Department of Quantitative Social Science

Languages, ethnicity, and education in London

Michelle von Ahn
Ruth Lupton
Charley Greenwood
Dick Wiggins

IOE
LONDON

Leading education
and social research
Institute of Education
University of London

DISCLAIMER

Languages, ethnicity, and education in London

Michelle von Ahn[*], Ruth Lupton[†], Charley Greenwood[‡], Dick Wiggins[§][¶]

Abstract. For the first time in 2008 the Annual School Census (ASC) required all schools to provide pupil information on the language spoken at home. Our analysis focuses on children attending state schools in London. Over 300 languages are spoken by London pupils, around 60% of London pupils are English speakers however, there are over 40 languages spoken by more than 1,000 pupils. Bengali, Urdu and Somali are the top three languages spoken in London, other than English. We show that English has a 'doughnut' shaped geographical distribution in London, being the predominant language in most of Outer London. Languages other than English are more common in Inner London. Most minority languages, such as Bengali, Urdu and Turkish, have one, two or three main clusters, reflected settled immigrant communities. However others, notably Somali, are widely dispersed. This has implications for service provision. Some of the ethnic categories that are widely used in analysis of Census data hide substantial linguistic diversity, particularly 'Black African' and 'White Other'. Within London, where these groups are numerous, language data provides a valuable disaggregation of these heterogeneous groups. Our work suggests that language spoken provides a means to better understand the relationship between ethnicity and educational performance.

JEL classification: I20,R23,Z13.

Keywords: Languages, London, ethnicity, educational performance.

[*]London Borough of Newham. E-mail: (Michelle.vonAhn@newham.gov.uk)

[†]London School of Economics and Political Science. E-mail: (; R.Lupton@lse.ac.uk)

[‡]Institute of Education, University of London. E-mail: (c.greenwood@ioe.ac.uk)

[§]Institute of Education, University of London. E-mail: (r.wiggins@ioe.ac.uk)

Section 1

The increasing linguistic diversity of the UK attracts much interest and debate among public service providers, educationalists and the public. The presence of languages other than English has been seen as both an asset and a liability in education (Tosi 1984; Mehmedbegović 2007), an economic opportunity (Martin 2000) and a major economic cost (BBC 2006), an expression of multiculturalism and a threat to community cohesion. More immediate pragmatic concerns include establishing whether we have the language skills to do business with the rest of the world, and where these skills are located, where and in what contexts is there a need for translation skills or language classes, and to what extent the languages of the five London Boroughs hosting the 2012 Olympics match those of our world visitors.

Yet remarkably, little is known about the numbers of people who speak different languages, and the implications of this dimension of population composition and change. Language data has not been collected in the Census of Population, and until 2007, School Census data only contained information about English as an Additional Language (EAL) not the individual languages. A 'model' question asking about the language spoken at home was inserted in 2007(DfES 2006 a and b) and from 2008, all schools were required to collect this data. Although this source only provides information about school pupils, not about all residents, it represents a major advance on our knowledge thus far.

London is by far the most linguistically diverse part of the UK (Figure 1), and was also the subject of an earlier analysis *Multilingual Capital* (Baker and Eversley 2000), based on pupil data collected from individual London Boroughs, which identified over 300 languages spoken by London school children. In this paper, we present selected findings of analysis of 2008 Annual School Census data for the capital to quantify and map the languages spoken in contemporary London, updating Baker and Eversley's work and showing how the situation has changed.   We show how an analysis of language rather than ethnicity alone can shed new light on patterns of educational inequality.

Data, Language Classification and Geography

Our analysis draws on data from the Annual School Census in 2008.   This covers pupils in all state schools, but not private schools**.** 1.1 million pupils are represented in the dataset. The data collection instrument provides a list of 322 languages, of which some are variants of other languages. For example, a person may be classified as speaking Bengali (main category) or Bengali (Sylheti) or Bengali (Chittagong/Noakhali) or Bengali (any other). This structure provides some inconsistencies as different local authorities have collected data at different levels – some using main categories only and some using variants – and for this reason, analysis of the variants is not reliable. Furthermore, there are even less detailed categories such as "believed to be English" and "other than English". For London as a whole, 7% of records have this insufficiency of language detail and there are tiny percentages of refusals and missingness (in total 1.4%). However, at Borough level, the ambiguity can range from 2.3% (Ealing) to

27.9% (Westminster). Clearly data quality issues need to be address if the data is to enable sophisticated analysis of change over time.

In line with the approach taken in *Multilingual Capital*, we classify languages based primarily on their location in the world (Dalby 1999), using a classification developed for this project by Eversley and Sanderson[1]. Eight 'geozones' are identified, as follows: Asia (South), Asia (East), Asia (West/central), Africa (North), Africa (West), Africa (East/Centra/Southern), European Union and Other Europe. In addition, the classification includes a category 'international/transnational' incorporating the major languages of Arabic, French, Portuguese and Spanish, which are spoken in many parts of the world, as well as 'other' languages.   Figure 2 shows this classification, highlighting languages which are spoken by more than 5000 pupils in London state schools.

We use postcode data for individual pupils in the Annual School Census to assign all records to 'Super Output Areas (SOAs, Vickers and Rees, 2007)' which themselves exist in varying levels of aggregation 'lower' with around 1500 residents and 'middle' consisting of around 7.5 k residents known respectively as LSOAs and MSOAs. This enables us to map languages by the home address of pupils, whereas Baker and Eversley were previously only able to calculate figures at Borough level based on the location of the school attended . For the London-wide maps that follow, we show data for London's 983 MSOAs. LSOA data is used in more detailed analysis that we have conducted for the London Borough of Newham.   We show numbers of pupils rather than percentages, given the unevenness of the data quality (denominators) across the boroughs. For each language, five categories are shown. It is important to note, therefore that the scales for the maps are different to one another. The primary purpose is to show the geographical distribution of each language, not to compare volumes.

Patterns or clustering of language

English speakers (including a small percentage classified as 'believed to be English', (<1%)) include over 60% of state pupils in London (663,584 in total). No other language is spoken by more than 5% of all pupils with a recorded language, and over forty languages are spoken by more than 1000 pupils – a picture of remarkable diversity.   The fifteen most prevalent languages other than English are shown in Table 1 (columns 2 and 3). Column 1 shows the ordering of languages in 1999 (according to Baker and Eversley).   The table indicates that while overall there has been little change in the relative importance of languages spoken in the nine years since 1999, there are some notable differences. Notably, Somali speakers have become more prevalent along with Tamil speakers which may well represent recent turmoil in their countries of origin. Polish and Albanian speakers appear in 2008 rankings for the first time in comparison to Greek, Cantonese and Creole speakers possibly reflecting recent expansion in EU membership.

---

[1] Antony Sanderson worked on the current project as a member of our Advisory Group. He was also a contributor to both editions of 'Multilingual Capital' and an advisor to the Department of Education and Skills on language classification.

Maps of individual languages demonstrate patterns of settlement and dispersal of minority communities and provide a basis for understanding how these patterns change between Censuses of Population.   Figure 3 clearly shows a doughnut pattern, with English speaking pupils found more in outer London (particularly in the East and South) than inner London. Minority language speakers tend to be concentrated in particular parts of the city. For example, Bengali speakers are heavily concentrated in Tower Hamlets, and Urdu speakers in three main areas: the neighbouring boroughs of Newham, Redbridge and Waltham Forest, Ealing/Hounslow and Merton/Wandsworth (Figure 4). However, note the much more dispersed distribution of Somali speakers (a similar size population overall to Urdu speakers) (Figure 5). Similar language maps can be found in the new edition of Multilingual Capital (Tinsley, Eversley, Mehmedbegović & Wiggins, 2010) and accessed via the LERU website (www.leru.org.uk)

Language and Ethnicity

While many languages 'attach' to particular ethnic groups, there are others (those that we have classified as 'international') for which knowing a person's language does not tell us about their country of origin or ethnic heritage.   Data on ethnicity, using the sixteen major ethnic categories used in the Census of Population, is also collected in the Annual School Census. This reveals that
57% of French speaking pupils are 'black' and a similar percentage of Arabic speakers are classified as 'other black' (15%), mixed (10%), white (9%) or Asian (8%). This suggests the need to analyse language and ethnicity data in these cases to understand the nuances of people's circumstances and needs. Notably these different populations have different geographical concentrations. White French speakers tend to reside in West London, Black French speakers in East London.

Correspondingly, language data can potentially offer a finer-grained understanding than has to date been available through the collection of ethnic categories.   Some ethnic groups are characterised by considerable linguistic homogeneity.   For example, 84% of pupils identified as Bangladeshi in London speak Bengali at home (with a further 12% categorized loosely as 'other than English' of which some will be Bengali speakers). 98% of White British and 95% of Black Caribbean children speak English at home However, other ethnic groups are very linguistically diverse, most notably 'Black African' and 'White Other'. 30% of Black Africans speak English at home, 20% Somali, 9% Yoruba, 6% Akan, 5% French, 2% Lingala, 2% Igbo and 2% Arabic. 179 other languages are spoken by fewer than 2% each of the London's Black African pupils. The main African languages spoken in London originate in different parts of the continent. Yoruba, Igbo and Akan are spoken mainly in West Africa, including Nigeria and Ghana. Lingala is spoken in Central Africa.   Among the 'other white' ethnic group, Turkish (14%) is the most common language, but 10% speak Polish, 8% Albanian or Shqip, 6% Portuguese, and 3% each Lithuanian, Greek and Spanish. 'Indian' is also a linguistically diverse category, with two major groups in Gujerati (29%) and Panjabi (22.6%), as well as Hindi, Urdu, Tamil and Malayalam speakers.   For these heterogeneous groups, the collection of data on language provides an opportunity for

finer grained understanding of who is living in London and their socio-economic circumstances, and how these are changing over time.

Preliminary Analysis of Language and Attainment

The usefulness of ethnic/language categories is demonstrated by a preliminary analysis of educational attainment data. Here, for simplicity, we concentrate only on results at Key Stage 2 (age 11). Using a sixteen category classification (DMAG (2003, 2005)) for ethnicity we see considerable differences between ethnic groups (Figure 6). Pupils of Chinese ethnicity are on average the highest attainers, with a median of 15.38 points. Black Caribbean, Black 'Other' and Black African pupils are the lowest attainers, with medians of 13.55,13.67 and 13.73 respectively.   Groups that are predominantly English-speaking appear throughout this distribution, from Black Caribbean at the bottom to White British and White Irish near the top.

Figure 6 highlights the wide distribution of scores within each ethnic group. The solid boxes in the chart show the 25th and 75th percentiles for each group, while the 'whiskers' show the range of attainment beyond this. Most groups have a gap of about 2.5 points between the 25th and 75th percentiles (slightly wider for the 'White Other' and 'Other' categories. However there are very high attaining pupils in all groups.

Figure 7 and Table 2 illuminate the spread of attainment within the two ethnic categories which had the greatest linguistic diversity – 'white other' and 'Black African'. Within 'White Other', five groups have particular low attainment. Median scores for Turkish, Portuguese, Lithuanian and Polish speakers (as well as people whose language is classified as 'other than English') would put them at the bottom of the distribution in Figure 6. While there are high attainers in these groups, there are also long tails of low achievement. By contrast, Italian, Greek and English speakers in the 'white other' ethnic category have few low attainers and median scores that place them close to the top of the overall distribution. Attainment patterns for White Greek speakers are similar to those of people who identify as having Indians ethnicity (a median of around 14.6 points).

The Black African category also contains a wide spread. Table 2 shows the three lowest attaining and three highest attaining Black African language groups, by comparison with some of the main ethnic groupings. Note that Lingala, French and Somali speakers tend to have very low attainment, well below that of the lowest attaining ethnic group overall (Black Caribbean). The attainment of Black African Igbo speakers is similar to that of White British students. These data suggest that some of the commonly used ethnic groupings may be too broad to be useful, and that language data can provide greater insight into which pupils may be in need of particular support.

Language, Ethnicity and Socio-Economic Circumstances

This data gives no indication that language itself is responsible for greater or lesser attainment. Most likely it points to the different socio-economic circumstances and

migration histories of people who have come to London at different times and from different parts of the world. The ASC data only contains limited fields of data on socio-economic circumstances namely, whether or not a pupil receives free school meals (FSM) and an index of deprivation describing the pupil's residential neighbourhood (IDACI). To enrich our understanding of the circumstances of different ethno-linguistic groups, we drew we are able to match and merge the ASC data on data for the London Borough of Newham with data for that Borough previously matched by our consultants, Mayhew Harper Associates Ltd (2009).

Schematically at a local authority level, data records are linked together via a property gazetteer (see Figure 8 below) using the General Practice registers as a base for reference. Addresses are cross-referenced and checked as to whom is present using various logical assumptions to include or exclude people.

Analysis of this data shows that, for Newham at least, there are marked differences in socio-economic circumstances within language and ethnic groups, which may well be driving attainment patterns (Table 3). On both poverty indicators (FSM and whether the family is in receipt of Council Tax Benefit (CTB)), Black African Somalis are by far the most disadvantaged group. They also have the highest proportion of single parent families and larger families. Other ethno-linguistic groups with large families tend to have low proportions of single parents, and vice versa. Ethnicity alone certainly gives a misleading picture. Yoruba speakers are relatively advantaged on these measures, in great contrast to Somalis. On the other hand, considering language alone would also be inadequate. White British English speakers in Newham appear more disadvantaged than Black Caribbean English speakers, and Pakistani Panjabi speakers more disadvantaged than Indian Panjabi speakers.

Next Steps

This findings paper has highlighted the potential of the ASC language data to help disaggregate Census ethnic categories and give greater insight into the geographic distribution and socio-economic circumstances of different ethno-linguistic communities. Annual analysis of the data could provide a vital inter-censal picture of settlement and migration, providing that data is consistently and accurately collected.

The ASC remains a state school exercise. In some parts of England and Wales and specifically London this is a significant gap. For example in Kensington and Chelsea less than 50% of children are believed to attend local state secondary schools. Some may attend state schools in neighbouring areas. This is a reason for seeking to do regional rather than local studies. The existence of specialist private schools for speakers of other languages such as the Lycee Francaise or the German School in South West London may lead to specific gaps in the data but in general the high percentage of children who do attend state schools makes the ASC an invaluable source of data.

The richer insight that can be gained by matching the ASC data to other local administrative data is clearly shown by Table 3. In practice, negotiating access to the data presents a major challenge both ethically and technically as well as the need to ensure data security and confidentiality. At least three agencies or providers are

involved, the DCFS, the Primary Care Trust and the local authority. As a result a fundamental component of our project has been to test the viability and value of such an exercise in the context of a single London Borough namely, Newham. This 'proof of concept' study will be reported separately.

Finally, in this paper we have concentrated on description only. A key question is whether the attainment patterns of different linguistic groups can be entirely explained by their socio-economic position, or whether language (in itself or as a marker of previous circumstances and experiences) has explanatory power in attainment. We will be exploring this further using more advanced statistical techniques and both KS2 and KS4 data.

References

Baker, P. and Eversley, J. (editors), (2000). Multilingual Capital: the languages of London's schoolchildren and their relevance to economic, social and educational policies. Battlebridge Publications, London.

BBC (2006) Translation costing public £100m BBC News 12 December 2006. http://news.bbc.co.uk/2/hi/uk_news/6174303.stm (accessed 25 June 2007)

Dalby, D. (1999). The Linguasphere Register of the world's languages and speech communities. Hebron (Wales). The Linguasphere Press [includes an index of more than 40,000 glossonyms, ethnonyms and toponyms prepared by Michael Mann]

Department for Education and Skills (2006a. School Census for maintained secondary schools, CTCs and academies in England Preparation and Guidance for 2007

Department for Education and Skills (2006b) Pupil Language Data Guidance for Local Authorities on schools' collection and recording of data on pupils' languages (in compliance with the Data Protection Act)

Data Management and Analysis Group (DMAG) Briefing. (2003/9). 2001 Census Key Statistics: Ethnicity, religion and country of birth. Greater London Authority.

Data Management and Analysis Group (DMAG) Briefing. (2005/39). London's changing population: diversity of a world city in the 21st century. Greater London Authority.

Everseley, J, Mehmedbegović, D, Sanderson, A, Tinsley, T, von Ahn, M and Wiggins, R.D. (2010). Language Capital: mapping the languages of London's schoolchildren. The National Centre for Languages (CILT) and The Institute of Education, London (in press).

Mayhew Harper Associates Ltd. (2009). Making Use of local administrative data for population estimates and service planning. Gordon Square Quantitative Methods Seminar, The Institute of Education, 27th March, 2009 (www.nkm.org.uk accessed 19th January 2010).

Martin C (2000) Foreword in Baker and Eversley (2000) op cit.

Mehmedbegović D (2007) 'Miss, who needs the languages of immigrants?': London's Multilingual Schools in Brighouse T and Fullick L, Education in a Global City: Essays from London. Institute of Education

Tosi A (1984) Immigration and Bilingual Education- a case study of movement of population, language change and education within Europe, Oxford: Pergamon Press.

Vickers, D. and Rees, P. (2007). Creating the UK National Statistics 2001 output area classification. J of Royal Statistical Society (A), 170(2): 379-403.

Tables

Table 1: The 'top 15' languages spoken other than English in London

Rank Order	Year 1999	Year 2008	Number 2008
1.	Bengali and Sylheti	Bengali	46681
2.	Panjabi	Urdu	29354
3.	Gujarati	Somali	27126
4.	Hindu/Urdu	Panjabi	20998
5.	Turkish	Gujarati	19572
6.	Arabic	Arabic	19378
7.	English based Creoles	Turkish	16778
8.	Yoruba	Tamil	16386
9.	Somali	Yoruba	13961
10.	Cantonese	French	13020
11.	Greek	Portuguese	11915
12.	Akan	Polish	10991
13.	Portuguese	Spanish	8647
14	French	Albanian/Shqip	8380
15.	Spanish	Akan	8117

Table 2: Lowest and Highest Attaining Linguistic Groups within the Black African Category contrasted with selected ethnic group scores, London 2008

Language/ Ethnic Group (by median score)	25th percentile	median	75th percentile
Black African- Lingala	10.56	12.58	13.78
Black African- French	11.18	13.01	14.27
Black African- Somali	11.40	13.02	14.31
Black Caribbean (average)	12.23	13.55	14.62
Black African (average)	12.25	13.73	14.85
Black African- English	12.84	14.13	15.21
Black African - Yoruba	13.02	14.19	15.13
Black African - Igbo	13.04	14.36	15.48
White British (average)	13.03	14.38	15.52

Table 3: Socio Economic Characteristics of Largest Ethno-linguistic groups in Newham (KS2 only)

	Pupils	3 or more children	Single parent	FSM	Council Tax Benefit
Black African English/Believed to be English	85	58%	21%	32%	38%
Black African Somali	88	89%	41%	91%	97%
Black African Yoruba	65	72%	15%	25%	26%
Black African Akan	47	68%	19%	30%	36%
Black African Other than English/unknown	118	80%	32%	59%	63%
Pakistani Urdu	203	77%	11%	27%	44%
Pakistani Panjabi	53	85%	15%	30%	53%
Bangladeshi Bengali	428	83%	7%	36%	64%
White British English	366	50%	39%	45%	61%
Indian Gujarati	122	56%	4%	15%	34%
Indian Panjabi	52	56%	8%	23%	31%
Black Caribbean English	175	45%	39%	27%	44%
Other Asian Tamil	54	61%	11%	9%	51%
Other Asian Tagalog Filiipino	40	50%	18%	5%	10%

Figures

Figure 1: English as first language by region

Figure 2: A Classification of Languages

Asian (South)	Asian (East)	Asian (W/C)	African (North)	Africa (West)	Africa (E/C/S)	European Union	European Other	Inter/trans national	Other	Unspeci-fied
Bengali	Chinese	Turkish	Somali	Yoruba	Lingala	Greek	Albanian / Shqip	Arabic		Carib-bean
Urdu	Viet-namese	Persian/ Farsi	Tigrinya	Akan/ Twi-Fante	Swahili/ Kiswahili	Italian	Russian	French		Oceania/ S/C
Panjabi	Japanese	Kurdish	Amharic	Igbo	Luganda	Dutch/ Flemish	Serbian/ Crotian/	Portu-guese		
Gujarati	Korean	Pashto/ Pakhto	Other	Other	Shona	German	Other	Spanish		
Tamil	Tagalog/ Filipino	Other			Other	Polish				
Hindi	Other					Lithuan-ian				
Malay-alam						Other				
Nepali										
Other										

Figure 3: Distribution of English Speaking Pupils in London

Count per MSOA
English/Believed to be English

- 1,090 to 1,922 (102)
- 891 to 1,090 (130)
- 714 to 891 (164)
- 529 to 714 (211)
- 32 to 529 (376)

© Crown Copyright
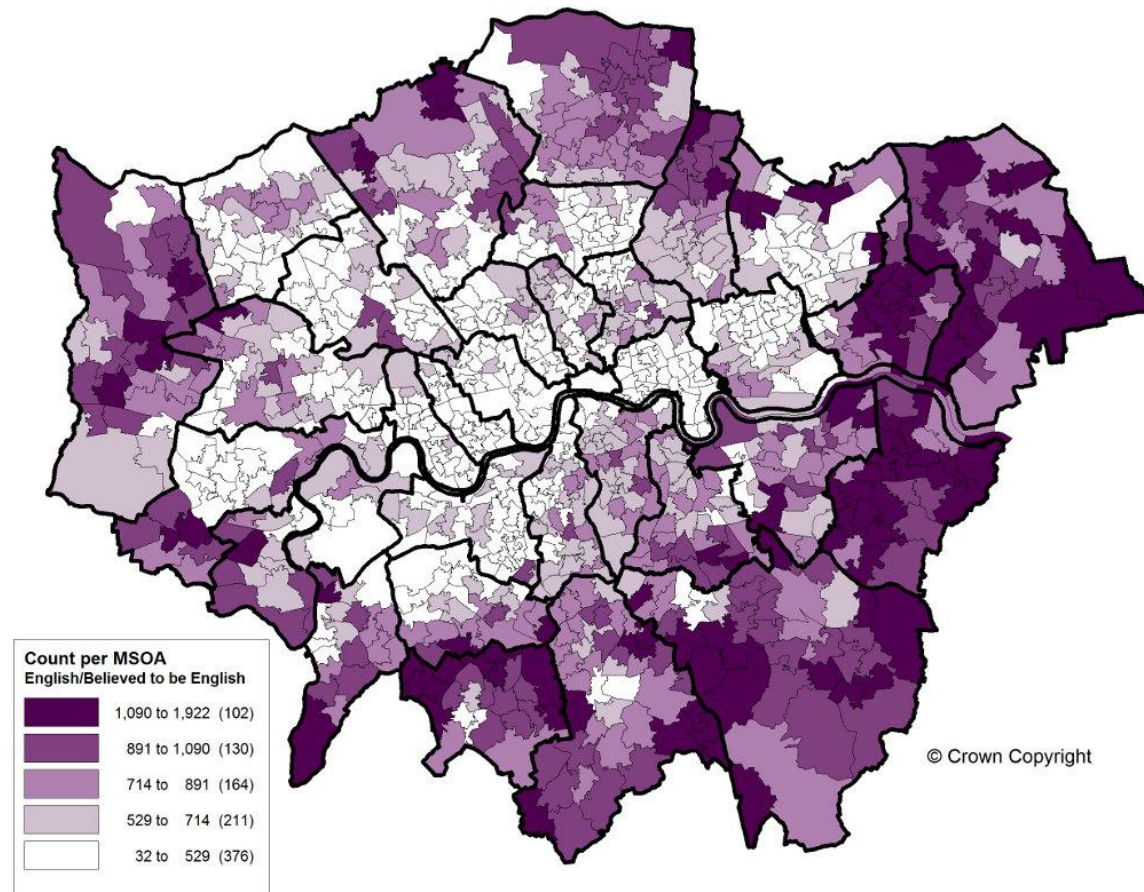
Figure 4: Distribution of Urdu-Speaking Pupils in London

Count per MSOA
Urdu

198 to 397 (24)
123 to 198 (37)
69 to 123 (63)
31 to 69 (128)
0 to 31 (731)

© Crown Copyright
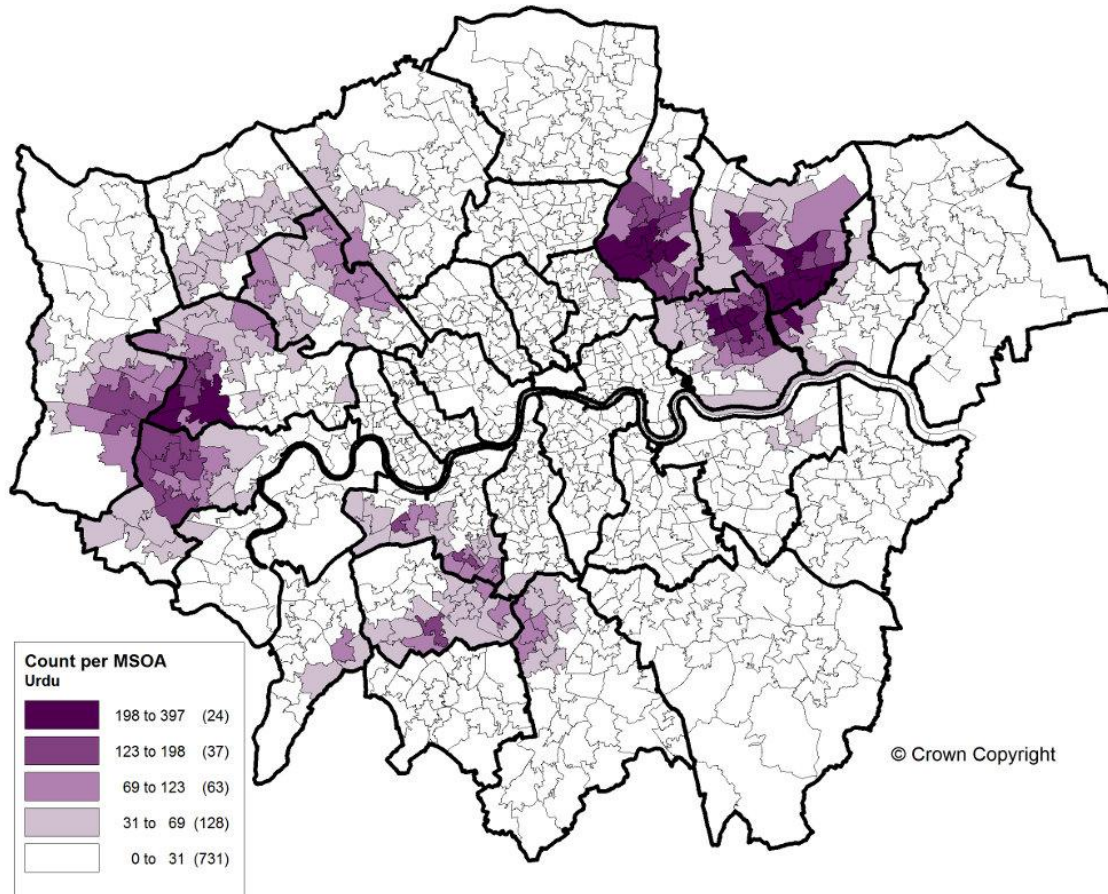
Figure 5: Distribution of Somali-Speaking Pupils in London

Count per MSOA
Somali

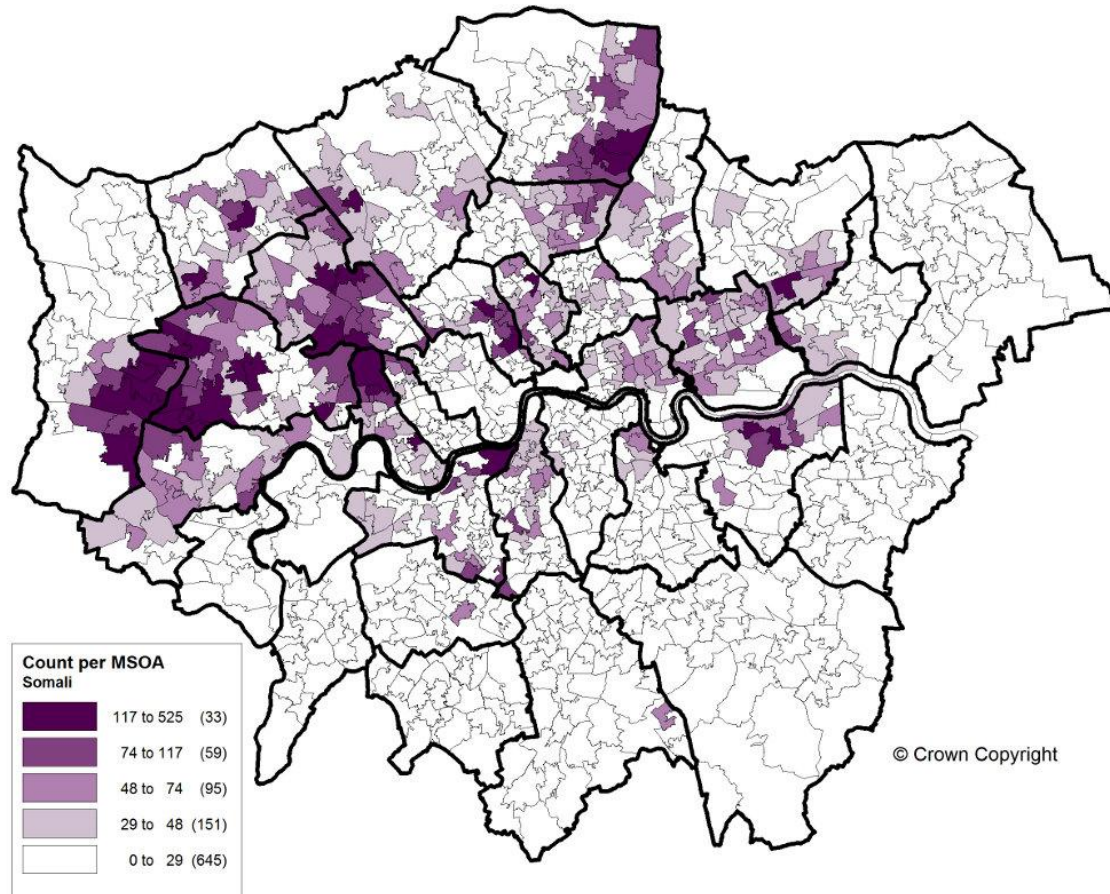117 to 525 (33)
74 to 117 (59)
48 to 74 (95)
29 to 48 (151)
0 to 29 (645)

© Crown Copyright

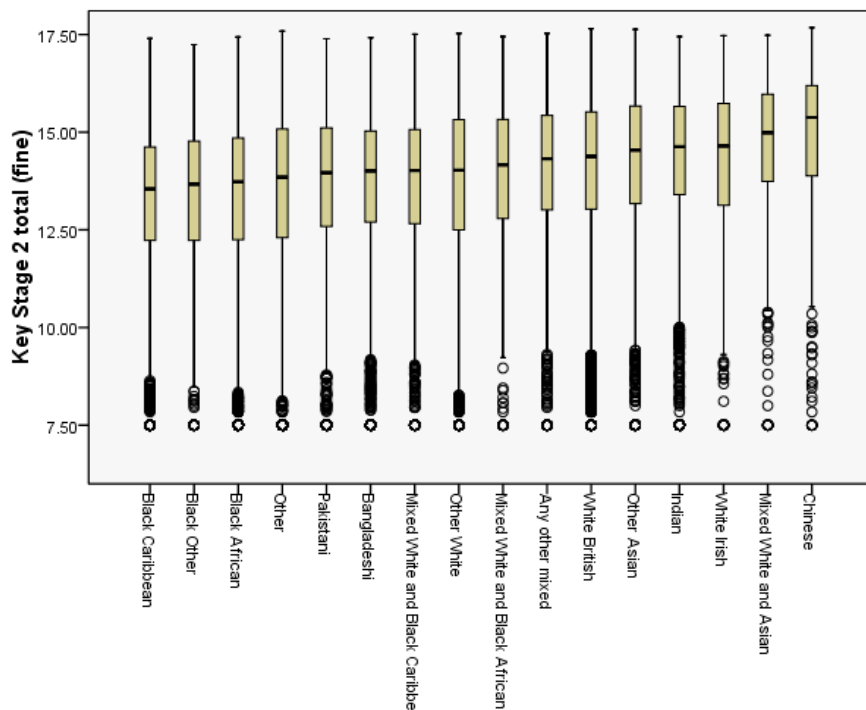Figure 6: KS2 total scores by ethnic category, London, 2008.

Figure 7: KS2 total scores by language category (10 largest), White Other, London, 2008.
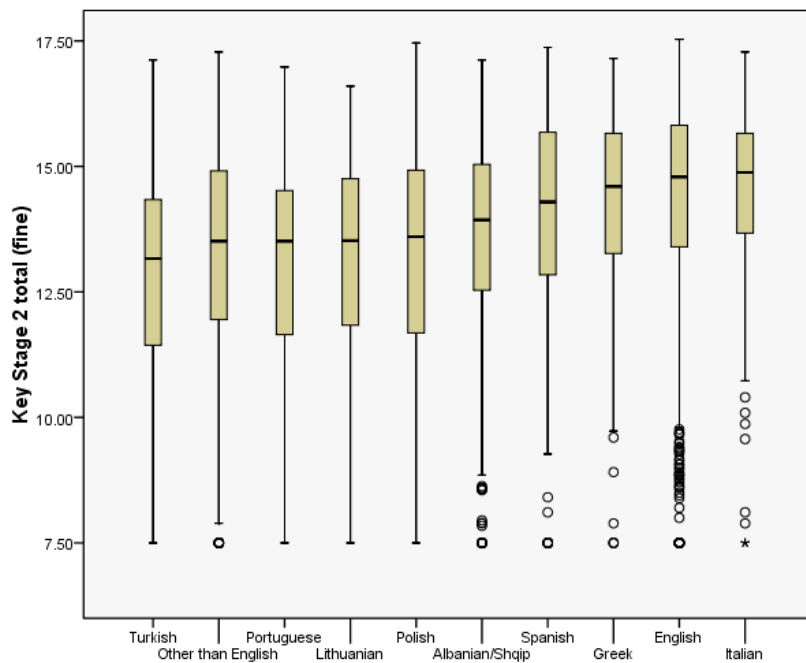
Figure 8: Data Matching