*Journal of Ethnic and Migration Studies*
*Vol. 35, No. 9, November 2009, pp. 1437–1460*

Routledge
Taylor & Francis Group

# Uncertainty in the Analysis of Ethnicity Classifications: Issues of Extent and Aggregation of Ethnic Groups

Pablo Mateos, Alex Singleton and Paul Longley

*Uncertainty is inherent in the conception and measurement of ethnicity, by both individuals themselves and those who seek to gather evidence of discrimination or inequalities in social and economic outcomes. These issues have received attention in the literature, yet rather little research has been carried out on the uncertainty subsequently created through the analysis of such measurements. We argue that, while general-purpose ethnicity classifications offer a method of standardising results, such groupings are inherently unstable, both in their upward aggregation and in their downward granulation. As such, the results of ethnicity analysis may possess no validity independent of the ethnicity classes upon which it is based. While this conclusion is intuitive, it nevertheless seems to pass unnoticed in the interpretation of research conducted in public policy applications such as education, health and residential segregation. In this paper we use examples based on the standard Census classification of ethnicity, alongside new rich ethnicity datasets from the education domain, in order to evaluate the sensitivity of results to the particular aggregation that is chosen. We use a case study to empirically illustrate the far-reaching consequences of this commonly overlooked source of uncertainty.*

*Keywords: Ethnicity Classifications; Uncertainty; Education; Aggregation*

## The Problem of Defining and Classifying Ethnicity

There has been a surge in population studies research on ethnicity over the last decade and a half (Bhopal 2007; Howard and Hopkins 2005), consistent with growing

Pablo Mateos, Alex Singleton and Paul Longley are respectively Lecturer in Human Geography, Research Officer in Spatial Literacy and Professor of Geographic Information Science at University College London. Correspondence to: Dr P. Mateos/Dr A. Singleton/Prof. P. Longley, Dept of Geography and Centre for Advanced Spatial Analysis, University College London, Gower Street, London WC1E 6BT, UK. E-mails: p.mateos@ucl.ac.uk; a.singleton@ucl.ac.uk; plongley@geog.ucl.ac.uk.

ISSN 1369-183X print/ISSN 1469-9451 online/09/091437-24 
DOI: 10.1080/13691830903125919

public interest in questions of group- and self-identity (Eriksen 2002), of shared origins and of migration in today's increasingly globally connected cities and societies (Castles and Miller 2003). This has been accompanied by some improvements in the collection, consistency and availability of population statistics by identity group, particularly ethnicity, especially following its widespread measurement in the round of censuses at the turn of the millennium (Morning 2008). However, although these efforts have helped to broaden our understanding of contemporary multicultural societies, effective use of statistics derived from such classifications requires a grasp of potentially far-reaching ontological and epistemological issues. There is consensus amongst many population statistics users that, 'while such [ethnicity and race] groupings may assist in making sense of individual outcomes, they are, of course, not fixed or preordained; rather they are social constructions imposed to order reality according to evolving ideas of human difference' (Ellis and Wright 2005: 15326). It is the problematic aspects of boundary setting associated with such 'evolving ideas of human difference' that lie at the core of the argument presented in this paper.

Identifying the defining characteristics of an ethnic group is often far from straightforward, both in conceptual and measurement terms, because ethnicity remains one of the most contested and unstable research concepts of social science (Nobles 2000). Ethnicity is socially constructed and relates to several dimensions of a person's identity—such as kinship, religion, language, shared territory, nationality or physical appearance (Bulmer 1996)—and as such is inherently contextual and likely to be transient. Such fluidity in the way individuals ascribe themselves to identity groups has been acknowledged ever since the Chicago School of the 1920s recognised that 'an individual may have many selves according to the groups to which he belongs and the extent to which each of these groups is isolated from others' (Park 1955[1921]: 181). Today the American Sociological Association describes race (in the US research context) as 'a social invention that changes as political, economic, and historical contexts change' (American Sociological Association 2002: 7). Thus, official ethnicity classifications and the way individuals ascribe themselves to one or more ethnic groups are highly contextual (to the time, place and purpose for which they were devised). They both depend on how those groups are perceived and come into contact with one another, in a particular society and point in time, as well as the priorities with which individuals identify with such constructed entities.

As a consequence of this renewed interest in ethnicity research, investigations have specifically addressed the problematic definition, classification and measurement of ethnicity in population studies and governmental statistics (Aspinall 2005; Bhopal 2004; Robbin 1999). Some researchers have questioned the actual need to subdivide populations according to ethnicity, or indeed any other commonly measured facets of identity such as race, nationality or indigenous group, consistent with an anti-essentialist critique (Brubaker 2004). A different type of critique has emerged from users of ethnicity statistics, who acknowledge the usefulness of classifications in the fight to reduce social inequalities, yet remain critical of the detail of the categories used in official ethnicity classifications. To cite some examples of this second critique,

many have argued for better definitions, labels and groupings, related to what it means to be 'White' (Bhopal and Donaldson 1998; Peach 2000), or 'Black' (Agyemang *et al.* 2005), who the 'Other' ethnic groups are (Connolly and Gardener 2005), or whether it is meaningful to use overarching groups such as 'Asian' (Aspinall 2003) or 'Hispanic' (Choi and Sakamoto 2005). Some of these demands have been at least partially accommodated by recent official ethnicity classifications and the quality of analysis is slowly being improved as a result.

Peter Aspinall (see the previous paper in this issue) has offered a comprehensive account of the problems associated with delimiting ethnic groups, the advances made in official ethnicity classifications and their future prospects, set in a policy-making context of accommodating the diversity and difference agendas. We will not repeat any of this debate in detail here, although our arguments should be interpreted in the context of Aspinall's paper. Central to his discussion of the best ways to improve ethnicity classifications is the trade-off between their validity and utility, crystallised through the concept of the *granularity* of any classifications. The number and detail of identifiable groups in a classification (its granularity) tends to grow over time, often in response to increasing public demands to acknowledge smaller groups. This augments the classification's validity, but also diminishes its utility, because of the difficulty in interpreting and comparing statistics that cross-classify a myriad of groups according to multiple dimensions of identity. This 'validity–utility' trade-off is likely to become more crucial in the near future, since the granularity of ethnicity classifications is increasing (the US 2000 Census allowed respondents to tick more than one race category, and in the UK the number of ethnic groups will grow from eight in 1991, through 16 in 2001, to a likely 19 in 2011), and additional questions on identity are being introduced in many countries' censuses and population statistics. New questions on religion, language, ancestry, nationality, migrant status, sexual preference, disability and other defining dimensions of minority groups that conceivably experience social inequality are found in today's population statistics and plans for future censuses.

Such movement towards increasingly complex measures of self-identity and diversity, manifest through greater granularity in ethnicity classifications and higher dimensionality in identity measurement, brings tremendous new research potential as well as new challenges. Amongst the former, it facilitates highly disaggregated research into the causes and consequences of inequalities by studying different combinations of fine-level classifications according to different but possibly interrelated dimensions of identity. Yet this in turn creates new methodological problems arising from the different ways in which these fine groups are cross-classified, aggregated, analysed and interpreted. This paper addresses some of these methodological challenges, in relation to the overarching concept of uncertainty in the classification of ethnicity. Our aim is to begin to discern the different sources of such uncertainty, which introduces vagueness or ambiguity in different stages of the way we interpret reality (Longley *et al.* 2005); how the real world is conceived (conception), how such conceptions are represented and measured (representation

and measurement), and finally, how those measured representations of reality are analysed (analysis). In the study of ethnicity, uncertainty in conception pertains to the ontological problem of defining ethnicity and how people perceive it; uncertainty in representation and measurement relates to the ambiguities that arise in creating ethnicity classifications and the way individuals are ascribed to them; and uncertainty in the analysis of those measurements refers to issues of the extent and aggregation of the selected ethnic classes. Our view is that, while issues of uncertainty in the conception and representation/measurement of ethnicity have received attention in the literature (e.g. Bhopal and Donaldson 1998; Rankin and Bhopal 1999; Robbin 1999), there has been relatively little focus upon the uncertainties arising from the kinds of decisions frequently made in the analysis of ethnicity classifications. We propose to contribute to such a debate by focusing on the uncertainties arising from the analysis of ethnicity classifications.

The rest of this paper is structured into three sections. The next one reviews the sources of uncertainty in the conception, representation and measurement of ethnicity, while the succeeding section develops some of the consequences of such uncertainty for analysis—specifically with regard to different aggregations. The third section develops a case study from the education domains, and demonstrates the existence of uncertainty in aggregation of fine ethnic groups. Finally, we offer some concluding remarks. More generally, we see the empirical analysis in this paper as contributing towards an understanding of the sources of uncertainty in the study of ethnicity analysis and the ways in which these uncertainties operate. We also see our work as key to the development of a flexible approach to ethnicity classification that is fit for purpose, with particular reference to public sector applications. We believe this to be a very significant research agenda, and what we present here is inevitably only an initial foray into issues of uncertainty in the conception, representation, measurement and analysis of ethnicity.

## Uncertainty in the Conception, Representation and Measurement of Ethnicity

Ethnicity is a social construction of seemingly indeterminant complexity, which defies watertight conception. As with other complex phenomena, humans make sense of ethnicity through simplifications, devised in the human mind and defined and represented through some sort of social consensus. This entails classification, which in turn implies the creation and use of appropriate measures for recording difference between classes.

One key area of application in which the uncertainties in the conception, representation and measurement of ethnicity are thrown into sharp relief is in public health research: specifically, in relation to the study of inequalities in health outcomes according to ethnic group, using established epidemiological methods. Research findings typically document stark inequalities in the health outcomes of different ethnic groups, and these are often viewed alongside different socio-economic, environmental, demographic, genetic, lifestyle, cultural and discrimination factors

(for reviews see Bhopal 2007; Nazroo 2003). However, a common auto-critique in this field is that the associative—hardly ever causal—explanations of observed health inequalities are highly dependent upon the bounding criteria of ethnicity categories and the methods used to allocate individuals into such groups—even to the point of questioning the validity of ethnicity as an organising concept in epidemiological research (Senior and Bhopal 1994). As a consequence, results from many different studies cannot be compared on equal terms, frustrating the demonstration of the common factors behind health inequalities.

Comstock *et al.* (2004) summarise very well the extent of the problem of comparability in ethnicity research in public health. They conducted a comprehensive review of 1,198 articles published in the *American Journal of Epidemiology* and the *American Journal of Public Health* from 1996 to 1999, and identified 219 different terms used to describe ethnic or racial groups in the US, which they struggle to group into eight core 'ethnic groups', dealing for example with 16 different ways of describing 'Black', 32 different ways of describing 'White' and 46 different ways of describing 'Hispanic. The fuzziness, incompatibility and degree of overlap between terms was very great, even though this large collection of articles was drawn from just two journals of the same scientific discipline in a single country in which research on ethnic and racial disparities has a long tradition. This issue poses a crucial problem that requires 'continued professional commitment [. . .] to ensure the scientific integrity of race and ethnicity as variables' (Comstock *et al.* 2004: 611). This problem of lack of standard definitions of ethnic groups has also been identified by other authors, and seen as an ontological problem that constitutes 'a problem with basics' (Bhopal 2004: 441). Important contributions along these lines by Peter Aspinall (2002, 2005, 2007), Raj Bhopal (2004, 2007), and other authors suggest that researchers in health and ethnicity should use comparable ethnicity classifications that explicitly define: the categorisations adopted; their context of use; the criteria used to justify their adoption; the method used to ascribe ethnicity to individuals; and also to provide precise explanations of observed differences in health outcomes according to ethnic group. According to these authors, most of these criteria currently remain unfulfilled in health and ethnicity research.

There is consensus in the research literature that, until researchers define consistent and comparable building blocks for ethnicity classification and develop a common method of ascribing individuals to these classes, the results of different studies cannot be generalised beyond the specific context of their respective research studies. In other words, their results are not independent of the definition of the ethnic groups that underpin the analysis that they develop. It is clear that the different interpretations of ethnicity listed in Comstock *et al.* (2004) will lead to substantially different results in terms of apparent population characteristics and group attributes, such as socio-economic status, education, employment or health, and even more worryingly, genetic characteristics. Despite this, the sources of uncertainty in the conception, representation and measurement of ethnicity are frequently not discussed at all, or where they are, scenarios as to how these uncertainties might operate are not

identified. Moreover, distinctions between the sources of uncertainty in classification and the ways in which they operate in the analysis of ethnicity classifications are not drawn.

## Uncertainty in the Analysis of Ethnicity Classifications

One pragmatic means of apparently circumventing the problem of uncertainty in the conception, representation and measurement of ethnicity lies in adherence to one of the small number of officially sanctioned ethnicity classifications that are widely used to code large public sector surveys, such as the Census ethnicity classification in the UK (Office for National Statistics 2003). This pragmatic, but usually uncritical, response addresses the problem of comparability between definitions of ethnic groups and the lack of commonality of methods used across studies. As such, it is widely used in health (Department of Health 2005) and education applications (Department for Education and Skills 2006). Greater adherence to a common classification should obviate the problem of having different boundaries between the ethnic groups used in different studies. Yet in practice, the use of these 'stable' official ethnicity classifications does not come without problems. Our contention is that, even when stable official ethnicity classifications and methods are used, the results of any analysis are still inherently uncertain. Official classifications also provide few clues as to how ethnic groups should be compared over time or between data sources (Platt *et al.* 2005), and more importantly how detailed ethnic categories should be aggregated into the coarser groups that might be appropriate to particular applications. Such factors introduce uncertainty in the *analysis* of ethnicity and are related to issues of the extent (size and number) and aggregation of groups. Three brief examples drawn from the UK Census of Population will be used here to illustrate this view.

The UK Census has collected data on ethnicity in 1991 and 2001, albeit using slightly different classifications of ethnicity and different levels of disaggregation (eight groups in 1991 and 16 in 2001). Therefore, the main problem in comparing the two datasets over time concerns how to match the two classifications and render them comparable. One of the major differences between the two classifications lies in the four 'Mixed' ethnicity categories (White and Black Caribbean, White and Black African, White and Asian, and Any other Mixed) which were first included in the 2001 Census. Three different approaches have been adopted in the literature to reallocate them into 1991 comparable categories: a group of researchers with representatives from the Office for National Statistics (ONS) recommends allocating them all into an expanded 'Other' macro-group (Bosveld *et al.* 2006; Platt *et al.* 2005; Simpson and Akinwale 2007); the Greater London Authority (GLA) advocates allocating them into their most closely related 'pan-ethnic' groups (Bains and Klodawski 2006); and researchers at the University of Leeds suggest splitting them into each of 'their parents' alleged ethnic groups' (Rees and Butt 2004). Table 1 shows the allocations of 2001 Census ethnicity groups into the 1991 categories proposed by

**Table 1.** Comparison of three approaches to allocate 2001 Census ethnic groups to 1991 categories

| 1991 Census ethnic group | Allocated 2001 Census ethnic groups | | |
|---|---|---|---|
| | ONS | GLA | Leeds |
| White | White British<br>White Irish<br>White Other | White British<br>White Irish<br>White Other | White British<br>White Irish<br>White Other<br>0.5*Mixed: White & Black Caribbean<br>0.5*Mixed: White & Black African<br>0.5*Mixed: White & Asian |
| Black Caribbean | Black Caribbean | Black Caribbean | Black Caribbean<br>0.5*Mixed: White & Black Caribbean |
| Black African | Black African | Black African | Black African<br>0.5*Mixed: White & Black African |
| Black Other | | Black Other<br>Mixed: White & Black Caribbean<br>Mixed: White & Black African | Black Other |
| Indian | Indian | Indian | Indian<br>0.5*Mixed: White & Asian *Proportion Indian |
| Pakistani | Pakistani | Pakistani | Pakistani<br>0.5*Mixed: White & Asian *Proportion Pakistani |
| Bangladeshi | Bangladeshi | Bangladeshi | Bangladeshi<br>0.5*Mixed: White & Asian *Proportion Bangladeshi |
| Chinese | Chinese | Chinese | Chinese |
| Other Asian (*) | | Asian Other<br>Mixed: White & Asian | Asian Other |
| Other groups | Any Other Ethnic Group | Any Other Ethnic Group | Any Other Ethnic Group |
| | Black Other<br>Asian Other<br>Mixed: White & Black Caribbean<br>Mixed: White & Black African<br>Mixed: White & Asian<br>Any Other Mixed | Any Other Mixed | Any Other Mixed |

(*) Other Asian was not included in the 1991 Census pre-set questions, although counts for write-in answer 'Other Asian' were reported by ONS and therefore this term has been adopted by two of the studies as a 1991 ethnic group.

*Source*: Three approaches: ONS, GLA and Leeds, compiled respectively from Platt *et al.* (2005: 44) and Bosveld *et al.* (2006: 30); Bains and Klodawski (2006: 4); Rees and Butt (2004: 176); see text for more details.

each of these three approaches, abbreviated as ONS, GLA and Leeds respectively. It is clear that these three different approaches will lead to substantially different results when comparing 1991 with 2001 ethnicities, in terms of the sizes and characteristics of the resulting groups.

In order to illustrate the uncertainty derived, we compared the division of the 1991 Census population according to ethnic group with that for the 2001 Census using the three approaches outlined above for two metropolitan areas in England (Greater London and West Midlands). We calculated the apparent crude population growth or decline according to ethnic group for each of the approaches, making no adjustments for changes in output area boundaries or definitions of resident population. The results of this comparison are presented in Table 2, which shows substantial differences between the growth rates calculated by each approach for groups such as Black Caribbean (ranging from 18.1 to 30.3 per cent in London and 5.8 to 27.7 per cent in the West Midlands), Black Other ($-100$ to 172.5 per cent in the West Midlands), and White ($-4.3$ to $-2.8$ per cent in London and $-6.2$ to $-5.1$ per cent in the West Midlands). The interpretation of these different rates can lead to conflicting claims concerning the temporal processes of migration by different minority groups. For example, the so-called phenomenon of 'White flight' could be pictured differently just by changing the basis of comparison of the two census ethnicity classifications, as manifest in two very different rates of apparent decline of 'White' group populations in London ($-4.3$ per cent or $-2.8$ per cent).

Another issue with the UK 2001 Census ethnicity classification is the way in which the Census agencies re-allocate the write-in answers collected in each of the five 'Other' categories (Other White, Other Black, Other Asian, Other Mixed, and Any Other ethnic group). This is achieved through a process in which the individual write-in answers are re-assigned to those categories which the Census agencies deem to provide the 'closest match' for the purpose of producing Census outputs using the official 16 ethnicity categories. A list of how the ONS does this for England and Wales in the 2001 Census (Office for National Statistics 2003) is shown in Table 3. No documentation is available as to how the ONS decided where to draw the line between groups identified as White (e.g. Cypriot) or otherwise (e.g. Afghani), groups identified as British (e.g. Cornish) or neither British nor Irish (e.g. Northern Irish, reclassified as White Other); and how they defined who is considered 'Asian', in essence intimating an association with the Indian Subcontinent (e.g. East African Asian, Sri Lankan, Tamil, Sinhalese, Caribbean Asian, British Asian, or Nepalese) versus the other parts of Asia that are left in the 'Other' category (e.g. Japanese, Vietnamese, Filipino, Malaysian, or Burmese). Knowing how these re-allocations of Census responses were made is crucial when interpreting the ethnicity counts in some areas of London, where 923,003 people (some 13 per cent of the total 2001 Census respondents or 30 per cent of the non-White British respondents) provided a write-in answer to the ethnicity question that did not match any of the 16 pre-set categories (source: Census commissioned table C0183—Ethnicity). Beyond the 2001 Census, many public sector agencies, such as the Department of Health and the Department

**Table 2.** Comparison of three approaches to measure 1991–2001 population growth by ethnic group in London and the West Midlands

| | London | | | | West Midlands | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1991–2001 population growth (%) | | | | 1991–2001 population growth (%) | | |
| Comparable 91–01 ethnic groups | 1991 population | ONS | GLA | Leeds | 1991 population | ONS | GLA | Leeds |
| White | 5,333,580 | −4.3 | −4.3 | −2.8 | 2,178,149 | −6.2 | −6.2 | −5.1 |
| Black Caribbean | 290,968 | 18.1 | 18.1 | 30.3 | 72,183 | 5.8 | 5.8 | 27.7 |
| Black African | 163,635 | 131.6 | 131.6 | 142.0 | 4,116 | 143.0 | 143.0 | 172.7 |
| Black Other | 80,613 | −100.0 | 105.3 | −25.1 | 15,716 | −100.0 | 172.5 | −43.7 |
| Indian | 347,091 | 25.9 | 25.9 | 31.0 | 141,359 | 11.1 | 11.1 | 13.3 |
| Pakistani | 87,816 | 62.6 | 62.6 | 69.2 | 88,268 | 56.3 | 56.3 | 59.4 |
| Bangladeshi | 85,738 | 79.5 | 79.5 | 86.8 | 18,074 | 60.9 | 60.9 | 64.1 |
| Chinese | 56,579 | 41.8 | 41.8 | 41.8 | 6,107 | 72.7 | 72.7 | 72.7 |
| Other Asian | 112,807 | −100.0 | 71.1 | 18.0 | 8,852 | −100.0 | 241.5 | 98.7 |
| Other Groups | 120,872 | 340.6 | 44.0 | 44.0 | 18,847 | 384.3 | −3.3 | −3.3 |
| Total | 6,679,699 | 7.4 | 7.4 | 7.4 | 2,551,671 | 0.2 | 0.2 | 0.2 |

*Note*: The Table shows 1991 Census total population per comparable ethnic group in Greater London and West Midlands county, followed by the population growth for the period 1991-2001 according to the three approaches described in Table 1.
*Source*: 1991 and 2001 Census Key Statistics table for England. The three approaches to allocate 2001 ethnic categories to 1991 ethnic groups, termed ONS, GLA and Leeds, are described in Table 1.

**Table 3.** UK 2001 census write-in answers and their re-allocated ethnic categories by ONS

| 2001 census write-in answer | Re-allocated 2001 census ethnic category |
|---|---|
| English, Scottish, Welsh, Cornish | White British |
| Northern Irish, Cypriot, Gypsy/Romany, Former USSR, Baltic States, Former Yugoslavia, Other European, White South African, American, Australian, New Zealander, Mixed White | Other White |
| British Indian, Punjabi | Indian |
| British Pakistani, Kashmiri | Pakistani |
| British Bangladeshi | Bangladeshi |
| British Asian, East African Asian, Sri Lankan, Tamil, Sinhalese, Caribbean Asian, British Asian, Nepalese, Mixed Asian (i.e. mixture of descriptions in the Asian section) | Other Asian |
| Caribbean and West Indian islands (and also Guyana) apart from Puerto Rican, Dominican and Cuban, which are Latin American | Black Caribbean |
| Nigerian, Somali, Kenyan, Black South African, Other Black African countries | Black African |
| Black British, Black American, Mixed Black | Other Black |
| Hong Kong | Chinese |
| Japanese, Vietnamese, Filipino, Malaysian, Aborigine, Afghani, Burmese, Fijian, Inuit, Maori, Native American Indian, Thai, Tongan, Samoan | Other Ethnic Group |
| Arab, Buddhist, Hindu, Iranian, Israeli, Jewish, Kurdish, Latin American (eg. Cuban, Puerto Rican, Dominican, Hispanic), Moroccan, multi-ethnic islands (e.g. Mauritian, Seychellois, Maldivian, St Helena), Muslim, Other Middle Eastern (eg. Iraqi, Lebanese, Yemeni), Other North African, Sikh, South American (includes Central American) | *Considered 'difficult to allocate answers' and left in the same 'Other' group where they were written in* |

*Source*: Office for National Statistics (2003: 53–4).

for Children, Schools and Families (DCSF), collect ethnicity data at much finer granularity than the 16 Census categories (for example 95 ethnic categories in the Pupil Level Annual School Census—PLASC), but there is a requirement that these are then mapped back to the official ethnicity categories for comparability purposes (Department of Health 2005).

Ensuring that all the write-in (free-text) responses in such datasets are then coded into the list of very fine ethnic categories in a consistent way at local level (e.g. across schools and hospitals), and over time, is much more difficult than just using a self-identification method with the Census 16 categories. This could be because of different interpretations of the ethnicity coding instructions by people with different levels of training and expertise, as well as knowledge of the ethnic groups themselves and the person being coded (e.g. pupil or patient). This is especially crucial for groups that do not fall neatly within Census categories, or that may be prone to local interpretation of their type of 'otherness'. Examples of these in-between groups

collected by the write-in answers to the Census include: Kosovan, Albanian, Moroccan, North African, Kurdish, Arab, Turkish, Turkish Cypriot, Iranian, Middle Eastern, Israeli and Latin American, whose write-in answers in London included 171,419 people or 2.4 per cent of the total population in the 2001 Census (source: Census commissioned table C0183—Ethnicity). The same issue applies to other groups that appear in the last row of Table 1, and is closely related to the aforementioned contextual nature of the groupings of ethnicity that might serve different purposes in different situations.

Finally, use of official Census classifications poses a further problem of aggregation of fine ethnic categories. When ethnic group counts are broken down by other population characteristics such as age, sex or occupation as well as by small area geography, the intersections of these cross-tabulations tend to produce very small people-counts that may create a risk of disclosure of information ascribable to individuals. In order to prevent this risk when supplying commissioned tables or anonymised records, the UK Census agencies frequently aggregate the 16 ethnic categories into macro-groups, also called pan-ethnic groups. There are typically five of these: White, Mixed, Asian, Black and Other. Moreover, many researchers face problems of data quality when using ethnicity data derived from transactional databases, such as in Hospital Admissions where the mixing together of both 1991 and 2001 categories is still common (London Health Observatory 2005). Researchers are then frequently obliged to adopt different types of data aggregation using some macro-group classification of ethnicity. It is striking to see how these operational issues result in much of today's ethnicity research in social science and health falling back upon the all-embracing and crude categorisation of White, Asian and Black in the UK, and White, Black and Hispanic in the US.

Through the examples offered so far, it is hoped that we have cast some doubt upon the stability of research results that rely upon the commonly accepted ethnicity classifications used in official statistics. It should be clear from these examples that such results are highly sensitive to changes in the number and sizes of ethnic groups used, and the ways in which they are aggregated. Official ethnicity classifications and the statistics that employ them are not exempt from these problems either, despite the fact that such classifications are usually taken for granted and perceived by many as socially accepted and fixed. What follows is a preliminary investigation into these issues of uncertainty in the analysis of ethnicity classifications, and their impact in the variability of results using an innovative ethnicity classification—the Pupil Level Annual School Census (PLASC) expanded ethnicity classification.

## Uncertainty in Ascertaining Educational Attainment by Ethnic Group

Maintained schools and colleges receiving public funding in England have had a statutory duty to supply pupil data to the Department for Children, Schools and Families (DCSF) on an annual cycle since 2002 (Jones and Elias 2006). These data are stored at the DCSF in the National Pupil Database (NPD) and the Pupil Level Annual

School Census (PLASC). The NPD dataset links aspects of individual student educational histories (e.g. attainment) at different stages of progression through the school system in England. PLASC is a unique survey of all students in publicly funded schools and captures a range of demographic data including variables such as: postcode, ethnicity, free school meals eligibility, disability status, language spoken at home, etc. Moreover, although the published PLASC data constitute the most granulated large-scale database in Britain, an expanded version of the dataset presents some 95 different ethnic categories. We have negotiated special access requirements to the 'expanded ethnicity categories' in PLASC through the PLASC/NPD User Group at the Centre for Market and Public Organisation (CMPO) at the University of Bristol, as these are designated sensitive data by DCSF.[1]

Recording of PLASC using the expanded typology of 95 detailed ethnicity categories has been mandatory for data collection by schools since 2003 (Godfrey 2004). This extensive typology of detailed ethnic groups, which is listed in Table 4, facilitates very detailed analysis of ethnicity factors influencing education attainment, since the 95 categories can be flexibly aggregated in different ways. The DCSF guidelines stipulate that these categories should be always 'nested back' into the 16 2001 Census ethnicity categories (Department for Education and Skills 2006) but the raw data also present enormous potential for investigating the effect of aggregations following different dimensions of ethnicity that one might want to analyse. In this exercise we have compared three different aggregations of the PLASC ethnicity categories: the 'PLASC Main Group', which comprises the official PLASC groupings based on the 16 2001 Census ethnicity categories plus two non-response categories (a total of 18 categories); and two different aggregations of 18 ethnic groups termed 'Grouping A' and 'Grouping B'. These different aggregations have been built by classifying the 95 expanded ethnicity categories into 18 groups. These have been arranged in ways constrained only by a (consciously subjective) understanding of 'closeness' between groups (along one or more of the dimensions of ethnicity) and a maverick objective to maximise the number of categories that fall into different groupings in each of the three types of aggregation produced. The final 18 categories defined for each of the three alternative aggregations are listed in Table 5, and the look-up table between each of the 95 expanded ethnicity categories and the three aggregations is presented in Table 4 (in the three columns labelled 'PLASC Main Group', 'Grouping A', and 'Grouping B').

A measure of educational attainment and a proxy for socio-economic status were calculated for each ethnicity category in each of the three aggregations. The former measure entailed calculating the average 'capped' result for GCSE exams for pupils in each ethnicity category. This is a measure established by the Department for Children, Schools and Families (DCSF) that sums the cumulative results of the pupil's best eight GCSE scores (Department for Children, Schools and Families 2006), and thus is independent of the number of GCSE subjects taken. With respect to the measure of socio-economic status, the percentage of pupils per ethnicity category

who were eligible for free school meals was calculated, since this is the most widely used income indicator in the literature (Sammons 1995).

The results of these two measures are listed in Table 4 for the original 95 ethnic groups, and in Table 5 for the 18 ethnicity categories in each of the three alternative groupings ('PLASC Main Group', 'A' and 'B'). Since the number of ethnicity categories remains constant (i.e. 18) across the three alternative aggregations, differences in the results of the analysis could be explained by the uncertainties inherent in aggregation, as opposed to changes in the extent of analysis (number and sizes of groups). As expected, the ethnicity aggregations that are substantially different in nature between the three groupings produce different results, which are summarised in Table 3. Three examples will be cited here. While the educational attainment of 'Any Other White' ranks number 9 in 'PLASC Main Group', in Grouping A the 'Western European' ranks 4, 'White Other' ranks 6, but 'Eastern European' ranks 15, while in Grouping B 'White Other' ranks 10 and 'White European' ranks 11. Therefore, this example shows striking differences in the educational attainment of 'White Other' groups, depending upon the definition of who is deemed to be White, European, or falling within Eastern or Western Europe. This can be seen in the assignments made in Table 4. While the broad category of 'Black African' ranks 14 in 'PLASC Main Group', when it is broken down in Grouping A it ranges from as high as rank 5 for 'Nigerian & Ghanaian', rank 12 for 'Black African Other', to rank 18 for 'Somali', while in Grouping B, it is again divided between rank 7 for 'Western African' and 17 for 'Black African Other'. Finally, when categories do not neatly nest into each other, as opposed to the previous two examples, the uncertainty generated by the different ways in which ethnicity categories can be aggregated into classes are manifest even more clearly. The different definitions underpinning the 'Other Ethnic Group' class in each of the three groupings result in its rank moving from 12 to 9 and to 16 respectively in the 'PLASC Main Group', Grouping A and Grouping B.

Similar differences are found in the percentage of pupils eligible for free school meals shown in Table 5. In the PLASC Main Group the highest percentage is in the Bangladeshi group (54.6), followed by Black African (36.9), Pakistani (35.8) and 'Any Other Ethnic Group' (34.5). However, in 'Grouping A' the group with the highest percentage is Somali (with an extreme value of 82.3), followed by 'Middle Eastern' (48.7) and Pakistani-Kashmiri-Bangladeshi (41.3), with all the other classes having values of less than 30 per cent. In 'Grouping B' these are Bangladeshi (54.6), 'Middle East & North Africa' (54.5), 'Black African Other' (39.3) and Kashimiri (38.5). It follows that separate analyses framed using the three different aggregations of the PLASC expanded ethnicity categories suggest very different conclusions about the level of educational attainment and experiences of economic deprivation according to ethnic group.

A further analytical test on the data presented in Table 5 was performed to identify whether there is any significant relationship between the average capped GCSE score and the percentage of pupils eligible for free school meals within each ethnic group;

**Table 4.** List of PLASC expanded ethnicity categories and the three aggregations used in this paper

| PLASC expanded category | PLASC main group | Grp A | Grp B | Total pupils | Avg. capped GCSE | % eligible FSM |
|---|---|---|---|---|---|---|
| White British | WB | BI | EOB | 428,809 | 347.2 | 10.8 |
| White English | WB | BI | EOB | 51,130 | 356.7 | 10.2 |
| White Scottish | WB | BI | CEL | 210 | 379.9 | 6.2 |
| White Welsh | WB | BI | CEL | 149 | 403.3 | 8.1 |
| White Cornish | WB | BI | EOB | 1,161 | 371.7 | 8.7 |
| Other White British | WB | BI | EOB | 5,211 | 343.0 | 8.7 |
| White Irish | WI | BI | CEL | 2,182 | 358.9 | 17.5 |
| Traveller Of Irish Heritage | WI | BI | CEL | 128 | 163.7 | 42.2 |
| Any Other White Background | AOW | WO | WO | 4,963 | 358.3 | 10.6 |
| Albanian | AOW | EE | WE | 68 | 262.2 | 58.8 |
| Bosnian-Herzegovinian | AOW | EE | WE | 24 | 355.0 | 33.3 |
| Croatian | AOW | EE | WE | 15 | 240.6 | 40.0 |
| Greek/Greek Cypriot | AOW | WSE | WE | 126 | 390.7 | 9.5 |
| Greek | AOW | WSE | WE | 89 | 375.8 | 14.6 |
| Greek Cypriot | AOW | WSE | WE | 272 | 359.9 | 14.0 |
| Italian | AOW | WSE | WE | 223 | 370.0 | 6.3 |
| Kosovan | AOW | EE | WE | 142 | 301.7 | 63.4 |
| Portugese | AOW | WSE | WE | 151 | 255.7 | 33.8 |
| Serbian | AOW | EE | WE | 7 | 393.7 | 0.0 |
| Turkish/Turkish Cypriot | AOW | ME | WE | 316 | 324.7 | 33.9 |
| Turkish | AOW | ME | WE | 563 | 296.6 | 49.7 |
| Turkish Cypriot | AOW | ME | WE | 240 | 303.5 | 32.9 |
| White European | AOW | WO | WE | 1,218 | 366.4 | 10.6 |
| White Western European | AOW | WSE | WE | 770 | 379.0 | 13.6 |
| White Eastern European | AOW | EE | WE | 670 | 320.8 | 20.1 |
| Other White | AOW | WO | WO | 2,555 | 363.7 | 17.0 |
| Gypsy/Roma | AOW | WO | O | 323 | 129.1 | 44.3 |
| White & Black Caribbean | MWBC | BC | WO | 5,077 | 308.0 | 25.0 |
| White & Black African | MWBA | BAO | WO | 1,271 | 339.7 | 23.3 |
| White & Asian | MWA | AO | WO | 2,260 | 385.9 | 14.4 |
| White & Indian | MWA | IND | WO | 82 | 427.6 | 11.0 |
| White & Pakistani | MWA | PKB | WO | 19 | 299.7 | 26.3 |

**Table 4.** (*Continued*)

| PLASC expanded category | PLASC main group | Grp A | Grp B | Total pupils | Avg. capped GCSE | % eligible FSM |
|---|---|---|---|---|---|---|
| White & Any Other Asian | MWA | AO | WO | 140 | 389.4 | 15.7 |
| Any Other Mixed Background | MAOM | MO | MO | 3,490 | 349.3 | 19.4 |
| Asian & Any Other Ethnic Group | MAOM | O | AO | 77 | 366.9 | 16.9 |
| Asian & Black | MAOM | BO | AO | 34 | 324.3 | 26.5 |
| Asian & Chinese | MAOM | CHN | AO | 2 | 255.0 | 100.0 |
| Black & Any Other Ethnic Group | MAOM | O | BNA | 78 | 282.3 | 26.9 |
| Black & Chinese | MAOM | CHN | BNA | 1 | 432.0 | 0.0 |
| Chinese & Any Other Ethnic Group | MAOM | O | O | 21 | 416.3 | 14.3 |
| White & Chinese | MAOM | CHN | WO | 27 | 383.1 | 14.8 |
| White & Any Other Ethnic Group | MAOM | O | WO | 299 | 395.1 | 9.4 |
| Other Mixed Background | MAOM | MO | MO | 651 | 337.6 | 30.3 |
| Indian | IND | IND | IND | 13,668 | 409.4 | 11.6 |
| Pakistani | PAK | PKB | PAK | 8,396 | 334.0 | 32.5 |
| Mirpuri Pakistani | PAK | PKB | KAS | 701 | 306.7 | 35.0 |
| Other Pakistani | PAK | PKB | PAK | 4,282 | 332.8 | 41.4 |
| Kashmiri Pakistani | PAK | PKB | KAS | 654 | 334.0 | 42.4 |
| Bangladeshi | BGD | PKB | BGD | 5,871 | 354.6 | 54.6 |
| Any Other Asian Background | OA | AO | AO | 2,600 | 357.2 | 19.7 |
| African Asian | OA | AO | AO | 161 | 373.7 | 15.5 |
| Kashmiri Other | OA | PKB | KAS | 109 | 326.3 | 37.6 |
| Nepali | OA | AO | RAP | 57 | 327.1 | 0.0 |
| Sinhalese | OA | AO | AO | 32 | 393.6 | 0.0 |
| Sri Lankan Tamil | OA | AO | AO | 357 | 418.6 | 20.2 |
| Other Asian | OA | AO | AO | 1,035 | 377.2 | 24.0 |
| Caribbean | BC | BC | BNA | 8,648 | 304.7 | 24.2 |
| African | BA | BAO | BAO | 4,125 | 329.6 | 26.7 |
| Angolan | BA | BAO | WA | 18 | 262.0 | 66.7 |
| Congolese | BA | BAO | WA | 74 | 249.6 | 71.6 |
| Ghanaian | BA | NG | WA | 628 | 352.2 | 26.8 |
| Nigerian | BA | NG | WA | 1,454 | 359.8 | 24.2 |
| Sierra Leonian | BA | BAO | WA | 79 | 293.8 | 44.3 |

**Table 4.** (*Continued*)

| PLASC expanded category | PLASC main group | Grp A | Grp B | Total pupils | Avg. capped GCSE | % eligible FSM |
|---|---|---|---|---|---|---|
| Somali | BA | SOM | BAO | 1,515 | 251.5 | 82.3 |
| Sudanese | BA | BAO | BAO | 41 | 359.5 | 41.5 |
| Other Black African | BA | BAO | BAO | 2,955 | 321.8 | 34.8 |
| Any Other Black Background | BOB | BO | BNA | 2,481 | 306.8 | 27.5 |
| Black European | BOB | BO | BNA | 51 | 318.8 | 21.6 |
| Black North American | BOB | BO | BNA | 2 | 186.0 | 0.0 |
| Other Black | BOB | BO | BNA | 167 | 300.3 | 22.2 |
| Chinese | CHN | CHN | CHN | 2,106 | 438.0 | 11.1 |
| Hong Kong Chinese | CHN | CHN | RAP | 110 | 479.7 | 4.5 |
| Malaysian Chinese | CHN | CHN | RAP | 3 | 428.0 | 0.0 |
| Singaporean Chinese | CHN | CHN | RAP | 2 | 429.5 | 0.0 |
| Taiwanese | CHN | CHN | RAP | 0 | n/a | n/a |
| Other Chinese | CHN | CHN | RAP | 112 | 437.8 | 18.8 |
| Any Other Ethnic Group | AOEG | O | O | 2,034 | 331.4 | 23.6 |
| Afghanistani | AOEG | ME | RAP | 315 | 300.6 | 54.3 |
| Arab | AOEG | ME | MENA | 270 | 373.2 | 43.7 |
| Egyptian | AOEG | ME | MENA | 35 | 441.9 | 31.4 |
| Filipino | AOEG | O | RAP | 112 | 382.3 | 4.5 |
| Iranian | AOEG | ME | MENA | 195 | 358.1 | 44.1 |
| Iraqi | AOEG | ME | MENA | 119 | 388.3 | 53.8 |
| Japanese | AOEG | O | RAP | 26 | 349.3 | 11.5 |
| Korean | AOEG | O | RAP | 40 | 368.1 | 2.5 |
| Kurdish | AOEG | ME | MENA | 261 | 264.8 | 67.0 |
| Latin American | AOEG | LA | O | 218 | 332.4 | 26.6 |
| Lebanese | AOEG | ME | MENA | 29 | 317.7 | 41.4 |
| Libyan | AOEG | O | MENA | 2 | 273.0 | 50.0 |
| Malay | AOEG | O | RAP | 3 | 429.0 | 0.0 |
| Moroccan | AOEG | O | MENA | 50 | 338.8 | 54.0 |
| Polynesian | AOEG | PI | RAP | 0 | n/a | n/a |
| Thai | AOEG | O | RAP | 18 | 206.3 | 11.1 |
| Vietnamese | AOEG | O | RAP | 240 | 346.0 | 62.9 |
| Yemeni | AOEG | ME | MENA | 144 | 285.5 | 75.0 |
| Other Ethnic Group | AOEG | O | O | 1,039 | 340.7 | 29.2 |

**Table 4.** (*Continued*)

| PLASC expanded category | PLASC main group | Grp A | Grp B | Total pupils | Avg. capped GCSE | % eligible FSM |
|---|---|---|---|---|---|---|
| Refused | RF | NS | NS | 6,962 | 342.0 | 12.6 |
| Information Not Obtained | INO | NS | NS | 8,316 | 312.1 | 11.2 |
| Ethnicity Data Missing | EDM | | | 3,382 | 60.2 | 0.2 |
| TOTAL | | | | 601,548 | 346.0 | 0.0 |

*Notes*: The three aggregations are: PLASC Main Ethnicity Categories, Grouping A and Grouping B. 'Total pupils' = pupils who took GCSE exams in State and maintained schools in England in 2006; 'average capped GCSE' is the average of the capped GCSE point score (see text for definition); '% eligible FSM' is the percentage of pupils in each ethnicity category eligible for free school means.

Abbreviated codes used: AO = Asian other, AOEG = Any Other Ethnic Group, AOW = Any Other White, BA = Black African, BAO = Black African other, BC = Black Caribbean, BGD = Bangladeshi, BI = British & Irish, BNA = Black Non-African, BO = Black other, BOB = Black: Other Black, CEL = Celtic, CHN = Chinese, EDM = Ethnicity Data Missing, EE = Eastern European, EOB = English & Other British, IND = Indian, INO = Information Not Obtained, KAS = Kashmiri, LA = Latin American, MAOM = Mixed: Any Other Mixed, ME = Middle Eastern, MENA = Middle East & North African, MO = Mixed other, MWA = Mixed: White & Asian, MWBA = Mixed: White & Black African, MWBC = Mixed: White & Black Caribbean, NG = Nigerian & Ghanaian, NS = Not Stated, O = Other, OA = Other Asian, PAK = Pakistani, PI = Pacific Islander, PKB = Pakistani-Kashmiri-Bangladeshi, RAP = Rest of Asia or Pacific, RF = Refused, SOM = Somali, WA = Western African, WB = White British, WE = White European, WI = White Irish, WO = White other, WSE = Western European.

**Table 5.** Educational attainment and eligibility for free school meals, calculated by three aggregations of the PLASC expanded ethnicity categories

| PLASC main group | Total pupils | Avg. capped GCSE | % eligible FSM | Grouping A | Total pupils | Avg. capped GCSE | % eligible FSM | Grouping B | Total pupils | Avg. capped GCSE | % eligible FSM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Chinese | 2,333 | 439.9 | 11.1 | Chinese | 2,363 | 439.1 | 11.2 | Chinese | 2,106 | 438.0 | 11.1 |
| Indian | 13,668 | 409.4 | 11.6 | Indian | 13,750 | 409.5 | 11.6 | India | 13,668 | 409.4 | 11.6 |
| Mixed: White & Asian | 2,501 | 386.8 | 14.4 | Asian other | 6,642 | 374.4 | 18.1 | Asian other | 4,298 | 367.9 | 20.5 |
| Other Asian | 4,351 | 366.7 | 20.7 | Western European | 1,631 | 363.9 | 14.3 | Rest of Asia or Pacific | 1,038 | 358.3 | 34.6 |
| Bangladeshi | 5,871 | 354.6 | 54.6 | Nigerian & Ghanaian | 2,082 | 357.5 | 25.0 | Bangladeshi | 5,871 | 354.6 | 54.6 |
| Mixed: Any Other Mixed | 4,680 | 350.1 | 20.4 | White other | 9,059 | 352.7 | 13.6 | Celtic | 2,669 | 353.7 | 17.2 |
| White British | 485,509 | 348.2 | 10.7 | British & Irish | 487,819 | 348.2 | 10.7 | Western African | 2,253 | 351.0 | 27.5 |
| White Irish | 2,310 | 348.1 | 18.8 | Mixed other | 4,141 | 347.5 | 21.1 | English & Other British | 485,150 | 348.1 | 10.7 |
| Any Other White | 12,735 | 347.2 | 17.4 | Other | 4,039 | 341.0 | 25.7 | Mixed other | 4,141 | 347.5 | 21.1 |
| Refused | 6,962 | 342.0 | 12.6 | Pakistani-Kashmiri-Bangladeshi | 20,032 | 338.8 | 41.3 | White other | 16,693 | 347.4 | 17.5 |
| Mixed: White & Black African | 1,271 | 339.7 | 23.3 | Latin American | 218 | 332.4 | 26.6 | White European | 4,894 | 341.8 | 22.6 |
| Any Other Ethnic Group | 5,150 | 333.8 | 34.5 | Black African other | 8,563 | 327.4 | 29.7 | Middle East & North Africa | 1,105 | 334.1 | 54.5 |
| Pakistani | 14,033 | 332.3 | 35.8 | Not Stated | 15,278 | 325.7 | 11.8 | Pakistani | 12,678 | 333.6 | 35.5 |
| Black African | 10,889 | 321.2 | 36.9 | Middle Eastern | 2,487 | 317.2 | 48.7 | Not Stated | 15,278 | 325.7 | 11.8 |
| Information Not Obtained | 8,316 | 312.1 | 11.2 | Eastern European | 926 | 313.7 | 30.1 | Kashmiri | 1,464 | 320.3 | 38.5 |

**Table 5.** (*Continued*)

| PLASC main group | Total pupils | Avg. capped GCSE | % eligible FSM | Grouping A | Total pupils | Avg. capped GCSE | % eligible FSM | Grouping B | Total pupils | Avg. capped GCSE | % eligible FSM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Mixed: White & Black Caribb. | 5,077 | 308.0 | 25.0 | Black other | 2,735 | 306.8 | 27.0 | Other | 3,635 | 316.6 | 27.2 |
| Black: Other Black | 2,701 | 306.6 | 27.0 | Black Caribb. | 13,725 | 305.9 | 24.5 | Black African other | 8,636 | 313.4 | 39.3 |
| Black Caribb. | 8,648 | 304.7 | 24.2 | Somali | 1,515 | 251.5 | 82.3 | Black Non-African | 11,428 | 305.0 | 24.9 |
| Ethnicity Data Missing | 4,543 | 139.8 | 2.4 | Ethnicity Data Missing | 4,543 | 139.8 | 2.4 | Ethnicity Data Missing | 4,543 | 139.8 | 2.4 |
| Total | 601,548 | 346.0 | 13.1 | Total | 601,548 | 346.0 | 13.1 | Total | 601,548 | 346.0 | 13.1 |

FSM = Free school meals, GCSE = General Certificate of Secondary Education. Each section of the table is individually sorted by the average capped GCSE result in descending order. The descriptions of the ethnicity groupings presented here are the long version of those represented in Table 2 through abbreviated codes.

**Table 6.** Regression results of GCSE results against eligibility for free school meals for three alternative aggregations of PLASC ethnicity categories

| Aggregation | Adjusted $R^2$ | | Coefficients | Standard error | t Stat |
|---|---|---|---|---|---|
| PLASC main group | 0.141 | Intercept | 381.3 | 18.6 | 20.5 |
| | | $X1_i$ | −135.9 | 71.3 | −1.9 |
| Grouping A | 0.556 | Intercept | 392.4 | 12.8 | 30.7 |
| | | $X1_i$ | −182.6 | 39.9 | −4.6 |
| Grouping B | 0.175 | Intercept | 381.5 | 16.9 | 22.5 |
| | | $X1_i$ | −116.1 | 55.4 | −2.1 |

*Note*: The categories 'Refused', 'Information Not Provided', 'Not Stated' and 'Ethnicity Data Missing' are not included in the regression analysis.

and if there is (cf. Shuttleworth 1995), whether this is sensitive to the aforementioned aggregation effects. A linear regression was calculated between average capped GCSE score as the dependent variable and the percentage of pupils eligible for free school meals as the independent variable. The different relationship between these two variables across the three different groupings of ethnicity is shown in the regression results of Table 6. The adjusted $R^2$ statistics are 0.141 for the PLASC Main Group, 0.556 for Grouping A and 0.175 for Grouping B. Together with the parameter values, standard error and *t* statistic shown in Table 6, this demonstrates that only Grouping A shows a strong relationship between these two variables. This means that, using the Grouping A aggregation of ethnicity classes, it can be argued that the lower the percentage of pupils eligible for free school meals an ethnic group has, the better the GSCE scores it gets. If this aggregation were not challenged, the argument would pass as valid, yet if different aggregations were used the result would not be apparent (in large part because of the high leverage effect of the Somali group in Grouping A). These results illustrate how changes in the composition of the ethnicity classes, resulting from aggregating the PLASC extended categories in different ways, have far-reaching impacts upon the results of the analysis. In short, the outcome of ethnicity analysis possesses no validity independent of the definitions of the ethnicity classes adopted in each study.

## Conclusion

A central contribution of this paper is that different types of uncertainty can impact upon the results of research studies on ethnicity, and that these are not independent of the definition of the ethnic groups that underpin the analysis that they develop. Although the issues associated with uncertainties in the conception, representation and measurement of ethnicity (i.e. going from an individual's subjective identity to some form of workable group measure) have been frequently debated in the literature, little attention has been paid to the study of uncertainties in the *analysis* of ethnicity classifications, which pertain to issues of the extent (number and size) and aggregation of ethnic groups (i.e. going from measures to results). As such, this

source of uncertainty could be conceived as an analogy to the Modifiable Areal Unit Problem (MAUP) in geography (Openshaw 1984) in that, by altering the number, size and 'arrangements' of ethnic group 'units', a range of different results for the same input data are obtained. We could even term the uncertainty in the analysis of ethnicity discussed in this paper as the 'Modifiable Ethnic Unit Problem' or MEUP, although this is somehow an overstretched and insensitive expression. As shown in this paper, the uncertainty in the *analysis* of ethnicity affects most studies regardless of their use of bespoke or official ethnicity classifications. This appreciation is only now becoming apparent because of the multiple analysis possibilities brought by an increase in the resolution of ethnicity classifications over time, as well as in the number of dimensions of different aspects of identity now being measured, such as religion, language, nationality, migration status, ancestry, etc.

Our intention in illustrating these uncertainties in the analysis of ethnicity is not to suggest in any way that they can be 'eliminated'. Rather, our discussion makes clear that different types of uncertainties are inherent in any classification of ethnicity and that our objective should be to manage their effects through best practices that are robust, well honed and open to scrutiny by other researchers.

If these concerns can be met, there is much that can be done to improve the ways in which official classifications of ethnicity are analysed across a range of application domains, and we have begun to illustrate this here through our case studies in educational attainment and free school meals. The motivation for ethnicity classification is likely to vary considerably between domains, as are the relative priorities assigned to comparability and transferability of research findings versus the drive for conclusive, focused results. This paper has used one of the most finely granulated ethnicity classifications available today in the UK—the PLASC expanded ethnicity categories—in order to demonstrate the existence of such extent and aggregation effects in the analysis of ethnicity, illustrate their impact, and define how these differ from previously identified sources of uncertainty in ethnicity research. With a similar view, in previous work we have used an even more disaggregated ethnicity classification, including over 180 groups based upon the origin of people's names (Mateos 2007). With this we hope to have opened a new avenue for applied research that will become ever more relevant as the ways in which we conceive, represent, measure and analyse the different dimensions of individual and communal identity become increasingly complex, in response to growing demands on policy-makers to better recognise difference and diversity in contemporary societies.

## Note

[1] In the analysis presented here, the NPD educational attainment data have been linked at the pupil level to the corresponding demographic data contained in the PLASC records. These data pertain only to those students taking the GCSE (General Certificate of Secondary Education) or equivalent in England in 2006, who are generally aged 15 or 16. These pupils are those who appear in the NPD dataset recorded as 'Key Stage 4 final candidate' in the 2005/2006 academic year, and those in PLASC 2006 for which matching records were found.

## References

Agyemang, C., Bhopal, R. and Bruijnzeels, M. (2005) 'Negro, Black, Black African, African Caribbean, African American or what? Labelling African origin populations in the health arena in the 21st century', *Journal of Epidemiology and Community Health, 59*(12): 1014–18.

American Sociological Association (2002) *Statement of the American Sociological Association on the Importance of Collecting Data and Doing Social Scientific Research on Race. Adopted by the elected Council of the American Sociological Association.* Available at: http://www.asanet.org/galleries/default-file/asa_race_statement.pdf. Accessed: 20 September 2006.

Aspinall, P.J. (2002) 'Collective terminology to describe the minority ethnic population: the persistence of confusion and ambiguity in usage', *Sociology, 36*(4): 803–16.

Aspinall, P.J. (2003) 'Who is Asian? A category that remains contested in population and health research', *Journal of Public Health Medicine, 25*(2): 91–7.

Aspinall, P.J. (2005) 'The operationalization of race and ethnicity concepts in medical classification systems: issues of validity and utility', *Health Informatics Journal, 11*(4): 259–74.

Aspinall, P.J. (2007) 'Approaches to developing an improved cross-national understanding of concepts and terms relating to ethnicity and race', *International Sociology, 22*(1): 41–70.

Bains, B. and Klodawski, E. (2006) *GLA 2005 Round: Interim Ethnic Group Population Projections.* London: Greater London Authority, DMAG Briefing 2006/22. Available at: http://www.london.gov.uk/gla/publications/factsandfigures/dmag-briefing-2006-22.pdf.

Bhopal, R. (2004) 'Glossary of terms relating to ethnicity and race: for reflection and debate', *Journal of Epidemiology and Community Health, 58*(6): 441–5.

Bhopal, R. (2007) *Ethnicity, Race, and Health in Multicultural Societies.* Oxford: Oxford University Press.

Bhopal, R. and Donaldson, L. (1998) 'White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health', *American Journal of Public Health, 88*(9): 1303–7.

Bosveld, K., Connolly, H. and Rendall M.S. (2006) *A Guide to Comparing 1991 and 2001 Census Ethnic Group Data.* Titchfield: Office for National Statistics, available at: http://www.statistics.gov.uk/articles/nojournal/Guidev9.pdf. Accessed: 12 May 2008.

Brubaker, R. (2004) *Ethnicity Without Groups.* London: Harvard University Press.

Bulmer, M. (1996) 'The ethnic group question in the 1991 Census of Population', in Coleman, D. and Salt, J. (eds) *Ethnicity in the 1991 Census. Volume 1. Demographic Characterisitics of the Ethnic Minority Populations.* London: Office for National Statistics/HMSO, xi–xxix.

Castles, S. and Miller, M.J. (2003) *The Age of Migration.* Basingstoke: Palgrave Macmillan (3rd edition).

Choi, K.H. and Sakamoto, A. (2005) *Who is Hispanic? Hispanic Ethnic Identity Among African Americans, Asian Americans, and Whites.* Austin: University of Texas at Austin, Population Research Centre Working Paper No. 04-05-07. Available at: http://www.prc.utexas.edu/working_papers/wp_pdf/04-05-07.pdf. Accessed: 22 February 2005.

Comstock, R.D., Castillo, E.M. and Lindsay, S.P. (2004) 'Four-year review of the use of race and ethnicity in epidemiologic and public health research', *American Journal of Epidemiology, 159*(6): 611–19.

Connolly, H. and Gardener, D. (2005) *Who Are the 'Other' Ethnic Groups?* London: Office for National Statistics. Available at: http://www.statistics.gov.uk/articles/nojournal/other_ethnicgroups.pdf. Accessed: 27 January 2006.

Department for Children, Schools and Families (2006) *Secondary Schools (GCSE and Equivalent) Achievement and Attainment Tables 2006.* Available at: http://www.dcsf.gov.uk/performancetables/schools_06/s8.shtml. Accessed: 14 November 2007.

Department for Education and Skills (2006) *Ethnic Minority Achievement. Data Collection—Ethnicity.* The Standards Site. Available at: http://www.standards.dfes.gov.uk/ethnicminorities/collecting/763919/. Accessed: 07 November 2007.

Department of Health (2005) *A Practical Guide to Ethnic Monitoring in the NHS and Social Care.* Available at: http://www.dh.gov.uk/assetRoot/04/11/68/43/04116843.pdf. Accessed: 23 September 2005.

Ellis, M. and Wright, R. (2005) 'Assimilation and differences between the settlement patterns of individual immigrants and immigrant households', *PNAS: Proceedings of the National Academy of Sciences of the United States of America, 102*(43): 15325–30.

Eriksen, T. (2002) *Ethnicity and Nationalism.* London: Pluto Press (2nd edition).

Godfrey, R. (2004) *Changes in Ethnicity Codes in the Pupil Level Annual Schools Census 2002–2003.* Canterbury: Christ Church University College. Available at: http://www.dfes.gov.uk/rsgateway/DB/STA/t000455/index.shtml. Accessed: 20 October 2007.

Howard, D. and Hopkins, P.E. (2005) 'Editorial: race, religion and the census', *Population, Space and Place, 11*(2): 69–74.

Jones, P. and Elias, P. (2006) *Administrative Data as Research Resources: A Selected Audit.* National Data Strategy. ESRC. Available at: http://www.rss.org.uk/pdf/Admin%20Data%20selected%20audit%20report%20v2.pdf. Accessed: 11 November 2007.

London Health Observatory (2005) *Using Routine Data to Measure Ethnic Differentials in Access to Revascularisation in London.* Available at: http://www.lho.org.uk/viewResource.aspx?id=9732. Accessed: 20 July 2006.

Longley, P.A., Maguire, D.J., Goodchild, M.F. and Rhind, D. (2005) *Geographic Information Systems and Science.* Chichester: Wiley.

Mateos, P. (2007) *An Ontology of Ethnicity Based Upon Personal Names. Implications for Neighbourhood Profiling.* London: University College London, Department of Geography, unpublished PhD thesis, available at: http://www.casa.ucl.ac.uk/pablo/papers/Mateos_2007_PhD_Thesis.pdf

Morning, A. (2008) 'Ethnic classification in global perspective: a cross-national survey of the 2000 census round', *Population Research and Policy Review, 27*(2): 239–72.

Nazroo, J. (2003) 'Patterns of and explanations for ethnic inequalities in health', in Mason, D. (ed.) *Explaining Ethnic Differences: Changing Patterns of Disadvantage in Britain.* Bristol: Policy Press, 87–104.

Nobles, M. (2000) *Shades of Citizenship: Race and the Census in Modern Politics.* Stanford: Stanford University Press.

Office for National Statistics (2003) *Ethnic Group Statistics: A Guide for the Collection and Classification of Data.* Available at: http://www.statistics.gov.uk/about/ethnic_group_statistics/downloads/ethnic_group_statistics.pdf. Accessed: 13 February 2006.

Openshaw, S. (1984) *The Modifiable Areal Unit Problem.* Norwich: Geo Books.

Park, R. (1955[1921]) *Society: Collective Behaviour, News and Opinion.* Hughes, E. *The Collected Papers of Robert Ezra Park.* Volume 3. Glencoe, III: The Free Press.

Peach, C. (2000) 'Discovering white ethnicity and parachuting plurality', *Progress in Human Geography, 24*(4): 620–6.

Platt, L., Simpson, L. and Akinwale, B. (2005) 'Stability and change in ethnic groups in England and Wales', *Population and Trends, 121*: 35–46.

Rankin, J. and Bhopal, R. (1999) 'Current census categories are not a good match for identity', *British Medical Journal, 318*(7199): 1696.

Rees, P. and Butt, F. (2004) 'Ethnic change and diversity in England, 1981–2001', *Area, 36*(2): 174–86.

Robbin, A. (1999) 'The problematic status of US statistics on race and ethnicity: an "imperfect representation of reality"', *Journal of Government Information, 26*(5): 467–83.

Sammons, P. (1995) 'Gender, ethnic and socio-economic differences in attainment and progress: a longitudinal analysis of student achievement over 9 years', *British Educational Research Journal, 21*(4): 465–85.

Senior, P.A. and Bhopal, R. (1994) 'Ethnicity as a variable in epidemiological research', *British Medical Journal, 309*(6950): 327–30.

Shuttleworth, I. (1995) 'The relationship between social deprivation, as measured by individual free school meal eligibility, and educational attainment at GCSE in Northern Ireland: a preliminary investigation', *British Educational Research Journal, 21*(4): 487–504.

Simpson, L. and Akinwale, B. (2007) 'Quantifying stability and change in ethnic groups', *Journal of Official Statistics, 23*(2): 185–208.