcode Directory Lower Layer Super Output Area derived from pupil's postcode.	
IDACISCORE_SPR0x	Income Deprivation Affecting Children Indices. IDACI Score	0-1
IDACIRANK_SPR0x	Income Deprivation Affecting Children Indices. IDACI Rank.	

Eleven variables were selected for use in the derivation of the schools classification, seven relating to the ethnicity of the pupils and four to their socio-economic characteristics. The variables were standardised using range standardisation, but no weighting was applied. The selected variables were:

Ethnicity variables

- % African
- % Caribbean
- % Indian
- % Bangladeshi
- % Pakistani
- % White British
- % Any Other White

Socio-economic variables

- % eligible for Free School Meals (FSM)
- % with English as an Additional Language (EAL)
- % with Special Educational Needs (SEN)
- mobility rate

The base population used in the percentage calculations is the total number of pupils in the school in the statutory years of primary education (Reception to Year 6), except for the mobility rate, for which it is the number of Year 6 pupils as at May 2007.

The reasons why these particular variables were chosen to characterise the schools' pupil populations are now considered. The primary pupil population of London, particularly inner London, is truly multi-ethnic. In line with the 2001 Census of Population, the PLASC data distinguish 16 different ethnicities. Binary variables for each of the 16 ethnic groups were derived from the original alphanumeric codes of the PLASC data. The particular ethnicities included in the analysis were selected on the grounds that they comprise the largest ethnic groupings within the pupil population and therefore the ones most likely to determine the overall character of a school's pupil population. Across London as a whole, the indigenous White British group is the largest ethnicity, accounting for just over a third (37.5%) of the total pupil population. Pupils of African origin form the next largest grouping (13.0%). Other minorities with significant representation in the total primary population are: Any Other White (9.3%), Caribbean (6.9%), Bangladeshi (5.2%), Indian (4.3%) and Pakistani (3.1%). A range of other minorities, including mixed races, and unclassified pupils account for the remaining 30% of the total population.

The socio-economic variables were also selected on the basis that they reflect fundamental characteristics of the school population which have been shown to significantly influence educational outcomes:

1. *Free School Meals (FSM):* eligibility for FSM is determined by a pupil's household being in receipt of specific statutory income benefits. Although it is widely recognised as a rather inadequate measure of socio-economic status (Hobbs and Vignoles, 2007), it is the only measure available at pupil and school level in the PLASC and is conventionally used in educational research and policy making. Deprivation is widely recognised as having a negative influence on educational attainment (Power *et al.*, 2002; DCSF, 2009a). In 2007, 28% of pupils in state maintained primary schools in London were eligible for FSM. Not surprisingly, this proportion is significantly higher in inner London schools than suburban

ones. Nevertheless, there are notable pockets of high FSM rates in some inner London districts in which wealthier families are located (such as: Hampstead; central parts of both Westminster and Kensington & Chelsea; Roehampton and Putney in the borough of Wandsworth).

2. *English as an Additional Language* (*EAL*)*:* clearly the proportion of pupils who are not native English speakers has a major impact on education policy and outcomes. In the state maintained sector, the proportion of pupils with EAL was 39% in 2007. However, this proportion varies considerably across London, being highest in the inner London boroughs, as one might expect, but also surprisingly high in suburban boroughs to the north and west. A binary EAL variable was derived from the original language group variable of the PLASC.

3. *Special Educational Needs* (*SEN*)*:* (encompassing language and communication difficulties; specific literacy difficulties, i.e. dyslexia; and emotional and behavioural difficulties): since 2001, there has been a national policy of 'inclusion' whereby children with special needs have been educated, as far as possible, in mainstream schools. Not surprisingly, the proportion of SEN children within a class or school can have a significant impact on the education of both the SEN children themselves and their mainstream peers. In May 2007, just under a quarter (23.2%) of all primary pupils in London were identified as having some form of special needs. The School Census distinguishes between the three different levels of provision defined in the Special Educational Needs Code of Practice (DfES, 2001). These levels of need and type of provision range from 'school action' through 'school action plus' to 'statemented'. All three categories have been aggregated in the current analysis to create a binary SEN indicator.

4. *Mobility*: the impact of children entering schools at non-standard times in their school career is a major issue in the state maintained sector at the moment as schools with high

mobility rates face extra challenges in seeking to meet Government targets of educational attainment (Ewens, 2005; Machin *et al.*, 2006; Strand and Demie, 2006). The mobility rate comprises three main sources: immigration; household moves; and parental choice. Many measures of mobility have been used by researchers, but in this study mobility rates are derived from the entry date recorded in the PLASC data by identifying all pupils in Year 6 who had enrolled within the last two years of their primary education (i.e. from the beginning of October or later in their Year 5). Just over a tenth (11.6%) of the London primary pupils who sat the KS2 tests in May 2007 fall into this definition of mobility. Nine schools were calculated to have a 100% mobility rate and were excluded from the analysis on the grounds that this rating was most likely to be a reflection of the school being newly-established rather than a true measure of the mobility of its pupil population. Schools with a high mobility rate are significantly more likely to also have a high proportion of pupils eligible for FSM, EAL or SEN (Table 2). This picture confirms the pupil-level findings of other researchers (Ewens, 2005; Machin *et al.*, 2006; Strand and Demie, 2006).

Table 2: Correlations between socio-economic variables

	EAL	SEN	Mobility Rate
FSM	0.624	0.518	0.279
EAL		0.246	0.264
SEN			0.210

Note: All correlations significant at the 0.01 level

The final database comprises 1,592 schools with a total of 562,728 pupils, as indicated in Table 3. State maintained primary schools are distributed somewhat unevenly throughout London: beyond the City of London where there is only one primary school, the number of schools per borough varies in inner London from 28 in Kensington & Chelsea to 68 in Southwark and in outer London from 29 in Kingston-upon-Thames to 73 in Barnet. This distribution reflects not just the distribution of young families in the capital, but also levels of family income, as the proportion of children attending state schools is lower in boroughs of higher average household income where a larger proportion of children attend private schools. Average school size is lower in inner London (a mean of 286 pupils) than outer London (a mean of 395). At the borough level, outside the City of London, average school size ranges from 229 in Kensington and Chelsea to 519 in Redbridge.

Table 3: Pupils in maintained primary schools in London by borough and school year, May 2007

London Borough	National Curriculum Year as at May 2007							Total		
	Reception	1	2	3	4	5	6	Pupils	Schools	Mean no. of pupils
City of London	30	30	28	31	28	29	26	202	1	202
Camden	1,481	1,530	1,495	1,462	1,408	1,417	1,403	10,196	38	268
Greenwich	2,701	2,591	2,720	2,685	2,691	2,655	2,575	18,618	64	291
Hackney	2,245	2,318	2,245	2,281	2,238	2,224	2,202	15,753	52	303
Hammersmith & Fulham	1,270	1,203	1,266	1,183	1,144	1,230	1,194	8,490	35	243
Islington	1,792	1,811	1,856	1,802	1,784	1,813	1,767	12,625	45	281
Kensington & Chelsea	937	923	906	889	939	924	891	6,409	28	229
Lambeth	2,622	2,675	2,622	2,540	2,474	2,487	2,404	17,824	58	307
Lewisham	2,843	2,753	2,878	2,720	2,764	2,831	2,727	19,516	65	300
Southwark	2,910	2,934	2,896	2,833	2,890	2,917	2,801	20,181	68	297
Tower Hamlets	2,925	2,895	2,861	2.732	2,833	2,820	2,662	19,728	65	304
Wandsworth	2,297	2,275	2,219	2,226	2,099	2,205	2,087	15,408	55	280
Westminster	1,455	1,444	1,465	1,367	1,389	1,437	1,346	9,903	37	268
Barking & Dagenham	2,418	2,394	2,309	2,341	2,348	2,444	2,335	16,589	35	474
Barnet	3,315	3,316	3,477	3,349	3,379	3,406	3,338	23,580	73	323
Bexley	2,502	2,615	2,639	2,778	2,705	2,706	2,804	18,749	53	354
Brent	3,034	3,001	3,087	2,919	2,975	2,992	2,930	20,938	49	427
Bromley	3,186	3,224	3,270	3,347	3,321	3,422	3,376	23,146	64	362
Croydon	3,728	3,818	3,967	3,793	3,949	3,947	3,893	27,095	72	376
Ealing	3,405	3,369	3,325	3,297	3,358	3,312	3,288	23,354	60	389
Enfield	3,581	3,583	3,616	3,562	3,656	3,714	3,641	25,353	57	445
Haringey	2,894	2,782	2,757	2,696	2,767	2,751	2,707	19,354	53	365
Harrow	2,309	2,273	2,314	2,410	2,385	2,513	2,441	16,645	35	476
Havering	2,389	2,496	2,557	2,637	2,679	2,766	2,770	18,294	50	366
Hillingdon	2,956	2,995	3,020	3,067	3,115	3,043	3,018	21,214	52	408
Hounslow	2,459	2,388	2,338	2,447	2,365	2,517	2,351	16,865	45	375
Kingston-upon-Thames	1,477	1,495	1,493	1,498	1,515	1,490	1,491	10,459	29	361

Merton	1,879	1,845	1,838	1,794	1,794	1,836	1,685	12,671	43	295
Newham	3,818	3,822	3,877	3,842	3,799	3,882	3,893	26,933	61	442
Redbridge	3,065	3,044	3,096	3,090	3,072	3,243	3,175	21,785	42	519
Richmond-upon-Thames	1,847	1,835	1,797	1,823	1,744	1,610	1,556	12,212	32	382
Sutton	1,809	1,854	1,892	1,903	1,973	1,974	1,926	13,331	33	404
Waltham Forest	2,826	2,781	2,801	2,727	2,726	2,762	2,685	19,308	45	429
Total	**80,405**	**80,312**	**80,927**	**80,071**	**80,306**	**81,319**	**79,388**	**562,728**	**1,592**	**353**

The PLASC data were matched with an extract from the KS2 results tables for May 2007. A standard extract was obtained but many of the variables were not relevant to the current analysis and were therefore discarded. Table 4 details the variables used either directly or to derive new variables.

Table 4: Variables used from NPD KS2 Tables, May 2007

NPD KS2 Variable	Description	Value
KS2_ENGLEV	National curriculum level awarded for English test	2-5, etc.
KS2_MATLEV	National curriculum level awarded for maths test	2-5, etc
KS2_SCILEV	National curriculum level awarded for science test	2-5, etc
KS2_LEVXENG	Achieved level 4 or above (expected level) in KS2 English	Binary variable
KS2_LEVXMAT	Achieved level 4 or above (expected level) in KS2 maths	Binary variable
KS2_LEVXSCI	Achieved level 4 or above (expected level) in KS2 science	Binary variable
KS2_LEVXENGMAT	Achieved level 4 or above (expected attainment) in KS2 English and maths	Binary variable
KS2_LEVXEMS	Achieved level 4 or above (expected attainment) in KS2 English, maths and science	Binary variable

In measuring performance for the current research, it was decided to concentrate on English and Maths results, excluding Science from the analysis. This is because it is widely accepted that English and Maths are the most crucial of the three ‘core subjects’ in which pupils to date have been tested at KS2. Indeed, at the time of the analysis, it had just been announced that 2009 was the last in which pupils would sit a test in Science at KS2 (Eason, 2009). The following three variables, encompassing the full range of possible test outcomes, were computed for each of the English and Maths results: low (below National Curriculum Level 4), average (Level 4) and high (Level 5). Level 4 is the target level set by the Government at KS2, whilst

Level 5 is the highest level which can be attained in the tests. As with the School Census data, the performance variables were then converted to percentages, but in this case the base population was taken to be the subset comprising the Year 6 pupils entered for the SATs in each school, rather than the total pupil population of the school.

The aggregated pupil data were then matched with details of the names, postcodes and governance of the schools obtained from an Edubase file, dated November 2006. Five types of school governance or 'Type of Establishment' (ToE) are distinguished in Edubase, only four of which are relevant to the primary sector. The vast majority of primaries in London are either community schools, managed by a governing board selected by the Local Education Authority (LEA), or Voluntary Aided (VA) schools, managed by a charitable foundation (which is most usually a religious organisation; principally, the Church of England (CofE) or the Roman Catholic (RC) church). The remaining minority of schools are divided between the other two ToEs: Voluntary Controlled (VC) schools (in which the land and buildings of the school are owned by a charitable foundation but the LEA is responsible for the governance of the school), and foundation schools. For the purposes of the current research, it was decided that the principal distinction to be made as regards school governance was between schools managed by the LEA (i.e. community and VC schools) and those managed by external organisations (VA and foundation schools) and so a new binary variable (GOVERN) was derived to indicate this distinction.

Nearly seven out of ten London primary schools (69.2%) are governed by LEAs, with the remaining third (30.8%) being governed by other organisations, principally faith-based. The pupil profile of these two types of schools is quite distinct. On average, LEA-controlled schools have a higher proportion of pupils eligible for free school meals (FSM), with English as an additional language (EAL) or, to a lesser extent, SEN (Table 5). It is suggested that, together, these variables identify the proportion of 'needy' children in a school. Thus, it follows that the most needy children tend to be found in community schools. Furthermore, LEA-controlled schools have a higher average mobility rate.

Table 5: Socio-economic characteristics of pupils by school governance

Governance	Statistic	FSM	EAL	SEN	MOBILE
LEA-controlled	Mean	31.6	42.3	24.6	12.7
	Std. Deviation	17.4	26.1	9.9	7.6
VA / Foundation	Mean	20.4	32.2	20.2	9.3
	Std. Deviation	15.2	22.7	8.4	7.0
Total	Mean	28.1	39.2	23.2	11.6
	Std. Deviation	17.5	25.5	9.7	7.6

Figure 1 shows the distribution of the two types of governance by London borough, from which it is evident that some boroughs have a much higher proportion of VA/foundation schools than others. Westminster has a majority of such schools (70%), and four other boroughs each have around half their schools controlled by organisations other than the LEA (Kensington and Chelsea, Camden, Kingston-upon-Thames, and Richmond-upon Thames). Conversely, the proportion of LEA-controlled schools is above 80% in the following four boroughs (in declining order): Newham, Waltham Forest, Havering, and Hounslow.

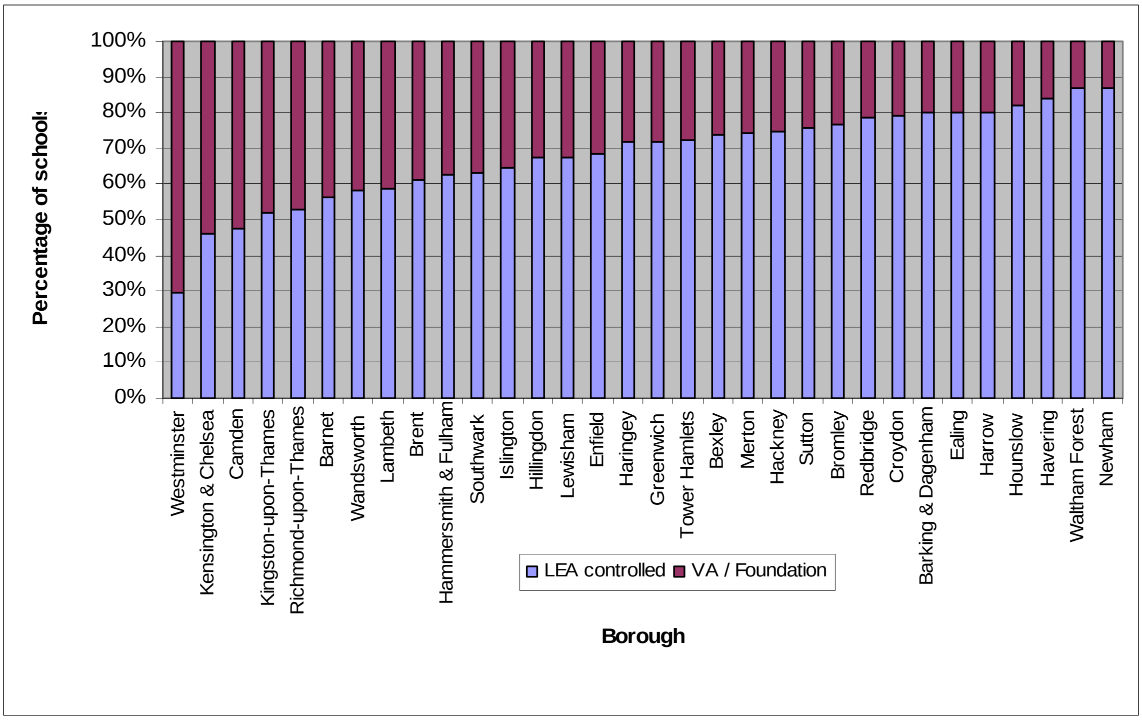


Figure 1: School governance by borough

Using GeoConvert, the school postcodes were then matched to grid references to allow the schools to be matched to their respective Lower Super Output Areas (LSOAs). Supplementary data comprising the IDACI and Index of Multiple Deprivation (IMD) scores and ranks for the LSOA in which the schools lie were then downloaded from the Department of Communities and Local Government (DCLG) website and appended to the database.

4. Deriving the Classification

The classification was derived by inputting the selected variables to the *k-means* cluster analysis algorithm in SPSS. This is one of the most commonly used methods in geodemographic classification of neighbourhoods such as MOSAIC (Harris *et al.*, 2005) and was also used in the census area classifications, in conjunction with Ward's clustering method (Office for National Statistics, 2004; Vickers *et al.*, 2003). The principal drawback of this method is that the number of clusters has to be pre-determined. However, this was overcome by creating classifications consisting of 4 to 22 clusters and then assessing their relative merits. The ultimate aim of the analysis is to define clusters which are distinct and where the individual schools are tightly concentrated within each cluster. Thus the alternative solutions were assessed in a two-stage process. First, all the solutions were compared using the following two measures: (i) the average distance from cluster centre (using the standardised data) to indicate compactness of the clusters, and (ii) the average difference from mean cluster membership to indicate the evenness of cluster size.

Second, the four most favourable solutions from the first stage of the assessment were further analysed for the homogeneity and separateness of the proposed clusters, using the raw data and the three measures of cluster validity detailed below. The results of the first stage of the assessment are displayed in Figure 2. Step reductions in both measures, indicating significant improvements in the clustering solution at particular points, resulted in the classifications with 7, 10, 14 and 16 clusters being 'short listed'.

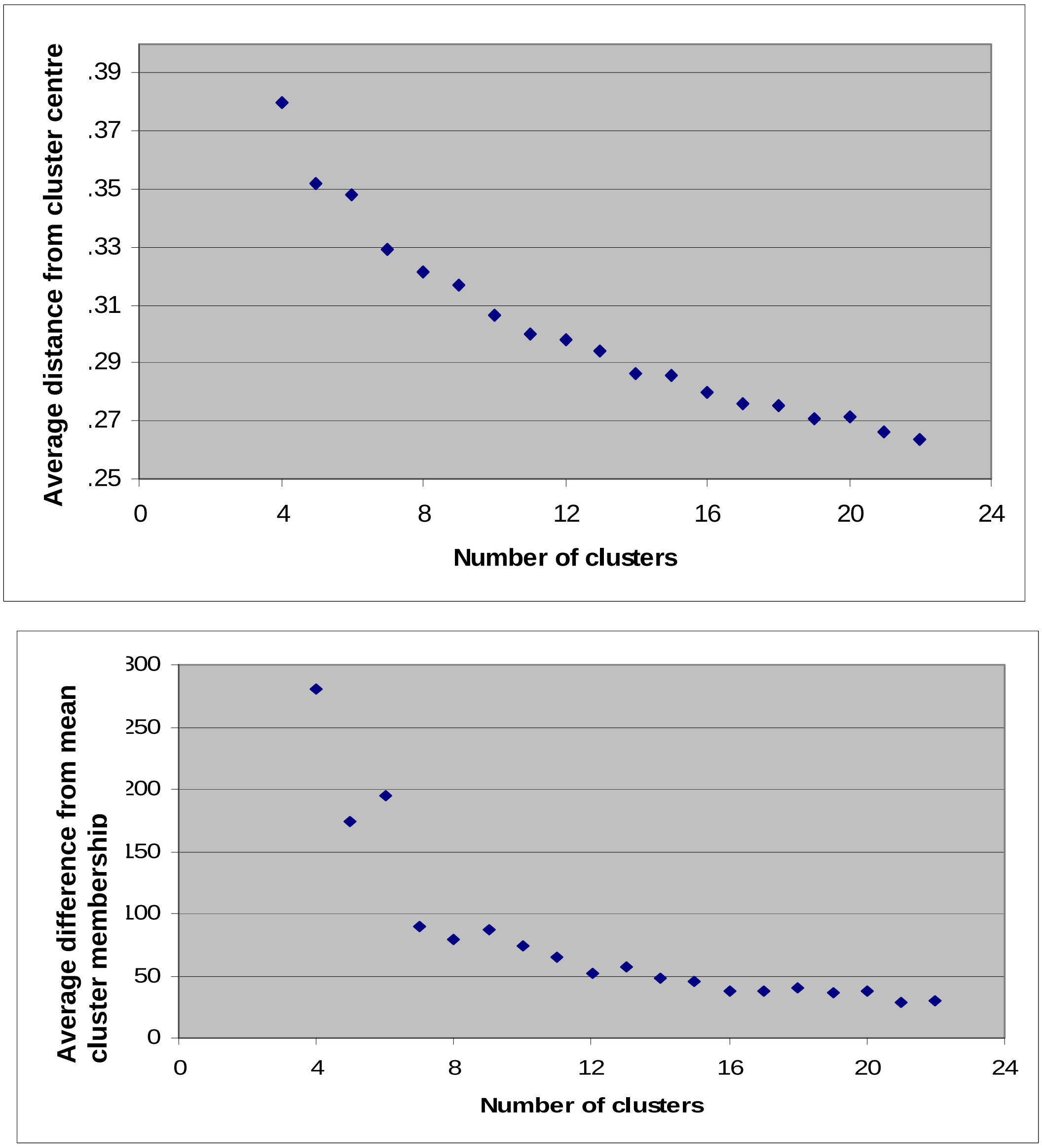


Figure 2: Evaluating alternative cluster solutions

The results of the second-stage assessment are shown in Table 6. Three measures of cluster validity have been used: the Silhouette Index (SI), the Davies-Bouldin Index (DBI) and the average distance to cluster centre. The SI measures how appropriately each individual case is assigned to its cluster (Rousseeuw, 1987). The silhouette width is calculated as follows: suppose a represents the average distance of a case from all other cases in its cluster and b represents the minimum of the average distances of the case from the cases of the other clusters, then the silhouette width s of the case is defined as:

$$s = (b - a) / \max[a, b] \qquad (1)$$

The value of *s* varies from -1 to 1. The closer the value is to 1, the better clustered the case is. Conversely, a value close to -1 means that the case is 'misclassified' and actually lies somewhere in-between the clusters. Once the 'silhouette width' of each individual case has been calculated, these can then be aggregated and averaged at the cluster and dataset levels to give the SI reflecting the compactness and separateness of the clusters.

The DBI is a function of the ratio of within-cluster distances to between-cluster separation: clusters with desirable characteristics will have low values (Davies and Bouldin, 1979). It is calculated as follows: for each pair of clusters: if a_i is the average distance of all cases in cluster *i* to the centre of that cluster, and b_{ij} is the distance between clusters *i* and *j,* calculate *R* as:

$$R = \max_{i \neq j} \frac{a_i + a_j}{b_{ij}} \qquad (2)$$

Then,

$$DBI = \frac{1}{N} \sum_{i=1}^{N} R_i \qquad (3)$$

The average distance to the cluster centre is a simpler measure of cluster validity (Everitt *et al.*, 2001) and is an entirely internal one, i.e. it does not take into account distances *between* clusters. It is therefore limited in terms of providing an overall assessment of the cluster solution but, nevertheless, still provides a useful measure of the compactness of the clusters. The values of these evaluation statistics are presented in Table 5 for each of the possible solutions indicating that the 14-cluster solution has the most favourable set of characteristics overall, and is therefore the best classification. The final classification explains 66.8% of the variance in the dataset, but rather more of the variance in ethnicity (71.5%) than in socio-economic factors (58.4%). The characteristics of the 14 clusters are detailed in Table 7.

Table 6: Evaluation of favoured clustering solutions

Number of clusters	Silhouette Index (high preferable)	Davies-Bouldin Index (low preferable)	Average Distance to Cluster Centre (low preferable)
7	0.362	2.461	51.975
10	0.316	4.955	77.298
14	0.307	1.582	22.499
16	0.260	4.506	22.115

Table 7: Cluster characteristics (indexed to global average = 100)

Cluster	No. of schools	LEA-controlled schools	VA / foundation schools	African	Caribbean	Indian	Bangla-deshi	Paki-stani	White British	Any Other White	FSM	EAL	SEN	Mobile
A2	148	129	35	56	41	44	19	39	177	56	102	40	126	111
A5	246	85	133	25	17	44	6	19	209	47	25	18	66	64
A9	173	71	165	35	38	70	15	45	145	137	36	54	69	72
B3	73	38	240	343	158	21	25	10	38	61	118	139	104	77
B8	111	82	140	144	374	47	23	74	47	76	120	88	115	97
B11	140	115	67	132	93	49	96	61	94	96	160	106	136	97
B13	60	60	189	64	78	60	37	29	67	310	75	123	75	88
B14	184	89	125	95	128	102	27	116	91	96	77	88	93	95
C1	76	139	13	75	57	793	85	423	22	43	82	194	91	111
C6	34	136	19	88	81	216	160	929	25	90	118	190	117	113
C7	126	130	33	193	197	49	87	52	30	201	181	159	140	126
C10	71	126	41	131	97	130	52	132	51	133	133	155	114	260
C12	97	121	54	142	84	114	258	165	37	109	161	180	105	112
D4	53	131	31	43	16	14	1425	32	19	29	204	216	92	91
Total/ Global Average	**1592**	**69.2**	**30.8**	**13.0**	**6.9**	**4.3**	**5.2**	**3.1**	**37.5**	**9.3**	**28.1**	**39.2**	**23.2**	**11.6**

As it can be difficult to effectively characterise as many as 14 clusters, the following four groups have been discerned:

A. Clusters containing schools with predominantly well-off White British pupils (Clusters 5, 2 and 9);

B. Clusters containing schools with ethnically-mixed European and second generation African and Caribbean migrant pupil populations (Clusters 14, 8, 3, 13 and 11);

C. Clusters containing schools predominantly controlled by the LEA with high proportions of first generation South Asian migrant pupils (Clusters 1, 6,7, 10 and 12);

D. Clusters containing schools with predominantly needy Bangladeshi pupil populations (Cluster 4).

The clusters and groups are graphically represented in Figure 3. In this diagram, the clusters are arranged to reflect, as closely as possible, the relative distances between them in multi-variate space, whilst also indicating their placement on the horizontal 'ethnicity' axis and the vertical 'socio-economic' axis. At one end of the spectrum, in the top left of the diagram, the clusters are characterised by suburban schools with predominantly well-off White British pupils. The continuum progresses through clusters of schools with varying proportions of ethnic minority Afro-Caribbean and South East Asian migrant pupils from poorer households. At the other end of the spectrum, in the bottom right of the diagram, is a cluster of schools, located almost exclusively in the Tower Hamlets, in which the majority of pupils are of Bangladeshi origin and very needy.

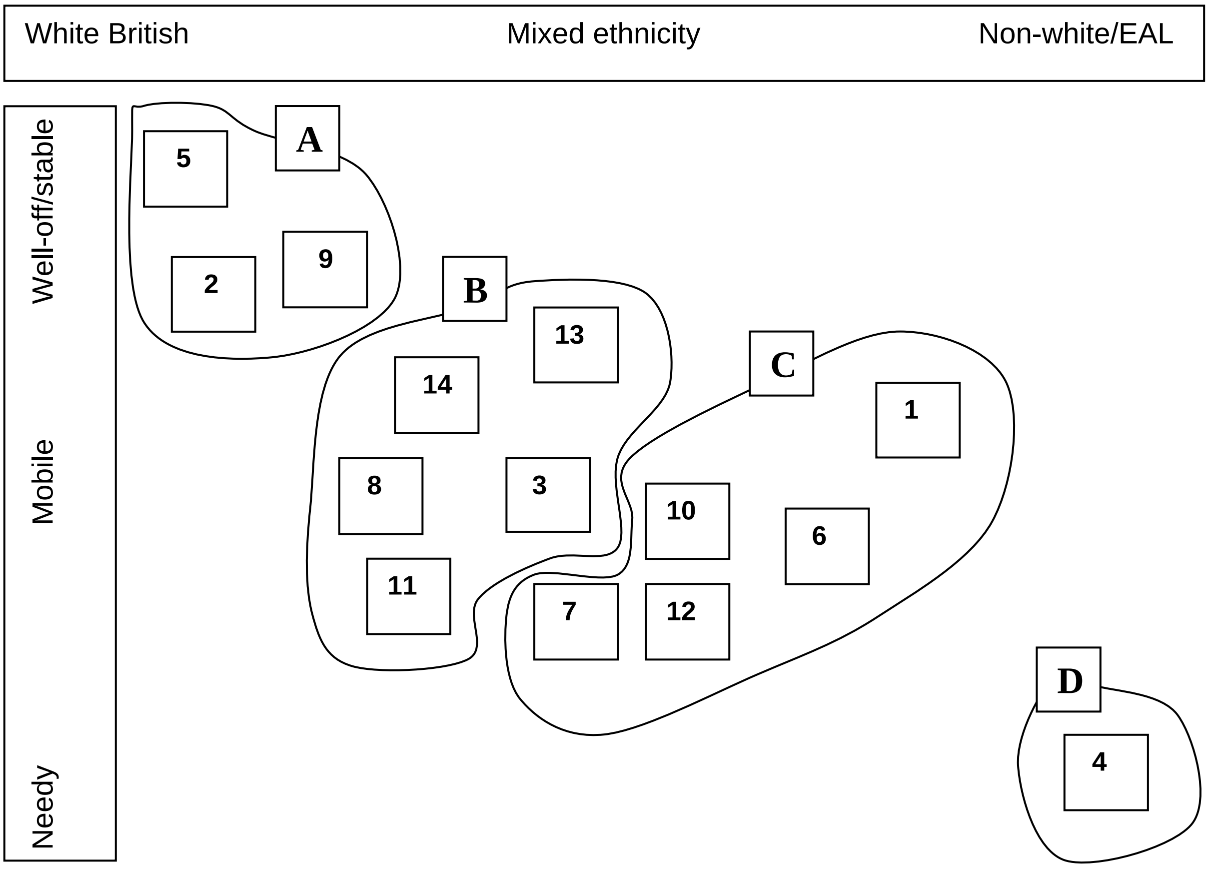


Figure 3: Cluster visualisation. Source: After Harris *et al.* (2005, Figure 6.3, p.170).

The radar charts in Figures 4-7 provide an alternative form of visualisation. In these diagrams each radial 'spoke' represents either one of the clustering variables or the categorical variables defining the governance of the schools, whilst each ring shows the variable values and the shaded plot connects the cluster averages for the variables.

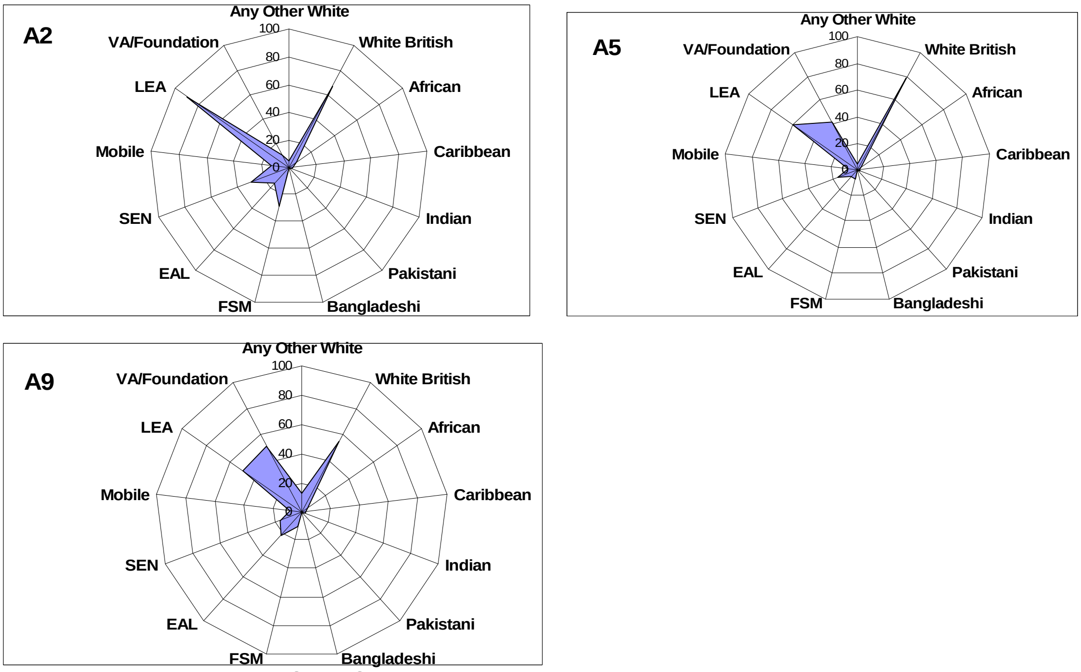
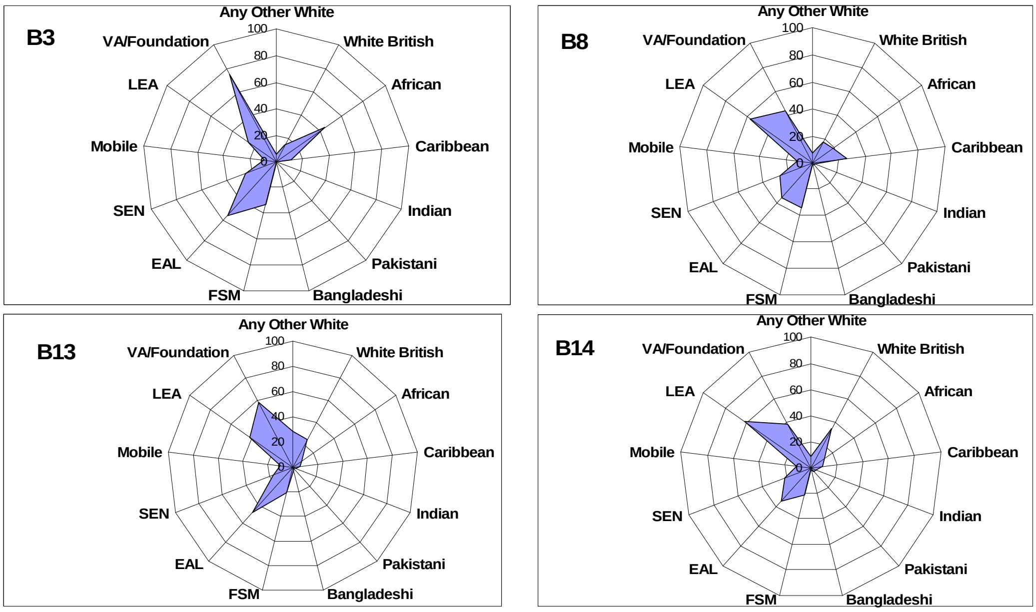


Figure 4: Cluster profiles for Group A



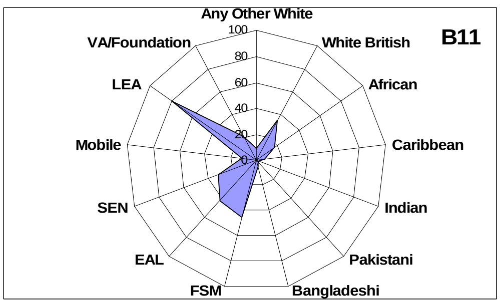


Figure 5: Cluster profiles for Group B

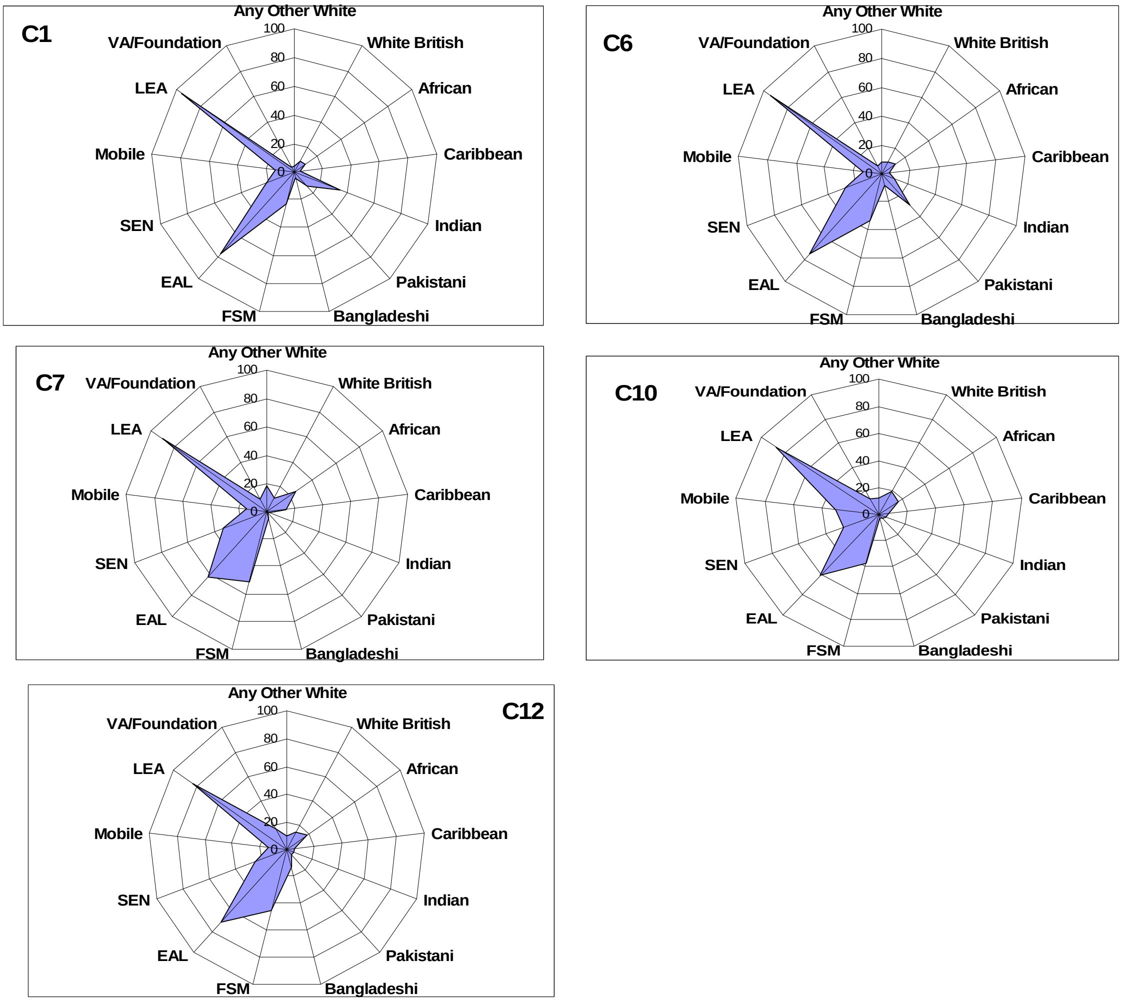


Figure 6: Cluster profiles for Group C

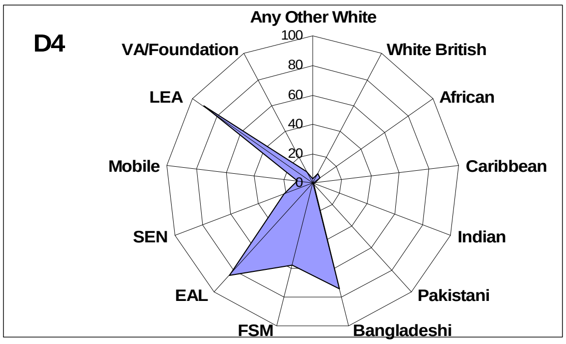


Figure 7: Cluster profile for Group D

Cluster D4, the only one in Group D, is the most distinctive, due to its very high proportion of needy (as indicated by high rates of FSM and EAL) Bangladeshi pupils. By contrast, the Group A clusters all have a high proportion of White British pupils. There is a relatively high proportion of well-off pupils in a disproportionate number of VA and foundation schools in clusters A5 and A9, although higher rates of FSM and SEN prevail in cluster A2 than others. The key characteristics of Group B clusters are a high proportion of VA and Foundation schools, especially in all but B11, and a multi-racial pupil population with relatively high proportions of African and Caribbean pupils. Finally, Group C clusters are predominantly LEA controlled with multi-racial pupil populations, reflected in a high rate of EAL, although the dominant ethnicity varies between clusters (e.g. C1 Indian, C6 Pakistani). Schools in C10 have a much higher average mobility rate than schools in other clusters or groups. Schools in these Group C clusters generally cater for children that are both economically and educationally needy, with relatively high rates of SEN and FSM as well as EAL (especially in C7 and C12).

5. Conclusions

The school classification presented here provides a useful multi-dimensional typology of London primary schools, which highlights the similarities and differences between them. Inevitably, there is much more analysis which could be done. Geodemographic analysis has its critics (Voas and Williams, 2001) but, as Harris *et al.* (2007) have recently suggested, the resultant classifications remain potentially valuable as a framework for analysis. Thus the classification presented here has the potential to provide the context for more detailed analyses of ethnic and social disparities between schools or, more specifically, to be used as a sampling frame for qualitative analysis and/or as a benchmarking tool by managers and policy makers.

The classification presented here does, of course, represent just a snapshot of London primary schools at one point in time: 2007. Its true value as an analytical and policy-making tool would be realised only if it were to be regularly updated and made available online. Such systems, which are already available in the US (Wisconsin Department of Public Instruction, 2009; State of California, 2007), enable more flexibility in the identification of similar schools which can then vary according both to the selected criterion and over time.

The opportunity for more detailed analyses of the relationship between schools and their neighbourhoods might be provided by the availability of more regularly updated local population statistics through the development of the Integrated Population Statistics System (IPSS) which is due to replace the decennial Census from 2011 onwards (ONS, 2003). This could be matched through geographical information systems (GIS) with the PLASC data. In the development of any future online system, however, it will be important to balance the need to update the classification to reflect changes in the socio-economic and ethnic composition of the pupil populations of schools, with the twin needs of maintaining a reasonably stable classification, against which temporal changes in performance of the various clusters of schools can be plotted, and enabling schools to compare themselves with a relatively constant set of peers. Some criteria would have to be decided by the researchers and users of the system against which any potential changes would be assessed (Sheldon *et al.*, 2002).

Acknowledgements

The data used in this analysis is from the PLASC and KS2 Tables of the National Pupil Database held by the Department for Education (formerly the DCSF). This research was conducted for a dissertation by Anne Gibbs in partial fulfilment of the requirements of a Masters Degree in Geographic Information Systems at the University of Leeds.

References

Davies, D.L. and Bouldin, D.W. (1979) A cluster separation measure, *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 1(4), 224-227.

DCSF (2009a) *Breaking the Link Between Disadvantage and Low Attainment: Everyone's Business*, DCSF-00357-2009, March. Downloaded from Department of Education website: http://publications.education.gov.uk/default.aspx?PageFunction=productdetails&PageMode=p

ublications&ProductId=DCSF-00357-2009&

DCSF (2009b) *Primary Families of Schools*, downloaded from http://fos.dcsf.gov.uk/ 28th July 2010.

Department for Education and Skills (2001) *Special Educational Needs Code of Practice*, viewed online at http://www.teachernet.gov.uk/_doc/3724/SENCodeOfPractice.pdf on 18th October 2010

Department for Education (2010) *The Importance of Teaching: The Schools White Paper 2010*, downloaded from the Department for Education website at http://publications.education.gov.uk/eOrderingDownload/CM-7980.pdf on 1st December 2010.

Eason, G. (2009) *All SATs Tests Might Go*, online news report downloaded from http://news.bbc.co.uk/1/low/education/8036755.stm Last accessed 28/9/09.

Everitt, B.S., Landau, S. and Leese, M. (2001) *Cluster Analysis*, 4th edition, Hodder Arnold, London.

Ewens, D. (2005) Moving home and changing school 1: widening the analysis of pupil mobility, *DMAG Briefing 2005/32*, Greater London Authority, November.

Harris, R., Sleight, P. and Webber, R. (2005) *Geodemographics, GIS and Neighbourhood Targeting*, Wiley, Chichester.

Harris, R., Johnston, R. and Burgess, S. (2007) Neighbourhoods, ethnicity and school choice: Developing a statistical framework for geodemographic analysis, *Population Research Policy Review*, 26, 553-579.

Hobbs, G. and Vignoles, A. (2007) Is free school meals status a valid proxy for socio-economic status (in schools research)?, *CEE DP 84*, Centre for the Economics of Education, London School of Economics, May.

Jones, P. and Elias, P. (2006) Administrative data as a research resource: a selected audit, *Economic and Social Research Council Regional Review Board Report 43/06*, Warwick Institute for Employment Research.

Machin, S., Telhaj, S. and Wilson, J. (2006) The mobility of English school children, *Fiscal Studies*, 27/3, 253-280.

Office for National Statistics (2003) '*Proposals for an Integrated Population Statistics System*', ONS, October.

Office for National Statistics (2004) '*National Statistics 2001 Area Classification for Local Authorities User Guide'*, downloaded from http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/downloads/User_Guide.pdf . Last accessed 20/4/09.

Power, S., Warren, S., Gillborn, D., Clark, A., Thomas, S. & Coate, K. (2002) *Education in Deprived Areas: Outcomes, Inputs and Processes*, Perspectives on Education Policy, (London, Institute of Education).

Rousseeuw, P.J. (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics*, 20, 53-65.

Sheldon, G., Hall, R., Brunsdon, C., Charlton, M, Alvanides, S. and Mostratos, N. (2002) *Maintaining Basic Command Unit and Crime And Disorder Partnership Families For Comparative Purposes*, Home Office, July.

Specialist Schools and Academies Trust (2010) Educational Outcomes 2009, viewed online at https://www.ssatrust.org.uk/newtechnologies/datadriventransformation/Pages/EO09overview.aspx#anchor on 26th June 2010

State of California (2007) *Overview of California's 2006 Similar Schools Ranks based on the API*', March. Viewed online at http://www.cde.ca.gov/TA/ac/ap/documents/simschl06b.pdf Last accessed 20/4/09.

Strand, S. and Demie, F. (2006) Pupil mobility, attainment and progress in primary school', *British Educational Research Journal*, 32(4), 551-568.

Vickers, D., Rees, P. and Birkin, M. (2003) A new classification of UK local authorities using 2001 Census Key Statistics, *Working Paper 03/03*, School of Geography, University of Leeds, Leeds.

Voas, D. and Williamson, P. (2001) The diversity of diversity: a critique of geodemographic classification, *Area*, 33(1), 63-76.

Wisconsin Department of Public Instruction (2009) *Wisconsin Information Network for Successful Schools*, online at http://dpi.wi.gov/winss/similar.html. Last accessed 20/4/09.