GREATERLONDONAUTHORITY

## Data Management and Analysis Group

# The National and London Pupil Datasets

An introductory briefing for researchers and research users

David Ewens

DMAG Briefing 2005/8
February 2005
ISSN 1479-7879

# DMAG Briefing 2005/8

## February 2005

# The National and London Pupil Datasets

An introductory briefing for researchers and research users

For more information please contact:

David Ewens
Data Management and Analysis Group
Greater London Authority
City Hall
The Queen's Walk
London SE1 2AA

Telephone: 020 7983 4656
Email: david.ewens@london.gov.uk



ISSN 1479-7879

# Contents

## 1. Summary

- The National Pupil Dataset (NPD) and the London Pupil Dataset (LPD) LPD are both new and, at this time, neither is well known outside a limited circle of researchers and research-users. This briefing provides an introduction to key features of both datasets. It points both to areas of strength and potential, and to some of the (many) pitfalls the data hold for the unwary.

- Comments in the briefing reflect experience at the GLA. This involves work with large, annual (i.e. recurrent), extracts from the NPD, analysed in the main in SPSS. Comments also reflect the place of the NPD in a wider, national, development of evidence-based policy and practice.

- In that wider context, data in the NPD are particularly appropriate for effect size and impact analysis, and for analysing what is sometimes called the political arithmetic or political economy of education. The latter can be summarized as 'who gets what, where, when and why in society?'

- The NPD first came into being 2002. It contains individual pupil records for ***all*** children in the maintained (state) school system and is updated annually. The information gathered can change from one year to the next, and is to some extent open to development. The NPD contains information such as pupil date of birth, gender, ethnicity, date admitted to school, special educational needs, key stage assessments and public examination results.

- Because data are for the population of children at maintained schools, the NPD overcomes the problems of smaller datasets and sample surveys, where limited numbers can at times make it difficult to reach meaningful conclusions. The annual update means information is 'current' in a sense that older or less frequently updated, population datasets are not. This is particularly important in areas, such as London, which experience large-scale change over comparatively short periods of time.

- In any one year, approximately 8,000,000 pupil records are added to the NPD. Pupil records from one year can be linked to pupil records from an earlier year. Data in the NPD are potentially longitudinal. A number of variables in the NPD also allow information to be linked to data from elsewhere, including information from the 2001 national census and from EduBase, the national education institution dataset. Combined information can be coded to regional, local authority, ward, and census output area levels.

- Access to data in the NPD is by application to DfES Analytical Services, and is for specific research purposes. The terms of access require each applicant to state what those purposes are. Further approval from DfES Analytical Services is needed if, having received data, an applicant develops additional research goals. Data extracts are anonymised, must

be held securely, and cannot be passed to others. Placing analyses of the data from the NPD in the public domain requires prior approval from DfES Analytical Services. Terms may be varied in cases where research is commissioned by the DfES itself.

- The terms and conditions set by DfES Analytical Services provide a framework within which research proposals can be developed. They also provide individual researchers with a measure of protection.

- The NPD is not a single dataset. It is a collection of separate but linkable data sets, or what is sometimes called a data warehouse. The datasets which go to make up the NPD include the Pupil Level Annual Schools Census (PLASC), and separate assessment datasets for key stages 1 to 4+. Extracts provided to the GLA reflect that structure. The LPD does not arrive as a single, 'ready for analysis', dataset.

- The LPD is largely, but not entirely, a subset of the NPD. As with any other NPD data release, the LPD is held securely and for specific research purposes.

- The LPD contains records for pupils who attend schools maintained by any of London's 33 separate local authorities, regardless of where those pupils live, or who live in London and attend maintained schools outside London. In the LPD, the NPD has provided the first pan-London pupil dataset in the capital's history. It is able to offer analyses relevant to the GLA's strategic interests in quality and equality issues, social inclusion and exclusion and regeneration. Arguably, it is difficult to see how those areas could be understood without taking educational issues into account.

- The purpose of the NPD as an education dataset is reflected in the documentation available to the researcher. There is no single, coded, 'researcher manual', which provides all the information someone who is not an education specialist might need to understand the data in the NPD.

- The NPD draws on information which is a side-product of the work of a wide range of senior staff in schools and in local authorities, including Chief Education Welfare Officers, school examination secretaries, national curriculum co-ordinators, and special needs co-ordinators. Each group is supported by separately published advice from specialists in the Department for Education and Skills, in the Qualifications and Curriculum Authority, in the Office for Standards in Education and elsewhere. Taken together, those separate documents are crucial elements in the documentation of the NPD and LPD. Neither the NPD nor the LPD are suitable for analysis by those who have not familiarised themselves with that documentation. Similarly, neither the NPD nor LPD is suitable for analysis by those who are unfamiliar with the maintained school system.

- While the LPD is an education dataset, and appropriate for analysis of issues such as school improvement, the data may have tangential benefits for those working in other areas, such as health studies, neighbourhood profiling and spatial analysis.

- In any one year, in excess of 1,000,000 individual pupil records meet the criteria for inclusion in the LPD. The 2004 LPD contains 5,000,000 plus key stage 4 individual subject entry records. A slimmed merged version of the 2002 and 2003 LPDs has the same longitudinal characteristics as the NPD. At 8+ Gigabytes, it is the single largest dataset held by the GLA.

- Experience of work with small files and spreadsheets, does not provide an adequate basis for work with either the NPD or LPD. Additionally, the NPD draws on recurrent but changing data collection exercises. Work with recurrent surveys requires skills over and above those required in one-off work with an individual survey, and that is compounded by the changing nature of the data in the NPD.

- Work on datasets of the size and complexity of the NPD is time-consuming, and this needs to be allowed for. One effect at national level is that the times at which data have been released to the GLA has varied considerably from one year to the next. The majority of the data in the NPD are alphanumeric. Time will be needed by those working with systems such as SPSS to recode and label data in numeric form.

- The size and complexity of the NPD raises questions about the computing capacity required to store and analyse data. Larger extracts from the NPD are at or beyond the limits of current desktop personal computers. Additionally, the impact of different computer operating systems on the processing and analysis of large datasets will need to be considered.

- Thanks are expressed to Jayesh Amin, of the London borough of Haringey, who kindly provided a blank copy of DfEE Form 7. This was the predecessor to PLASC, and extracts are included here as appendices to illustrate the limitations of pre-NPD data.

## 2. Introduction. What can the National Pupil Dataset (NPD) do for you?

The National Pupil Dataset is amongst the most important national innovations in data collection in the recent past. Its potential is considerable and the scope for development is also considerable. Nonetheless, it is not necessarily well know. Mention of it to date tends to prompt the questions 'what is the NPD, what is it for, what can it do?

The NPD is a national dataset containing pupil level records of each child attending a maintained school. There was more than one reason for establishing the NPD, and each has a bearing on what it can now do. However, despite a complex evolution, the NPD is well placed to provide analyses of

1. education outcomes for different groups, in wider social context; that is, to answer 'who gets what, where, when and why-type questions, and to
2. provide analyses of education improvement.

Both are central to quality and equality of life issues, to social inclusion, exclusion, regeneration and to the wider public interest. All of these are major concerns for the GLA, for areas of national policy, and for more than one tradition of scholarly research. In a capital city with 33 separate local education authorities (LEAs), the NPD offers the first ever pan–London pupil level dataset. Analysis of this comparatively large dataset is likely to provide more definite answers than can be inferred from smaller dataset held by individual local authorities. Additionally, analyses can group data to match the geography of individual jurisdictions such as Learning and Skills Council areas, as well London boroughs. Those analyses may prove valuable in assessing the impact of the policies of bodies with education responsibilities within their area of jurisdiction. There may also be some tangential benefits from the NPD for other areas of concern, such as spatial modelling or health studies.

In broad terms, the development of the NPD is part of the development of evidence-based policy and practice in the United Kingdom. The link between the NPD and evidence-based practice is touched on at several points later in the briefing, and the potential implications of the NPD for regional government such as the GLA has been referred to above and will be developed further below. However two examples are given here, highlighting key issues affecting work with the NPD in terms of a national trend towards evidence-based policy.

Trevor Knight, then at the Department of Education and Skills (DfES) Analytical Services, has pointed to the value of data, of the type found in the NPD, in analyses of the impact of policy and practice.

> *The concept of effect (*sizes) *or impact assessments … is pivotal in developing a clearer appreciation of the factors which appear important in developing and evaluating policy, and to which pupil and other groups those effects are or are not important.*
> *(*In Ian Schagen and Karen Eliot (eds), *But what does it mean? The use of effect sizes in educational research.* NFER, 2004, page 77).

Effect size measures are more complex than simple descriptive statistics but complex measures may be the 'right' ones to use in situations which are themselves complex. The review of aspects of social selection and London schools, in the 2003 DMAG briefing *Cross-border pupil mobility,* illustrates this point.

Social selection in schooling implies that socially advantaged pupils are more likely than other pupils to attend particular types of school or be placed in particular types of class within a school. The factors associated with this are likely to be complex. At the point of secondary transfer, for instance, patterns of transfer to particular types of secondary school may be associated with pupil attainment at the end of primary school, type of primary school attended, pupil ethnicity, gender, socio-economic status, religion and home neighbourhood type.

The *Cross-border pupil mobility* briefing uses descriptive statistics to show, for example, pupil characteristics tabulated with type of school attended, allowing for whether the school attended is in the pupil's home LEA area or not. One graph in the briefing shows pupil ethnicity, by type of school attended, by whether the school is in the pupil's home LEA area or not, and by pupil level of attainment in key stage 2 at the end of primary schooling. It is not the simplest graph to read. If other factors, such as pupil gender or free school meal entitlement had been introduced in that graph, it may well have become impossible to interpret.

At this point a policy-maker might reasonably ask whether there are any statistical tests which could simplify that picture by identifying key, possibly related, underlying variables which together statistically 'explain' whether social selection in schooling exists and, perhaps, if so whether it is by schools of pupils or by parents of schools. Such tests exist, and they clearly have a value where research users already know that a multiplicity of factors influence particular outcomes. While this may appear counter-intuitive to some, one aim of statistical analysis can be to simplify the complex.

An acceptance of complexity is echoed in a recent series of articles in the British Journal of Sociology (Issues 1 and 3, vol. 55, 2004), on the future of evidence-based policy analysis. Taken together, Philip Davies from the Cabinet Office, Hugh Lauder, Phillip Brown, and A.H. Halsey include in their BJS discussion a concern with the application of statistical techniques which provide policy makers with explanations rather than mere descriptions. That concern overlaps with the point made by Trevor Knight. Both points of view implicitly accept that in an era of evidence-based policy and practice, analyses will be needed which are appropriate to situations of complexity. These will go beyond descriptive statistics and administrative arithmetic.

This is not to say that a statistical technique is 'good' if it is complex, though this inference might be drawn from a casual reading of the BJS papers. A technique is 'good' by virtue of being appropriate to the task in hand. There is, indeed, a tradition of work in Britain which uses either descriptive statistics or virtually no statistics at all, and which has added considerably to understanding. Work

carried out by Michael Young and Peter Willmott at the Institute for Community Studies, and by Charles Madge, Tom Harrison and the diary keepers of Mass Observation, provide cases in point. Nonetheless, both Trevor Knight and those writing in the BJS, correctly point to a world of evidence-based policy-making and practice which is increasingly at ease with, and requires analyses appropriate to, complex situations. Information in the NPD can support that type of analysis, though not necessarily to the extent that might be wished. Section 12, for example, points to limitations in the data as far as the analysis of school effectiveness is concerned.

While there are limitations in the NPD in its current form as far as the analysis of school effectiveness is concerned, it is nonetheless particularly appropriate for analyses of 'who gets what, where, when and why'. Such analyses have a bearing on issues such as equality and inequality, social inclusion, social exclusion and regeneration. Each is a policy priority at the GLA (and in a number of organisations elsewhere). The BJS discussion also includes a concern with analyses of who gets what in society and why, which may in part reflect A.H. Halsey's long-standing interest in what is sometimes called the political economy, and sometimes the political arithmetic, of education. Taken together, these factors explain why the first major reason given at the start of this section for analysing data from the NPD is a political economy of education-type reason.

In addition to advocating statistical analyses which explain rather than merely describe, and restating the case for the political economy of education, the BJS papers also refer to three further steps which might be taken to strengthen analyses currently available to policy-makers. These are that analyses need to become more multi-disciplinary, more organised, and better communicated. The first two of these three points are particularly appropriate to work with the NPD.

In some circumstances, a multidisciplinary approach may offer a more rounded response to a policy issue, and in any event no single dataset will ever enable the analyst to say 'everything' about 'everything'. At least as far as analyses of school effectiveness are concerned, the NPD is no exception to this rule. However, public sector organisations such as the GLA are also required to carry out public consultations and user surveys. There may be a distinct advantage to such organisations in dovetailing work on the NPD with complementary strands of work on the views and actions of children, parents, teachers and those in local authorities. As far as the value of an organised approach is concerned, this briefing stresses that nature of the NPD actually requires this. Organisation is not a merely a desirable but ultimately optional extra (see sections 10 and 11).

A distinctive characteristic of the NPD is that it has, from the outset, been intended to make a contribution to understanding of the larger education issues, and is very much a part of the broader development of evidence-based policy and practice. It is certainly not an 'administrative dataset', to be understood as separate from that broader development, and it may be able to contribute tangentially to evidence-based policy issues beyond those concerned with the political economy of education.

## 3. What is the National Pupil Dataset (NPD)? What is the London Pupil Dataset (LPD)?

The LPD is mainly, but not entirely, based on a subset of the NPD. Data are released to the GLA for analysis of pan-London and strategic issues, such as social inclusion, social exclusion and mobility. Records are of pupils in the maintained sector who live in London, regardless of where they attend school, or who attend school in London regardless of where they live.

The NPD is what is commonly called a data warehouse. It holds individual pupil level and other information gathered in a range of separate data collection exercises each year, such as the Pupil Level Annual Schools Census (PLASC) and the separate exercises for collecting pupil level key stage assessments and public examination results. Researchers new to the NPD have at times referred to it in terms of one of those component data collection exercises, and more particularly have confused PLASC with the NPD as a whole. Clearly, a rose is still a rose by any other name, but asking the DfES for data from PLASC, when what is required is assessment data, may result in disappointment, confusion or both.

Within the NPD, data are held in separate, linkable, files, and information received by the GLA is in that form, rather than in the form of a single 'flat' file. The data collection exercises used to populate the NPD have a statutory basis, which improves the quality of returns from schools, and data are collected for all children in maintained schools. That is, information in the NPD relates to the total population of children in those schools, rather than to a sample.

The NPD came into being in 2002, and has been updated annually since. In excess of 8,000,000 million individual pupil records are added each year, covering a range of items including pupil age, gender, ethnicity, special educational needs, free school meal entitlement, key stage assessments and public examination results, home postcode and school attended. Because it is updated annually, and records from one year can be linked to records from another, it can provide cross-sectional *and* longitudinal views of education.

Table 1 shows the years the DfES were able to provide key stage assessment information for those pupils on roll in January 2002 and in the LPD, and also shows that summer 2003 key stage assessments have been made available by the DfES for pupils on roll in January 2003. This information is crucial in measuring aspects of pupil progress through the school system, which in turn is key to assessing the impact of policies aimed at lifting levels of attainment in low achieving areas such as the inner city. Arguably, these types of consideration are close to those expressed in the BJS articles referred to in section 2, and place the development of the NPD within a wider development of evidence-based policy and practice.

Approximately 8,000,000 million records are added to the NPD each year, and 1,000,000 each year to the LPD. Given the number of records, both datasets avoid the problem encountered in smaller sample surveys where limited numbers can make it impossible in some instances to arrive at statistically

meaningful conclusions. Interestingly, it has now been decided that another national exercise measuring attainment amongst young people will no longer rely on a long-standing sample survey. It will now use a combination of individual student public examination records, Individualised Learner Records held by the Learning and Skills Council, and information from the National Information System for Vocational Qualifications held by the DfES (see DfES Statistical First Release 48/2004).

**Table 1. LPD assessment information available for pupils on roll in 2002, and assessment data from 2003***

| | School year and age group for which assessment information is available | | | | | |
|---|---|---|---|---|---|---|
| | 1996/7 | 1997/8 | 1998/9 | 1999/2000 | 2001/2 | 2002/3 |
| Pupil age at 31st August prior to start of school year | 13 | 14 | 15 | 16 | 17 | 18 |
| | 12 | 13 | 14 | 15 | 16 | 17 |
| | 11 | 12 | 13 | 14 | **15** | 16 |
| | **10** | 11 | 12 | **13** | 14 | **15** |
| | 9 | **10** | 11 | 12 | **13** | 14 |
| | 8 | 9 | **10** | 11 | 12 | **13** |
| | 7 | 8 | 9 | **10** | 11 | 12 |
| | 6 | 7 | 8 | 9 | **10** | 11 |
| | | **6** | 7 | 8 | 9 | **10** |
| | | | **6** | 7 | 8 | 9 |
| | | | | **6** | 7 | 8 |
| | | | | | **6** | 7 |
| | | | | | | **6** |

**Key – Years and broad pupil age groups for which assessment information is available**

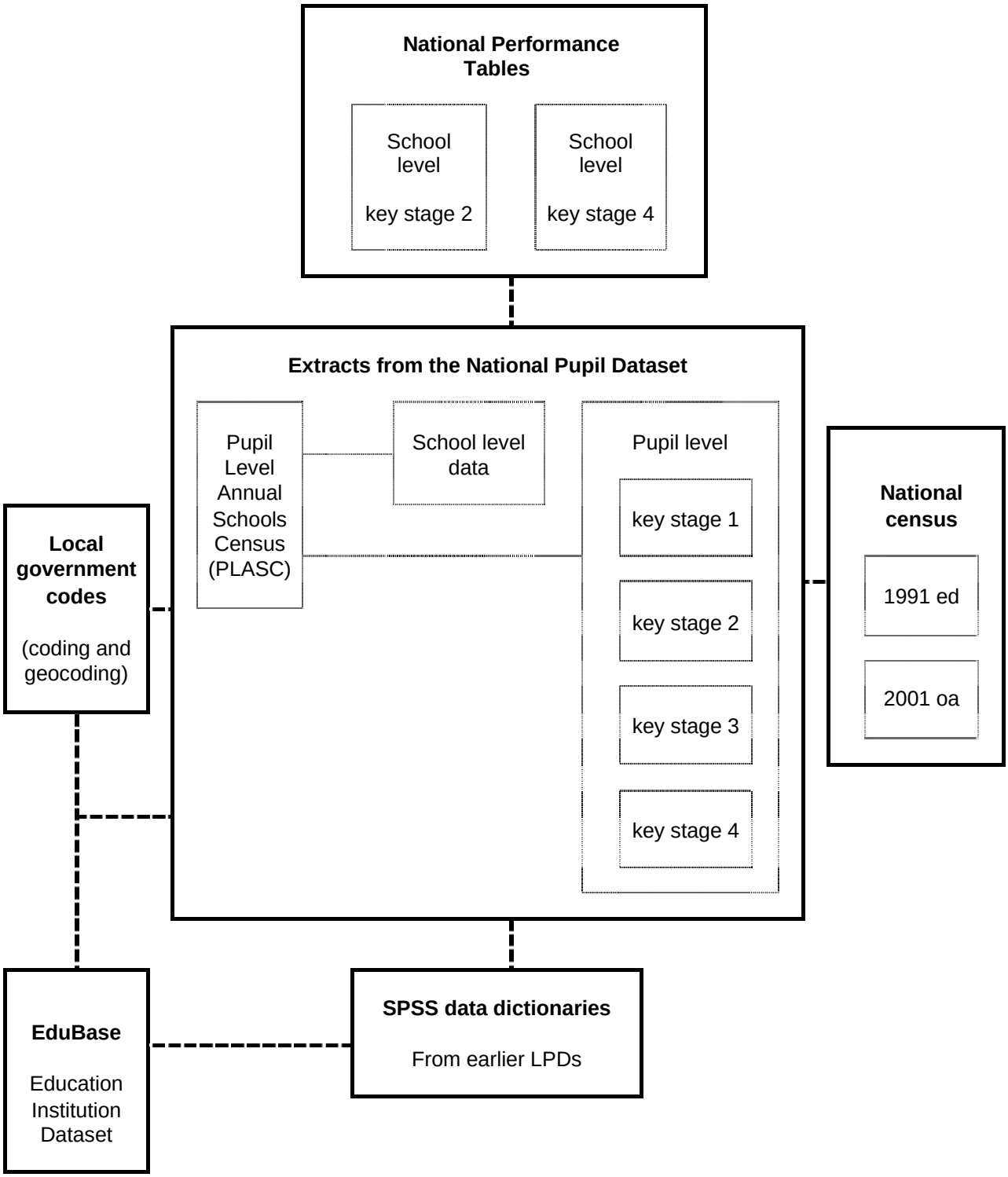
| | |
|---|---|
| Key stage 1 available | **6** |
| Key stage 2 available | **10** |
| Key stage 3 available | **13** |
| GCSE/GNVQ available | **15** |

Equally importantly, the annual updates of the NPD and LPD mean that they provide contemporary information in a way that older studies such as the much used National Child Development Study, which is based on a single cohort of people born in Britain in one week in March 1958, cannot. The same point would apply to more recent and recurrent surveys, which are carried out at longer intervals than applies to the NPD.

The contemporary nature of data in the NPD and LPD is particularly important in areas such as London, which are characterised by large scale, and at times rapid, change. By way of an example, the ethnic composition of schools in London today bears little resemblance to the situation when the generation born in March 1958 attended school. The experience of the British generation of 1958 may provide only a limited guide to the future of children in London's schools today.

A further strength of the NPD is the presence of variables that allow data to be linked to other datasets, including the 2001 national census. Figure 1 gives a simplified schematic of the LPD, and shows the range of datasets which have been drawn on in the GLA. Appendix table A2 shows in more detail the variables in the initial DfES 2002 LPD extract from the NPD prepared for the GLA.

**Figure 1. A simplified schematic of the London Pupil Dataset**



As with the NPD, LPD data allow for the creation of derived variables, such as pupil length of time on roll, distance from home to school, or level of progress from one key stage to the next. Given the relatively limited number of pupil level variables currently available in PLASC, this may have a particular significance for those accustomed to working with some of the more extensive local education authority (LEA) pupil datasets, or with school effectiveness datasets. As noted above, this is discussed further in section 12.

The LPD is held under terms and conditions set by DfES Analytical Services, designed, in part, to ensure that data confidentiality is preserved. Pupil level records from the LPD are not released to others, and those seeking pupil level data should approach the DfES directly. There may be scope for joint work where a research proposal, formulated in advance, has been given approval by the DfES and the GLA. These points are discussed further in section 13.

While the NPD and the LPD lend themselves to quantitative analysis, there is no reason why they should not be used in tandem with qualitative information, as suggested in Section 2. Either can suggest lines of inquiry for, and can supplement, the other. The type of work carried out at the Institute for Community Studies and by Mass Observation may not be as remote from today's policy needs as might be inferred from a casual reading of the BJS papers.

## 4. The advantages of pupil level data

'Pupil level data' means that each record is for an individual pupil. Strictly speaking the NPD is not populated entirely with pupil level data. For example, the majority of pupils are entered for more than one subject at GCSE. The DfES has now provided the GLA with more than 5,000,000 2003 GCSE records. Each record is a for an individual subject entry, such as an entry for English, or history or geography. A pupil entered for ten subjects would have ten separate records in this dataset. However, this information can be aggregated to provide a single record for each pupil. Additionally, school level data can be attached to pupil level records. The key point remains that records in the NPD are distinctively organised around individual pupil level data.

Appendix table A3 is an extract from the January 2000 Department for Education and Employment (DfEE) Form 7 questionnaire for secondary schools. Form 7 surveys extended to all maintained schools, and were largely replaced by PLASC in 2002. As the appendix table shows, Form 7 was used to gather aggregated information on basic items such as the total number of boys and girls in particular age groups at the time of the census.

Aggregated information is not necessarily useful where the requirement is to relate data which have already been grouped under separate headings. Data from the January 2000 Form 7 could not, for example, be used to provide a summary of pupil ethnicity by age. That summary can be provided using pupil level data, as long as data are gathered at that level on ethnicity and age, or on an age related measure such as date of birth.

Aggregated data provide answers to a specific set of predetermined questions, such as how many boys and girls there are on roll in schools. Pupil level data tend to be useful either where the full range of questions cannot be anticipated in advance, or where the range of questions leads to a questionnaire too large for respondents to complete. Indeed, one justification given for collecting pupil level data is that it reduces the burden on schools by removing the need for an LEA or the DfES to carry out multiple surveys. The mantra 'collect once, use often' would have been well known to those giving that justification.

However, even where information has been gathered under the heading the research analyst or research user is interested in, it still may not be possible to provide a borough, regional or national summary from aggregated school-level figures. A simple example is that averaging the average percentage of pupils gaining 5 or more higher grade passes in public examinations in individual schools does not give the true national average.

Equally importantly, analyses of data aggregated to the level of, for example, the school, may well conceal important variations between groups of pupils within schools. A simple example would be a single figure showing the average level of educational performance in a social diverse area, will mask low levels of attainment in some neighbourhoods and higher levels of attainment in others. A further example, which would be understood immediately and intuitively by teachers, is that the presence or absence of a limited number of school phobic children on a the roll of a school, which is otherwise successful in tackling truancy, can have a marked effect on attendance figures reported in national performance tables. A slightly more complex example, which also reflects the way in which aggregated data mask diversity, occurs when free school meal entitlement is correlated with education attainment.

A correlation coefficient indicates whether a relationship between two variables is more or less close to being linear. Average school examination scores vary less than the examination scores of individual children. A correlation of average school examination scores with average school levels of free school meal entitlement will give a stronger coefficient than a correlation based on pupil level data. Where the research user is interested in the relationship between individual pupil entitlement to free school meals and attainment in public examinations, a correlation based on aggregated school level data would need to be handled with caution, and could be misleading.

Additionally, while it is reasonable to expect schools to maintain records of children on their rolls, it would be unreasonable to expect them to provide some of the more complex statistical analyses needed where these can be provided more economically centrally, using pupil level data. Those 'central' analyses, whether carried out by the DfES, an LEA or another agency, can be used to provide the national or local comparator, against which individual schools can compare data for pupils on their roll. At one level this is an example of the 'collect once, use often' approach to pupil level data. More importantly, the return of information to schools, 'adding value' to their work with data, is consistent with the current emphasis on developing evidence-based policy and practice. The DfES *Pupil Attainment Tracker* provides a case of information

collected from schools being analysed to provide the national comparator, which is then returned to schools for use with local data in a tailor-made computing system (see www.standards.dfes.gov.uk/performance/pat/ ).

The appendix tables showing the type of information available from Form 7, and the range of information available now, point to a major change at national level in the data available for analysis. Realizing the potential this brings will require care and effort by both research analysts and research users. The remainder of the briefing focuses on these two points.

## 5. "Data in the NPD are strengthened by the involvement of a range of specialists" versus "Where is the researcher's user manual?"

A range of specialists in schools provides the data in the NPD as an adjunct to their main educational responsibilities, and can be expected to use that information in evidence-based work on school improvement. The point that these individuals are not first and foremost administrative data providers is key to understanding the NPD, and should also inform the approach taken by researchers seeking the co-operation of schools, LEAs or national agencies.

On balance the involvement of educational specialists has constructive implications for the quality of the data in the NPD, but it also has implications for data documentation. Data documentation exists, but may not be in the form expected by some researchers. The different data streams, which feed into the NPD, are discussed further below, but three points are raised now to illustrate what research analysts can expect by way of data documentation.

The Pupil Level Annual Schools Census (PLASC) is one key source of pupil data in the NPD, and Department for Education and Skills data definitions for the 2004 survey are currently available at www.teachernet.gov.uk/_doc/6029/PLASC2004-DataDefinitions2004v5.0.xls Appendix table A1 shows DfES 2003 PLASC data definitions. Item P54 refers to pupil special educational needs, and indicates that 'School action' is to be coded as 'A', and 'School action plus' is to be coded as 'P'. Special educational needs co-ordinators in schools will be aware of what those terms mean, and that they are set out in the DfES policy document *Special Educational Needs: Code of Practice*. This is currently available at www.teachernet.gov.uk/_docs/3724/SENCodeOfPractice.pdf ). However, there is no reference to the code of practice in the data definition document, nor is there any other explanation in that document of what 'school action' means to help a researcher new to this area.

For some, the year to year change in variables collected through surveys such as PLASC, will add further complexity. Those who work on NPD data for more than one year, face the possibility that understanding data definitions used in the first year will not provide a complete understanding of data definitions in subsequent years. Adding to this, it is now proposed that a school census should take place three times each year rather than once, and that different

data would be gathered at the different times of the year (see www.teachernet.gov.uk/docbank/index.cfm?id=8178).
A further potential change for the future was signalled in a January 2005 note from the DfES to local authority Chief Education Welfare Officers. This sets out the requirement that schools maintain pupil level attendance records in school management information systems from September 2006, with a view to it being collected in the January 2007 School Census. Chief Education Welfare Officers can be expected to be familiar with the background to this change. Researchers who do not have that understanding may be less certain. The change in monitoring pupil attendance is unlikely to be the only change to be introduced. Letters from the DfES to LEAs, of the sort sent to Chief Education Welfare Officers, are available at www.teachernet.gov.uk/LEAmailing, and other key information relevant to data being collected is currently available at www.teachernet.gov.uk/management.

The three examples of documents relating to data in the NPD point to code lists, documents giving advice to specialists in schools, and letters from national agencies flagging developments in particular aspects of data collection. In practice these are typical of the documentation available for the rest of the NPD. Those accustomed to working with research datasets, where a codified user manual is a main starting point, may well conclude that similar documentation should be available with the NPD. A problem in achieving this is that those who have the knowledge needed to provide such a manual have no incentive to do so. At this point some researchers may cast around for a less challenging area of work. A more positive response would be to explore the value of the existing documentation.

## 6. The advantages of the documentation available to education specialists

Section 2 places the NPD in the context of a broader development of evidence-based policy and practice. The implication is that the main purpose of the NPD is other than meeting the needs and expectations of individual researchers, including the need of some for a codified user manual.

Further key points are that the data flows which feed into the NPD tend to predate it, were designed for purposes other than populating the NPD, and that the provision of the data is a spin-off from the work of those involved, rather than being their main function. Guidance for a range of education specialists, which ultimately leads to data being made available for the NPD, is set in the context of the main, educational, responsibilities of those specialists. Research analysts who are familiar with that guidance will have a deeper understanding of the education issues involved than will those who do not. They are also likely to have a greater understanding of some of the pitfalls which await the unwary. Some of those pitfalls may appear trivial, but ignorance of them can lead to embarrassing errors.

## 7. Public examinations and key stage assessments. Pitfalls and guidance.

The importance of examination results to candidates and to schools means that both will, ultimately, check results. This has a beneficial influence on the quality of data. The complexity of the examination system and the curriculum means that schools will designate a senior member of staff to act as examinations secretary and that other senior members of staff will be appointed to act as curriculum leaders. These will receive printed guidance on arrangements which affect them, whether from the DfES, the Qualifications and Curriculum Authority or from examination boards.

In the case of public examinations, the NPD taps into a pre-existing data flow. This was originally intended to allow examining boards to receive candidate information and scripts from examination centres, and to return examination results to candidates at those centres. Examination centres are generally schools, and candidates are generally, *but not always*, pupils on roll in those schools.

Candidates taking examinations at any one centre can be on roll at other schools or at no school. Where the analyst's aim is to provide a measure of school performance based on the results of pupils on roll in a school, rather than on the results of candidates at particular centres, then some records will need to be discarded and others will need to be reallocated. DfES contractors, LEAs and schools spend time cleaning examination data before it appears in national performance tables. That work begins when school staff are on vacation, and may involve persuading head teachers that excellent results from candidates not on roll are not properly part of that school's record.

The consequence, which may surprise those not familiar with the examination system, is that 7,579 pupil examination records in version 2 of the 2002 LPD are not attached to any school. Version 2 of the 2002 NPD was assembled before schools and LEAs had finally removed guest candidates' results, reallocated results for consortium candidates, and checked the accuracy of grades. An attempt to use that data to recreate individual school examination performance tables would result in awkward errors.

Key stage co-ordinators in schools, whose work results in assessment information being fed into the various and separate assessment datasets in the NPD, also receive specialist printed guidance. In this case it is from the Qualifications and Curriculum Authority (QCA), which produces separate 100-page *Assessment and Reporting Arrangements* reports for each key stage. These cover a range of issues, including the assessment codes to be used. The 2005 QCA booklets are currently available at www.qca.org.uk/2627.html. Further advice on how assessment totals are calculated nationally are available in the DfES *Autumn Package* for schools, available at (www.standards.dfes.gov.uk/performance/ap/ ). Neither of these two sources of information provides all information on the national curriculum, but they do provide more than would be found in a simple code book, and both are readily accessible.

## 8. PLASC – a multiplicity of specialists. The case of ethnicity data and local agreements.

Section 5 pointed to special needs co-ordinators in schools and to Chief Education Welfare Officers in local education authorities as other people who receive specialist printed advice. The implication is that even within PLASC, which is a single data flow feeding into the NPD, there is more than one group of specialists providing information. Their responsibilities have a bearing on the data available for research, and an understanding of what this involves is important if data in the NPD are not to be misunderstood.

Specialist staff who manage and monitor funded ethnic minority projects provide a case in point. There is a single ethnicity variable in PLASC, and we might suppose that information was provided in a common framework for all pupils in all schools. This has not been so. The classification used in 2002 used fewer categories than the more complex 2003 survey. From 2003 the PLASC classification matched that of the 2001 national census, and LEAs negotiated 'extended' ethnic categories with the DfES. Extended categories matched back to less detailed 'subcategories', and could be used instead of those subcategories. The DfES asked schools to act on advice from their LEA when recording pupil ethnicity, which meant that recording arrangements varied from one LEA to another.

Table A4 shows the ethnic subcategories and extended categories used in 2003 PLASC, and the number of pupils in each. The dual heritage subcategory 'White and Asian', could, after negotiation, be used as a label in its own right or be divided into the 'extended' categories 'White and Pakistani', 'White and Indian' and 'White and any other Asian'. Nationally, some children with a dual White and Pakistani heritage would have been recorded as such while others would have been recorded under the more general 'White and Asian' heading. We cannot assume that the total of those listed as 'White and Pakistani' is the actual total of that group in the school population.

Nonetheless, we might have hoped that individual LEAs would have identified extended categories relevant to their area, and that practice would have been consistent within an LEA. If that were so, then pupil level data could be used to identify the major concentrations of those in extended categories such as 'White and Pakistani', even if some with that heritage are not recorded as such elsewhere.

Tables A5 and A6 show that the use of divisible or more precise categories varies both between, *and* within, LEAs. Information grouped under extended categories can only give an imperfect view of the ethnic composition of the school roll in some LEAs. Further, in London, children from the same neighbourhood can be educated in different LEAs, each with potentially different recording procedures. Information on numbers in the extended ethnicity categories may provide a poor guide to the finer detail of ethnicity amongst residents in particular areas.

This is not to say that the extended ethnicity information is of no value. If need be, it can be reclassified under broader subcategory headings which are themselves still more detailed than those used in the 1991 census. Nonetheless, it provides a further example of the point that not knowing the context in which information in the NPD has been gathered can lead to major misunderstanding.

## 9. The NPD – a multiplicity of national agencies, and more pitfalls for the unwary

A further complexity for some researchers arises from the involvement of more than one agency in the provision of advice to LEAs and schools, and their differing use of English.

> *'I don't know what you mean by "glory",' Alice said.*
> *Humpty Dumpty smiled contemptuously. 'Of course you don't –*
> *till I tell you. I meant "there's a nice knock-down argument for you!"*
> *'But "glory" doesn't mean "a nice knock-down argument",' Alice objected.*
> *'When I use a word,' Humpty Dumpty said, in a rather scornful tone, 'it means just what I choose it to mean – neither more nor less'.*
> (Lewis Carroll, *Through the Looking Glass,* Chapter 6).

The DfES for example has consistently required that pupil age be recorded as it would have been at the 31st August prior to the start of the school year, which is how it appears in the NPD. DfES reports reflect that practice. Public examination performance tables, for example, provide information on results for all 15 year-olds, meaning all pupils on roll who were 15 at the start of the school year in which examinations took place. The QCA has consistently referred to pupil age at the time when an assessment was taken. Additionally, assessment results are for those assessed and not for all pupils in an age group. The result is, depending on whose report is being read, that key stage 1 results are therefore either for pupils aged 6, or pupils aged 7, though the pupils in question are the same children. This is a small point, but there have been cases where politicians have pointed to the very low level of examination success amongst 16 year-olds in some areas, without realizing that the relevant figures were those for 15 year-olds.

The reality, then, is that documentation is available for data from the NPD, but available as advice for specialists in schools, and in a variety of documents. It is not available as a single codified user guide for research analysts. Additionally, those responsible for providing that diverse documentation are targeting a range of specialists in schools, for whom the provision of data is a subsidiary spin-off of their work as educationists. The available documentation contains information on coding, but is not mainly about coding. Further, agencies involved in providing advice can use the same terms to mean different things.

The processes through which information is made available to the NPD, results in a dataset which differs from standard research datasets of the type held in the UK Data Archives at the University of Essex, and which also differs from other widely known datasets such as the National Census. There is no single

'form' for collecting information, designed by a single group of people, specifically to be used in the collection of information from the generality of the population.

The point was made earlier that the involvement of education specialists in the data flows which feed the NPD should improve data quality. We might add that establishing the provision of data as a subsidiary responsibility of education specialists, ties the NPD closer to the ebb and flow of education in schools than would otherwise be the case. This may well appeal to research analysts concerned with the education and policy issues evident in, for example, the national curriculum and in more specific policies such as Excellence in Cities.

## 10. Data management – time

Large datasets take time to assemble and to clean. In the January 2003 PLASC in England, 22,457 primary, secondary and special schools provided a total of 7,710,940 pupil records. Of these, 1,062,780 pupils attended schools in London. Checking that number of records for accuracy is a substantial undertaking. Linking PLASC records with pupil assessment records at key stages 1, 2, 3 and 4 will likewise be a substantial undertaking, and checking for duplicate records is also time consuming.

Analysing that volume of data also takes time and, for practical purposes, would be impossible without computing facilities and the appropriate software. The expectations cited earlier from within DfES Analytical Services, and from within the Cabinet Office, indicate that analyses which are *never* more than arithmetical counts will lack credibility. Analysts will need to use computing systems which are capable of the level of work now required. The LPD is analysed using SPSS, and what follows in this and the next section reflects that experience. Additionally, comments in both sections reflect the point that the GLA works with data from successive years of the NPD.

The majority of the data in the NPD are either alphanumeric, using numbers and text for a case, or numbers for some cases and text for others, or are wholly in a text format. SPSS works more efficiently with numeric data, and in many instances will only work with numeric data. Output can be made meaningful to the reader by attaching labels to numeric codes.

There is more than one way of achieving this, but all involve time, and some involve considerable amounts of time. Coding and labelling can be achieved by merging a dataset from the NPD with 'lookup tables', which contain numeric codes with labels attached. This is potentially the most time-consuming approach, but may in some instances be the only approach available. In a second approach, numeric codes can be added directly to NPD data using pre-written syntax files, perhaps with labels being attached using the data dictionary from a file created in an earlier year. This is also time consuming, but can be can be less so than merging files where the link is based on alphanumeric data. However, the second approach depends on NPD data remaining the same from one year to next, and experience suggests that change should be expected. In

some instances it may be possible to use the SPSS Automatic Recode facility. This is, comparatively, quick, but in practice there are few opportunities to use it with NPD data. Preparing the LPD for analysis is a project in itself, and time needs to be allowed for that. Table A1 gives a rough measure of the extent to which data are alphanumeric and will need to be recoded.

Those with experience of planning and carrying out projects are likely to agree that successful work requires a focussed set of priorities. Project management, as understood for example in the management of a building project, will also assume that a project will have a time-line, with phases being completed on time according to a pre-determined sequence. The assumption that research should be carried out in phases at specific times leads to difficulties for work with the NPD.

The timing of data release by the DfES can vary, unpredictably, from one year to the next, and this puts obstacles in the way of any simple critical path analysis version of project management. Research users and research analysts need to be aware of this and are cautioned against making commitments, in advance of NPD data being received, about the time when analyses will be delivered.

The times at which data are released also have implications for the types of analysis which are likely to be needed. School governing bodies and head teachers look for analyses of 'their own' assessment data in the period immediately after the release of assessment results. It is during this period that analyses for individual schools stands its best chance of influencing practice. Later in the year, discussion within schools moves to other issues. The GLA did not receive 2003 assessment data until 2005, which is well outside the window of opportunity for providing school-level analyses to schools. Those using NPD data will not be in a position to provide reports with the degree of immediacy some might like to see. While this hardly means that there is no scope for valuable analysis of the NPD and LPD, it remains the case that the time at which NPD data can be released by the DfES, will have an effect on the range of analyses wanted from researchers and analysts.

## 11. Data management - computing capacity

The 1980's were marked by the arrival of the (comparatively) low cost desktop personal computer. This brought to individuals and to smaller organisations a computing capability which had previously been the preserve of large organisations, or of those with large budgets, using mainframe computers. Experience since the 1980's has been that desktop pcs increased in power, and increased the scope for analysis in both the public and private sector. If only temporarily, the NPD has in a sense thrown that trend into reverse. The NPD, taken as a whole, is well beyond the reach of today's personal computers.

The NPD contains records for approximately 8,000,000 pupils in any one year, of which approximately one in eight are in the LPD. For projects which work with linked data from more than one year, datasets will be larger than datasets for a

single year, as will datasets which incorporate information from elsewhere. At the least, research analysts and research users will need to be aware that the computing capacity available for a project needs to be taken into account at the project design stage. The merged 2002 2003 LPD, which contains a restricted list of variables, is in excess of 8 Gigabytes.

Faced with what are, by today's standards, potentially very large datasets, one option is to re-learn lessons from earlier days of computing, when constraints on file size were greater than they are now. One possible solution is for analysts to be highly selective in the data they choose to use at any one moment in time. A large SPSS file, for example, can be split into separate files, with subsets of data tailored to different aspects of the work being covered. If necessary, those files can be remerged if each record has been given a unique identifier, perhaps created through the SPSS compute id=$CASENUM procedure, before the file is split.

The drawback is that subsets of data tend to diverge as each is modified in different ways. This can lead to awkward discrepancies, for instance in totals from the different subsets of data. Additionally, the storage capacity for data will be sizeably greater than if files are not split.

Another possibility is that more time-consuming analyses are run overnight, possibly in a batch mode, or that analysts work on other computers while analysis is under way. A further possibility is to use database software other than SPSS. Relational database software reduces the demands placed on computing capacity at any one moment in time. BathData, for example, organizes the large volume of information used in national school performance tables and also works closely with the LEA-funded National Consortium for Examination Results (NCER, see www.ncer.org and www.bathdata.com/national_data.php ). The NCER has successfully released examination data to LEAs as relational databases for well over a decade, and more recently has also released key stage assessment data in the same form.

There is no standard relational database software which is pre-programmed with the range of statistical tests available in, for example, SPSS. Creating those programmes would have resource implications as great as, or greater than, the resource implications involved in purchasing high specification computers and appropriate statistical software. It is perhaps worth remembering that the American political scientist Norman H. Nie spent a considerable amount of time in the late 1960's programming statistical tests in Fortran, concluded that making these available to others would save them an equally considerable amount of time and, with others, produced ….. SPSS!

There is at present no low cost solution available to those working with larger NPD extracts.

## 12. The NPD does not contain everything. The 'curious' absence of school effectiveness data.

The decision to establish the NPD followed the decision to include a measure of individual pupil progress in national performance tables, partly based on *'The Value Added National Project Final Report. Feasibility studies for a national system of Value Added Indicators* (School Curriculum and Assessment Authority, 1997). National performance tables give information on comparative levels of achievement in different schools. The inclusion of value added measures in those tables may be seen as providing a 'fairer' account of a school's achievement, because value added takes account of the reality that different schools have different types of intake. Some schools, for example, will have pupils with low levels of attainment when they first join, and this will have a bearing on their subsequent level of attainment.

However, a key conclusion in earlier research suggests that value added measures would be relevant to the main task of comparative performance tables in another sense. That conclusion was that schools with 'similar' intakes, nonetheless have different levels of subsequent attainment (see Michael Rutter, Barbara Maughan, Peter Mortimore and Janet Ouston *Fifteen Thousand Hours, Secondary Schools and their effects on Children,* Open Books, 1979).

Work to identify factors which 'make a difference' in schools, and work to quantify the extent of the difference a school can make, progressed after 1979 under the general heading of school effectiveness research. Over a period of more than 20 years, school effectiveness researchers have been accustomed to working with a range of information, including pupil social class and pupil family size, as well as a range of outcome measures such as pupil attendance, behaviour and delinquency. Other standard measures in school effectiveness research have included teacher expectations, teacher style and measures of school ethos. Additionally, and partly influenced by school effectiveness research, LEAs at various times, and in various places, have gathered information on pupils joining and leaving schools between the dates of the January pupil counts, on pupil religion, and on pupil home language and pupil fluency in English.

None of this information was included in the 2002 and 2003 NPD, despite its potential relevance for assessing whether a school was more, or less, effective, and despite the relevance of that to the key aim of national performance tables. Type of pupil SEN was not included in PLASC until January 2004, and information on children's looked after status, though collected in 2003, was of such poor quality that it has not been released. Detailed information on language spoken was to have been collected in 2005, but that aim has not been realized. Information on fluency in English has not been collected and will not be collected in 2005.

The terms of reference given to the National Value Added Project (NVAP) may have played a key role in restricting the range of data initially available within the NPD. NVAP consultations took place with teachers and LEAs, but not with parents, and it was a requirement that data collection arrangements did not

place an undue burden on schools. The focus was to be on key stage 2 and key stage 4. This is, it was to be on points in time when pupil were assessed.

Crucially, measures of value added were to be used for public accountability purposes. National 'raw score' school performance tables had been introduced in 1992 and, as noted above, these had been widely criticised by educationists as being 'unfair'. Perhaps inevitably in this context, educationists came to see value added as potentially correcting the unfairness of league tables, rather than seeing it in terms which might have been expected from of a reading of school effectiveness research. Interestingly, Trevor Knight, who was a member of the Value Added National Project Advisory Group, more recently wrote that

> *There is now a widespread feeling amongst schools and LEAs (and the academic community) that some contextual measures have a statistically significant and educationally justified impact on the explanation of variation in outcomes at all key stages.*
> *(*In Ian Schagen and Karen Elliot (eds) *But What does it mean?* NFER, 2004, page 76*)*

A wider range of data was available in the NPD in 2004 than in 2002. There is scope for change.

## 13. The wider use of data from the NPD – protecting researchers

DfES Analytical Services acts as a gatekeeper for those who wish to have access to data from the NPD. Applications for access must include a statement about the research project in hand, and the use of data is restricted to the stated aims of that project. Additionally, it is a requirement that data are held securely, rather than as an open access resource. Approval from DfES Analytical Services is required prior to the publication of any analyses of data from the NPD. None of this should be new to, or should alarm, professional researchers. Despite this, there are at least two problems that can occur with or for research analysts.

The first is evidenced in the claim by (some) London LEA officers that their role should entail automatic access to data from other LEAs. In practice, data in the NPD are collected on a statutory basis, rather than on the basis of agreement between LEAs. The requirement for schools to provide a PLASC return, for example, is based on section 537A of the 1996 Education Act. The statutory framework surrounding the collection of data for the NPD does not place ownership of the data with local authorities. While individual LEAs have access to data from the schools they maintain, access to data from other LEAs is properly by application to DfES Analytical Services. The opportunity to apply for access to the wider data exists for LEA staff equally with staff in universities and elsewhere. Despite this, (some) LEA staff have seen current arrangements as obstructive. This is an overly negative view, as an assessment of the second problem indicates.

The second problem follows on from the 'automatic access' fallacy, and might be called the technological fallacy. In the 'Information Age', some will assume

that if the technology for data analysis and transfer exists, then it should be used. By contrast, professional research analysts can be expected to understand, and in most cases act on, agreements involving considerations such as data confidentiality. The GLA has a substantial presence in research and statistical analysis in London, and has a code of ethics and standards which *all* staff are *required* to observe. The code covers issues such as maintaining the confidentiality and security of information.

It is reasonable to expect that organisations which wish to receive pupil level data should have equivalent codes. However, where that is not the case, the DfES Analytical Service's terms and conditions can be invoked if necessary, and will provide the research professional with at least a measure of protection. Additionally, the requirement by DfES Analytical Services for a specific research proposal from those seeking access to data also sets a limit to the demands which can legitimately be placed on the researcher. DfES terms and conditions of access to NPD data provide professional research analysts with a framework within which research proposals can be developed, *and* with a measure of protection.

## 14. Conclusion

The NPD belongs to the age of evidence-based policy and practice, and is relevant to the impact analysis welcomed from within the Cabinet Office and from within DfES Analytical Services. It is also relevant to analyses of who gets what in society, when where and why, which will be of interest to agencies such as the GLA where equality and inequality issues, and social inclusion and exclusion are policy priorities. That is, the NPD and LPD can both be approached to provide analyses of the political economy or political arithmetic of education. Additionally, while both datasets are relevant to work on school improvement, the data available are not as extensive as might be expected by those grounded in school effectiveness research. The NPD and LPD also contain information relevant to spatial analysis and regeneration. The partially longitudinal nature of the NPD and LPD is a particular strength, and there may be scope for developing the NPD as new research user needs emerge.

The NPD and LPD hold data on the population of children attending maintained schools, and therefore avoid pitfalls associated with sample surveys. Additionally, because each is updated annually, the NPD and LPD provide a more up to date view than either older surveys or recurrent, but less frequently updated, surveys. This is important in areas such as London, where large-scale change can take place over short periods of time. Since both hold pupil level data, the NPD and LPD allow for more accurate and sensitive statistical analyses than is the case where data have been aggregated to school, neighbourhood, or borough level.

The NPD benefits from the involvement of a range of specialists at national, local and school level. This should add a further quality check on data. However, the resulting multiplicity of legislation, codes of conduct, and reporting arrangements will present a challenge to some researchers. For those who

have come to expect that a codified user manual will be available as their guide to data, the challenge may simply be too great. For others, this briefing has pointed to documents which shed light on data in the NPD. These are readily accessible, though it should be remembered that individual documents are not kept on the Web in perpetuity.

In addition to documentation available via the Internet, other advice may be available where researchers are able to establish effective working relationships with teachers, LEA Officers or civil servants at the DfES. Numerically, those individuals are the main users of pupil level data of the type reaching the NPD. Effective working relationships are more likely to develop with such people where researchers are aware that the majority of these individuals are not primarily data providers or contributors to an 'administrative' dataset. The NPD is ***not*** an administrative dataset. It is a product of an era of evidence-based policy and practice.

Researchers new to data of this type will need to accept that acquiring an understanding of the data, and of its pitfalls, will involve effort and initiative on their part, and are cautioned against assuming that others have a duty to provide a UK Data Archive-type research user manual. Changes in the data collected in the NPD each year mean that an awareness of where explanations of data can be found, and establishing effective working relationships with those who use information in the NPD, are both investments worth making.

The size and complexity of the NPD means that data checking is time-consuming, with pitfalls along the way. One result is that data have been released from the NPD to the GLA at widely varying times of the year. Additionally, once released, raw data still requires time-consuming recoding and labelling, with further time required for the creation of derived variables. Each of these factors needs to be allowed for. The size of the datasets means that the NPD raises issues of computing capacity for those who wish to work with it. Desktop personal computers may not provide that capacity if work is with larger datasets. The NPD will present challenges for those unaccustomed to working with large, complex datasets.

Data are released from the NPD under terms and conditions set by DfES Analytical Services. These may be viewed as constraining what researchers can do with, and research users can get from, the NPD. The alternative, and more positive, view is that the terms and conditions which apply provide a framework within which serious research proposals can be assembled, ***and*** provide a measure of protection for individual researchers.

## Appendix. Key reference tables.

### A1. PLASC 2003 data definitions

| Ref | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| P1 | Unique Pupil Number (UPN) | UPN | 13 | Format: AnnnnnnnnnnnA (for temporary UPN). Any alphabetical character but cannot be I, O, or S. To calculate the check letter:<br>1. Multiply the individual digits by their weights as follows:<br>digit 2 by weight 2; digit 3 by weight 3; digit 4 by weight 4; digit 5 by weight 5; digit 6 by weight 6; digit 7 by weight 7; digit 8 by weight 8; digit 9 by weight 9; digit 10 by weight 10; digit 11 by weight 11; digit 12 by weight 12; digit 13 by weight 13.<br>2. Sum the individual results, divide the total by 23, and take the remainder.<br>3. Calculate the check letter from the result as follows:<br>0 = A; 1 = B; 2 = C; 3 = D; 4 = E; 5 = F; 6 = G;<br>7 = H; 8 = J; 9 = K; 10 = L; 11 = M; 12 = N; 13 = P;<br>14 = Q; 15 = R; 16 = T; 17 = U; 18 = V; 19 = W; 20 = X;<br>21 = Y; 22 = Z.<br>For calculating the check letter (see above) any alphabetical character at digit 13 is accorded the same numerical value as listed in paragraph 3 above, i.e. A=0, B=1, C=2 etc. |
| P2 | Pupil's former UPN | FormerUPN | 13 | Format: AnnnnnnnnnnnA (for temporary UPN). For calculating the check letter (see above) any alphabetical character at digit 13 is accorded the same numerical value as listed in paragraph 3 above, i.e. A=0, B=1, C=2 etc. |
| P3 | Pupil Surname | Surname | 30 | Any character string. The legal last name of the child (as written). |
| P4 | Pupil Forename | Forename | 15 | Any character string. The full first name of the child (not common contractions). |
| P6 | Pupil Middle Names | MiddleNames | 25 | Any character string. |

## A1. PLASC 2003 data definitions, continued

| Ref | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| P7 | Pupil Date of Birth | DOB | 10 | Format CCYY-MM-DD valid date. Implied age at previous 31 August would normally be in the range:<br>Primary schools 2-12; Middle schools 7-14;<br>Secondary schools 10-19; Special schools 2–25. |
| P8 | Pupil Gender | Gender | 1 | Table values: F = Female M = Male. |
| P9 | Pupil former Surname | FormerSurname | 30 | Any character string. If multiple values, the most recently used. |
| P11 | Pupil Preferred Surname | PreferredSurname | 30 | Any character string. The last name (as written) commonly used in the school. |
| P13 | Pupil Ethnic Code | Ethnicity | 3 | Table values: WBRI = BRITISH; WENG = ENGLISH; WSCO = SCOTTISH; WWEL = WELSH; WIRI = IRISH; WIRT = TRAVELLER OF IRISH HERITAGE; WOWB = OTHER WHITE BRITISH; WOTH = ANY OTHER WHITE BACKGROUND; WALB = ALBANIAN; WBOS = BOSNIAN-HERZEGOVINIAN; WCRO = CROATION; WGRE = GREEK/GREEK CYPRIOT; WGRK = GREEK; WGRC = GREEK CYPRIOT; WKOS = KOSOVAN; WITA = ITALIAN; WPOR = PORTUGESE; WSER = SERBIAN; WTUR = TURKISH/TURKISH CYPRIOT; WTUK = TURKISH; WTUC = TURKISH CYPRIOT; WEUR = WHITE EUROPEAN; WEEU = WHITE EASTERN EUROPEAN; WWEU = WHITE WESTERN EUROPEAN; WOTW = OTHER WHITE; WROM = GYPSY/ROMA; MWBC = WHITE AND BLACK CARIBBEAN; MWBA = WHITE AND BLACK AFRICAN; MWAS = WHITE AND ASIAN; MWAP = WHITE AND PAKISTANI; MWAI = WHITE AND INDIAN; MWAO = WHITE AND ANY OTHER ASIAN BACKGROUND; MOTH = ANY OTHER MIXED BACKGROUND; MAOE = ASIAN AND ANY OTHER ETHNIC GROUP; MABL = ASIAN AND BLACK; MACH = ASIAN AND CHINESE; MBOE = BLACK AND ANY OTHER ETHNIC GROUP; MBCH = BLACK AND CHINESE; MCOE = CHINESE AND ANY OTHER ETHNIC GROUP; MWOE = WHITE AND ANY OTHER ETHNIC GROUP; MWCH = WHITE AND CHINESE; MOTM = OTHER MIXED BACKGROUND; AIND = INDIAN; APKN = PAKISTANI; ABAN = BANGLADESHI; AOTH = ANY OTHER ASIAN BACKGROUND; AMPK = MIRPURI PAKISTANI; AOPK = OTHER PAKISTANI; AKPA = KASHMIRI PAKISTANI; AAFR = AFRICAN ASIAN; AKAO = KASHMIRI OTHER; ANEP = NEPALI; ASNL = SINHALESE; |

## A1. PLASC 2003 data definitions, continued

| Ref | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| Pupil ethnicity (continued) | | | 3 | ASLT = SRI LANKAN TAMIL; AOTA = OTHER ASIAN; BCRB = CARIBBEAN; BAFR = AFRICAN; BOTH = ANY OTHER BLACK BACKGROUND; BANN = ANGOLAN; BCON = CONGOLESE; BGHA = GHANAIAN; BNGN = NIGERIAN; BSLN = SIERRA LEONIAN; BSOM = SOMALI; BSUD = SUDANESE; BAOF = OTHER BLACK AFRICAN; BEUR = BLACK EUROPEAN; BNAM = BLACK NORTH AMERICAN; BOTB = OTHER BLACK; CHNE = CHINESE; CHKC = HONG KONG CHINESE; CMAL = MALAYSIAN CHINESE; CSNG = SINGAPOREAN CHINESE; CTWN = TAIWANESE; COCH = OTHER CHINESE; OOTH = ANY OTHER ETHNIC GROUP; OAFG = AFGHANISTANI; OARA = ARAB; OEGY = EGYPTIAN; OFIL = FILIPINO; OIRN = IRANIAN; OIRQ = IRAQI; OJPN = JAPANESE; OKOR = KOREAN; OKRD = KURDISH; OLAM = LATIN AMERICAN; OLEB = LEBANESE; OLIB = LIBYAN; OMAL = MALAY; OMRC = MOROCCAN; OPOL = POLYNESIAN; OTHA = THAI; OVIE = VIETNAMESE; OYEM = YEMENI; OOEG = OTHER ETHNIC GROUP; REFU = REFUSED; NOBT = INFORMATION NOT OBTAINED (Default);<br>May be blank if age at previous 31 August <5,<br>Table values (old, only valid for exclusions): 10 = White, UK heritage; 11 = White, European; 12 = White, other (known); 19 = White, type not known; 20 = Black, Caribbean heritage; 21 = Black, African heritage; 22 = Black, other; 30 = Indian; 40 = Pakistani; 50 = Bangladeshi; 60 = Chinese; 90 = Other (known); 98 = Parent/pupil preferred not to say; 99 = Ethnic group information not sought. |
| P14 | Source of Pupil Ethnic Code | EthnicitySource | 1 | Table values: C = Child; P = Parent; S = Current School;<br>T = Previous School; O = Other. |
| P25 | Pupil Free School Meal Eligibility | FSMeligible | 1 | true/false or 1/0. |

## A1. PLASC 2003 data definitions, continued

| Ref | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| P28 | Pupil In Care Indicator | InCare | 1 | true/false or 1/0. Is the child "looked after" on the census day? See Additional Explanatory Notes (3a). |
| P29 | Pupil In Care - Caring Authority Code | CareAuthority | 3 | Must not be blank if <InCare> or <InCareAtCurrentSchool>= true. Table values = LEALIST.XLS. See Additional Explanatory Notes (3b).<br>XXX allowed if <InCare>=false and <InCareAtCurrentSchool>=true. |
| P30 | Pupil In Care while at current school Indicator | InCareAtCurrentSchool | 1 | true,false or 1/0. Has the child ever been in care while at this school? |
| P34 | Pupil Enrolment Status | EnrolStatus | 1 | Table values (Default=C): C = Current (Single registration); M = Current main (Dual registration); S = Current subsidiary (Dual registration); G = Guest pupil. Note all values allowed for all school types. |
| P37 | Pupil Date of Entry | EntryDate | 10 | Format CCYY-MM-DD. Valid date before or equal to ReferenceDate. |
| P44 | Pupil Boarder Indicator | Boarder | 1 | Table values: B = Boarder, nights per week not specified; 6 = Boarder, six nights or less a week; 7 = Boarder, seven nights a week; N = Not a boarder<br>Primary, Middle and Secondary Schools: B or N. Special schools: 6, 7 or N. |
| P42 | Pupil Part-Time Indicator | PartTime | 1 | true/false or 1/0. For dual-registered pupils, time in other schools should be taken into consideration; e.g. 3 full days in one school and 2 full days in another school should not be classified by either school as part-time. *Change: PartTime=true now allowable for all ages* |
| P45 | Pupil's actual NC Year Group | NCyearActual | 2 | The yeargroup whose curriculum the child is following (as distinct from the yeargroup appropriate to his/her date of birth)<br>Acceptable values: N1, N2 (Nursery), R (= reception), 1-14. Should be<br>N1 if age at previous 31 August < 3,<br>N2 if 3 <= age at previous 31 August < 4,<br>R if 4<= age at previous 31 August < 5. |
| P46 | Pupil Class Type Indicator | ClassType | 1 | Table values: N = Nursery Class; O = Other Class |

## A1. PLASC 2003 data definitions, continued

| Ref | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| P50 | First Language | FirstLanguage | 3 | Table values: ENG=English, ENB=Believed to be English, OTH=Other, OTB=Believed to be Other plus (for PLASC 2003 onwards) REF=Refused, NOT=Information not obtained. Possibly values from new Language lookup table to be permitted. |
| P54 | Pupil SEN Stage | SENstatus | 1 | N (No special provision), A (School Action), P (School Action Plus), Q(School action plus and statutory assessment), S (Statemented). Default N. For exclusions only, allow 0-5 (Default 0). |
| P68 | Pupil Exclusion Start Date | StartDate | 10 | Format CCYY-MM-DD. In previous academic year (dates 01/09/2001 to 31/08/2002 inclusive).<br>Exclusion start date is the school day after<br>(1) the appeal committee's confirmation of permanent exclusion, or<br>(2) expiry of the time allowed for an appeal to be made, or<br>(3) the taking up by the pupil of a place elsewhere (if this occurs earlier). |
| P81 | Postcode | Postcode | 8 | Valid Postcode formats are: An nAA; AAn nAA; AnA NAA; Ann nAA; AAnA nAA; Aann nAA, where 'A' is an alphabetical character and 'n' is a numeric character. Double spaces should be converted to a single space.<br>Convert any 'O' at start of second postcode block to '0' (zero). |
| P01X | Connexions Assent | ConnexionsAssent | 5 | Yes, No, UNS=Unsought (Default), SNR=Sought, No Reply. Do parents consent to pupil data being shared with Connexions? |
| P135 | Post Advanced Level Indicator | <PostAdvanced> | 5 | true, false or 1/0. Whether taking a post A-level (or equiv) course<br>To be omitted for pupils not in year (NCyearActual) 12 or above (a*lso applies to P136, P138, P139, P140, P141, P142)* |
| P136 | Number of GCE A Level Subjects being studied | <Alevel> | 1 | Number of A levels (Including AS, A2). No separate category A/S. General Studies qualifications to be omitted. |
| P138 | Number of GCSE Subjects being studied | <GCSE> | 2 | Number of GCSEs |
| P139 | GNVQ Level | | 1 | 0=Not taking, 1=Foundation, 2=Intermediate, 3=VCE (3,6 or 12 units) |

## A1. PLASC 2003 data definitions, continued

| Ref | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| P140 | GNVQ Precursor Level | <PreGNVQ> | | 0=Not taking, 2=BTEC First or C&G Dip. of Voc. Ed (intermediate) 3=BTEC National or C&G Dip. Of Voc. Ed. (National) |
| P141 | NVQ Level | <NVQ> | 1 | 0=Not taking, 1=Level 1 NVQ, 2=Level 2 NVQ, 3=Level 3 NVQ |
| P142 | Other Post-16 Courses Indicator | <Other> | 1 | N=Not taking, B=International Baccalaureat, O=Other |
| S1 | LEA Number | LEA | 3 | Format = nnn. (character field)Table values = LEALIST.XLS |
| S2 | DfES Establishment Number | Estab | 4 | Format = nnnn. (character field)Table values = ESTABLST.CSV |
| S03X | Special School Status | SpecialStatus | | 1=Maintained, 2=Non-maintained. |
| S5 | School Name | Name | | 100 |
| S7 | Phase | Phase | 2 | Table Values:<br>PS = Primary<br>MP = Middle (Deemed Primary)<br>MS = Middle (Deemed Secondary)<br>SS = Secondary (including CTC and City Academy) |
| S01X | Type | Type | | 01 = first school, 5-8; 02 = first school, 5-9; 03 = first school, 5-10; 04 = first and middle school, 5-12; 05 = middle school, 8-12, deemed primary; 06 = middle school deemed secondary; 08 = middle school, 10-13, deemed secondary; 09 = comprehensive upper school, 12-15/16; 10 = comprehensive upper school, 12-18; 11 = comprehensive upper school, 13-16; 12 = comprehensive upper school, 13-18 11; 18 = infant and junior school, 5-11; 21 = comprehensive all-through, 11-16; 22 = comprehensive, 11-13, automatic transfer; 26 = junior comprehensive, 11-14, automatic transfer; 27 = junior comprehensive, 11-16, optional transfer at 13; 28 = junior comprehensive, 11-16, optional transfer at 14; 29 = senior comprehensive, 13-16, automatic transfer; 30 = senior comprehensive, 13-18, optional transfer; 31 = 32 = senior comprehensive, 14-18, optional transfer; 33 = senior comprehensive, 14-18, automatic transfer; 36 = non-comprehensive secondary – modern; 37 = non comprehensive secondary - technical; 39 = non-comprehensive secondary – other; |

## A1. PLASC 2003 data definitions, continued

| Ref | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| SO1X | School type - continued | | | 41 = middle school, 10-14, deemed secondary; 42 = first school, 5-7; 43 = first school, 7-10; 44 = comprehensive upper school, 14/15-18; 45 = middle school, 9-12, deemed primary; 46 = comprehensive, middle and upper, 10-16; 47 = City Technology College; 48 = comprehensive upper school, 11-16; 49 = City Academy; 52 = maintained special (day); 53 = maintained special (boarding); 54 = maintained special (hospital); 55 = non-maintained special (day); 56 = non-maintained special (boarding); 57 = non-maintained special (hospital); 61 = ex-GM special (day); 62 = ex-GM special (boarding); 63 = ex-GM special (hospital) |
| S9 | Minimum Year Group | LowestNCyear | 2 | Values = N1, N2, R, 1-14. The lowest year group for which the school customarily makes provision. *Exceptional* pupils/situations eg 'early entry' or 'held back', should not influence these values. |
| S10 | Maximum Year Group | HighestNCyear | 2 | Values = N1, N2, R, 1-14. See above. |
| S11 | Intake Type | Intake | 4 | Table Values:<br>COMP = Comprehensive<br>SEL1 = Selective (Grammar)<br>SEL2 = Selective (Secondary Modern)<br>SEL3 = Selective (Technical)<br>SEL4 = Selective (Religion)<br>SPEC = Special |
| S12 | Status | Governance | 2 | Table Values: CO = Community VA = Voluntary Aided VC = Voluntary Controlled FO = Foundation IN = Independent NM = Non-Maintained CT = City Technology College CA =City Academy |
| S24 | School Email address | Email | 60 | Format: string1@string2<br>string1 and string2 may include any characters except @. May be missing if School does not have an e-mail address. |
| S02X | Head Teaching Indicator | HeadTeachingIndicator | 1 | 1 = None; 2 = Averages less that 50%;<br>3 = Averages between 50% and 100%; 4 = 100% |
| S34 | Special School Organisation | Accommodation | 1 | Table values: D = Day Pupils (mainly) B = Boarding Pupils (mainly) H = Hospital Special School |

## A1. PLASC 2003 data definitions, continued

| Ref. | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| S35 | Maximum Day Pupils | MaxDayPupils | 4 | Maximum number of Day Pupils for whom the school is approved to make provision.<br>Must be >0 if <Accommodation> = D or H |
| S36 | Maximum Boarding Pupils | MaxBoarders | 4 | Maximum number of Boarding Pupils for whom the school is approved to make provision.<br>Must be missing if <Accommodation> = H<br>Must be >=0 if <Accommodation> = D<br>Must be >0 if <Accommodation> = B |
| S37 | Minimum Age - Boys | MinMaleAge | 2 | Minimum age for which the school is approved to make provision for Boys.<br>Must be greater than or equal to 2 and less than 15.<br>Must be present if school is coeducational or boys only. |
| S38 | Minimum Age - Girls | MinFemaleAge | 2 | Minimum age for which the school is approved to make provision for Girls.<br>Must be greater than or equal to 2 and less than 15<br>Must be present if school is coeducational or girls only. |
| S39 | Maximum Age - Boys | MaxMaleAge | 2 | Maximum age for which the school is approved to make provision for Boys.<br>Must be greater than or equal to 2 and less than 15<br>Must be present if school is coeducational of boys only. |
| S40 | Maximum Age - Girls | MaxFemaleAge | 2 | Maximum age for which the school is approved to make provision for Girls.<br>Must be greater than or equal to 2 and less than 15.<br>Must be present if school is coeducational or girls only. |
| S41 | Special School Provision Type | Provision | 4 | Type of special need for which the school is formally approved to make provision.<br>Table values:<br>VI = Visual Impairment<br>HI = Hearing Impairment<br>SPLD = Specific Learning Difficulties<br>SPCH = Speech and Language Disorders<br>AUT = Autism<br>EBD = Emotional and Behavioural Difficulty<br>MSI = Multi-sensory Impairment<br>PD = Physical Difficulties<br>MLD = Moderate Learning Difficulties<br>SLD = Severe Learning Difficulties<br>PMLD = Profound and Multiple Learning Difficulties<br>OTH = Other |
| S94 | Survey Reference Date | ReferenceDate | 10 | Must be 2003-01-16 for ASC return but should be variable to allow other returns/comparisons. |

## A1. PLASC 2003 data definitions, continued

| Ref. | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| S101 | Person Completing Survey | CompleterRole | 2 | Table values:<br>HD = Headteacher<br>DH = Deputy Headteacher<br>OT = Other Teacher<br>SS = School Secretary / Clerical Assistant / Administrative Assistant<br>BR = Bursar<br>ON = Other Non-Teacher |
| S102 | Survey Completion Time | Hours | 2 | Number of whole hours. 0-29 minutes to be rounded down and 30-59 minutes to be rounded up. Range = 1-99. Must not be blank if <Completer> not blank. |
| S138 | Free Meals Taken | FreeMeals | 3 | The number of Free School Meals actually taken on the census day. |
| S109 | Admission appeals lodged | Lodged | 3 | Integer. Total number of admission appeals lodged by parents in the academic year. |
| S110 | Admission appeals withdrawn | Withdrawn | 3 | Total number of admission appeals withdrawn by parents before reaching an Independent Appeals Committee. Integer. Default to 0 (zero) if <Lodged> = 0 (zero). |
| S111 | Admission appeals heard by Independent Admissions Committee | Heard | 3 | Total number of admission appeals heard by an Independent Appeals Committee. Integer. Default to 0 (zero) if <Lodged> = 0 (zero). |
| S112 | Admission appeals heard by Independent Admissions Committee - decided in parent's favour | Upheld | 3 | Total number of admission appeals decided in the parent's favour. Integer. Default to 0 (zero) if <Heard> = 0 (zero). |

## A1. PLASC 2003 data definitions, continued

| Ref. | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| S113 | Admission appeals heard by Independent Admissions Committee – rejected | Rejected | 3 | Total number of admission appeals rejected. Integer. Default to 0 (zero) if <Heard> = 0 (zero). |
| S166 | Class Reference | *Name* | 30 | *Any String* |
| S06X | Class Type | Category | 1 | O (=Other) or N (=Nursery). See P 46 above. |
| S07X | Class Yeargroup | Yeargroup | | Acceptable values: M = mixed year groups, N1, N2, R (= reception), 1-14. This should refer to the curriculum followed by the class and hence the values of NCyearActual for pupils in the class, rather than their dates of birth. See P45. |
| S08X | Class Key Stage | KeyStage | | Acceptable values: F(Foundation=Nursery or Reception), 1, 2, 3, M (=Mixed) |
| S171 | Number of Teachers in the Class | Teachers | 2 | Integer |
| S172 | Number of Adult Non-Teachers in the Class | NonTeachers | 2 | Integer |
| S176 | ASC Activity Category | ASCactivity | 2 | Annual Schools Census categories Table values: PE = PE or Games MD = Music, Singing or Drama TV = Watching TV or listening to Radio FL = Foreign Language GS = General Studies PS = Private Studies RE = Religious Education PL = Personal, Social or Leisure Education EN = English, Literacy or Reading MA = Mathematics or Numeracy SC = Science TE = Technology, IT or Computing HU = Humanities (including Geography, History, Economics or Business Studies) AR = Art, Craft or Design GP = Group project work (including mixed activities in other categories and library work) OT = Other Primary schools: PE, MD, TV, RE, EN, MA, SC, TE, HU, AR, GP, OT (all except FL, GS, PS and PL) Middle and Secondary schools: PE, MD, FL, GS, PS, RE, PL, EN, MA, SC, TE, HU, AR, GP, OT (ie all except TV) |

## A1. PLASC 2003 data definitions, continued

| Ref. | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| S179 | Pupils from the host school in the class | HomePupils | 3 | Integer. |
| S180 | Pupils from other schools in the class | GuestPupils | 3 | Integer. Default = 0 (zero). |
| S185 | Part-Time pupils not at school | PartTimeNotIn | 4 | Number of Pupils from this school who were part-time pupils not scheduled to be in attendance at the school's selected time.<br>Default to 0 (zero). Note definition of part-time pupil in P42 |
| S186 | Private Study pupils | PrivateStudy | 4 | Number of Pupils from this school who were engaged in private study at the school's selected time.<br>Default to 0 (zero). |
| S187 | Pupils at Another School | AtOtherSchool | 4 | Number of Pupils from this school who were attending another school at the school's selected time. Default to 0 (zero). |
| S188 | Pupils on Work Experience | WorkExperience | 4 | Number of Pupils from this school who were on work experience, etc at the school's selected time. Default to 0 (zero).<br>Must be 0 if <Phase> = MP, MS or PS. |
| S189 | Pupils at FE Colleges | FEcollege | 4 | Number of Pupils from this school who were attending a course at an FE college at the school's selected time. Default to 0 (zero).<br>Must be 0 if <Phase> = MP, MS or PS. |
| S317 | Teachers not teaching | NotTeaching | 3 | Teachers employed at the school and in school at the selected time but not teaching a class then. Default to 0 (zero). |
| S318 | Teachers at Other Schools | AtOtherSchool | 3 | Teachers employed at the school but engaged with a class at another school at the selected time. Default to 0 (zero). |
| S319 | Part-Time Teachers not at School | PTout | 3 | Part-time teachers employed at the school but not in school at the selected time. Default to 0 (zero). |
| S322 | Teachers from Other schools | FromOtherSchool | 3 | Teachers from other schools engaged with a class at this school at the selected time. Default to 0 (zero). |

## A1. PLASC 2003 data definitions, continued

| Ref. | Data item | XML tag | Maximum length | Format/table/defaults/data item validation |
|---|---|---|---|---|
| S303 | Gender of Teachers | *Gender* | 1 | *Acceptable values = M or F* |
| S304 | Tenure of Teachers | Tenure | 1 | Table Values:<br>F = Full-Time (Average 32.5 hours per week)<br>P = Part-Time (Average less than 32.5 hours per week) |
| S305 | Number of Teachers | Teachers | 2 | Integer |
| S09X | Special assistants | Category | 2 | Table values:<br>CC = Child Care Staff<br>SA = Special Assistant with one or more of: CCETSW, SRN, RSCN, NNEB, Declaration of Recognition of Experience in residential care of children. |
| S341 | Category of Non-Teaching Staff | Category | 2 | Table Values:<br>QA = teacher assistants with Diploma in Childcare and Education (previously NNEB) or equiv. (primary and special schools) UA = teacher assistants without Diploma in Childcare and Education (previously NNEB) or equiv. (primary and special schools)<br>TA = teacher assistants (middle and secondary schools) CC = child care staff<br>SN = special needs support staff EB = minority ethnic pupils support staff - bilingual assistants (all schools) EO = minority ethnic pupils support staff – other<br>ME =matrons/nurses/medical staff LI = librarians TE = technicians IT =IT technicians<br>SS = other education support staff AO = admin officers/secretaries BU = bursars<br>AC = other admin/clerical staff Primary schools: QA, UA, SN, EB, EO, ME, LI, TE, IT, SS, AO, BU, AC (ie not TA, CC) Middle and secondary schools: TA, SN, EB, EO, ME, LI, TE, IT, SS, AO, BU, AC (ie not QA, UA, CC) Special Schools: QA, UA, CC, SN, EB, EO, ME, LI, TE, IT, SS, AO, BU, AC (ie not TA) |

Note: In some instances the maximum length column is left bank in the DfES version. In this and other presentational issues, table A1 follows the DfES format.

## Table A2. National Pupil Dataset extract, 2002 GLA file descriptions

| File No | File Name | Description |
|---|---|---|
| 1 | Pupils.txt | The information stored within this file includes PLASC records from academic year 2001/2002, for all pupils who either have a home postcode in London, irrespective of where they attend school or attend a maintained school in London, irrespective of where they live. |
| 2 | KS4C.txt | This file contains Key Stage 4 records for students in the Pupil.txt file who sat Key Stage 4 exams. The records are recorded on a pupil level at one record per pupil. However, only Key Stage 4 candidate details and summary information are held in this file. Individual exam results are shown in the KS4R.txt file. |
| 3 | KS4R.txt | Key Stage 4 individual exam results are stored within this file, at one file record per exam, for the pupils in the Pupil.txt file who sat Key Stage 4 exams. |
| 4 | KS4Qual.txt | Description of Key Stage 4 qualification codes held in the KS4R.txt file. |
| 5 | KS4Subs.txt | Description of Key Stage 4 subject codes held in the KS4R.txt file. |
| 6 | KS3.txt | Key Stage 3 results for pupils in the Pupil.txt file who sat Key Stage 3 exams. |
| 7 | KS2.txt | Key Stage 2 results for pupils in the Pupil.txt file who sat Key Stage 2 exams. |
| 8 | KSl.txt | Key Stage 1 results for pupils in the Pupil.txt file who sat Key Stage 1 exams. |

Each file containing pupil data links on a field called ''PupilMatchingRef'.

**National Pupil Datasets Matching Column Names**

The NPD contains column names which are longer than 8 characters. To allow the data to be imported into 16 bit analytical software packages, which cannot accept variable names longer than 8 characters, the variables were renamed on export from the database. The following tables were drawn by the DfES to help match the less meaningful variable names with the more meaningful NPD column names. Please see the tables below for the relevant matching column names.

| PUPIL.txt variables | |
|---|---|
| Academic Year | pi ac |
| PupilMatchingRef | pl_pmr |
| AgeAtStartOfAcademicYear | age s yr |
| MonthPartOfAgeAtStartOfAcademicYear | mth age |
| Gender | pi gend |
| LEA | pi lea |
| Estab | pi estab |
| SchoolID | pl_schid |

**Table A2. National Pupil Datasets extract, 2002 GLA file descriptions, continued**

| | |
|---|---|
| DateOfJoining | pi join |
| HomePostCode | pi post |
| PartTimeInd | pl_pti |
| NurseryClassInd | pl_nci |
| BoarderInd | pl_board |
| ActualNCYearGroup | pi ncyr |
| EthnicGroup | ethnic |
| SourceOfEthnicity | eth see |
| MotherTongue | m tongue |
| FreeSchoolMeals | fsm |
| SENStage | senstage |
| PermExclusionInd | pl_pei |
| DateOfLeavingSch | do left |
| RegistrationType | pl_reg |

KS1. txt variables

| | |
|---|---|
| Academic Year | kl ac |
| PupilMatchingRef | kl_pmr |
| AgeAtStartOfAcademicYear | age_s_yr |
| MonthPartOfAgeAtStartOfAcademicYear | mth age |
| Gender | kl gend |
| LEA | kl lea |
| Estab | kl estab |
| SchoolID | kl schid |
| SourceCountry | kl land |
| LanguageOfSchool | kl_lang |
| EnglishReadingTask | kl read |
| EnglishComprehensionTest | kl comp |
| English WritingTest | kl writ |
| EnglishSpellingTest | kl_spell |
| MathsTaskTest | kl math |
| EnglishTASubj ectLevel | kleng_ta |
| EnglishTASpeakListen | kllistta |
| EnglishTAReading | klreadta |
| English! A Writing | klwritta |
| MathsTASubjectLevel | klmathta |
| MathsTAUsingApplying | klmat ua |
| MathsTANumbersAlgebra | klmat na |
| MathsTAShapesMeasures | klmat sm |
| Science! ASubj ectLevel | klsci ta |
| Science! AExperimentInvestigate | klsci ei |
| ScienceTALifeProcLivingThings | klsci It |
| ScienceTAMaterialProperties | klsci mp |
| ScienceTAPhysicalProcesses | klsci_pp |

**Table A2. National Pupil Dataset extract, 2002 GLA file descriptions, continued**

| KS2.txt variables | |
|---|---|
| Academic Year | k2 ac |
| PupilMatchingRef | k2_pmr |
| AgeAtStartOfAcademicYear | age s yr |
| MonthPartOfAgeAtStartOfAcademicYear | mth age |
| Gender | k2 gend |
| LEA | k2 lea |
| Estab | k2 estab |
| SchoolID | k2 schid |
| SourceCountry | k2 land |
| LanguageOfSchool | k2_lang |
| LanguageOfTest MA | k2 ma |
| LanguageOfTest SC | K2 SC |
| EnglishTA | k2eng ta |
| MathsTA | k2mat ta |
| ScienceTA | k2sci ta |
| WelshTA | k2wel ta |
| EnglishFinalTestLevel | k21ev e |
| MathsFinalTestLevel | k21ev m |
| ScienceFinalTestLevel | k21ev s |
| WelshFinalTestLevel | k21ev we |
| EnglishReadingTestLevel | k21ev r |
| English WritingTestLevel | k21ev wr |
| EnglishReadingTestMark | reading |
| English WritingTestMark | writing |
| EnglishHandwritingTestMark | hwriting |
| EnglishSpellingTestMark | spelling |
| EnglishTotalMark | tot 2e |
| EnglishTier | tier 2e |
| EnglishMainTestLevel | main 2e |
| EnglishExtensionTestMark | ext 2e |
| EnglishExtensionTestLevel | extlev2e |
| MathsTestAMark | papl_m |
| MathsTestBMark | pap2 m |
| MathsMentalArithmaticMark | mastat2 |
| MathsTotalMark | tot 2m |
| MathsTier | tier 2m |
| MathsMainTestLevel | main 2m |
| MathsExtensionTestMark | ext 2m |
| MathsExtensionTestLevel | extlev2m |
| ScienceTestAMark | papa_2s |
| ScienceTestBMark | papb 2s |
| ScienceTotalMark | tot 2s |
| ScienceTier | tier 2s |
| ScienceMainTestLevel | main 2s |
| ScienceExtensionTestMark | ext 2s |
| ScienceExtensionTestLevel | extlev2s |

**Table A2. National Pupil Dataset extract, 2002 GLA file descriptions, continued**

| KS3.txt variables | |
|---|---|
| Academic Year | k3 ac |
| PupilMatchingRef | k3_pmr |
| AgeAtStartOfAcademicYear | age_s yr |
| MonthPartOfAgeAtStartOfAcademicYear | mth_age |
| Gender | k3 gend |
| LEA | k3 lea |
| Estab | k3 estab |
| SchoolID | k3 schid |
| SourceCountry | k3 land |
| LanguageOfSchool | k3_lang |
| LanguageMA | k3 ma |
| LanguageSC | k3 sc |
| EnglishTA | k3eng_ta |
| MathsTA | k3mat ta |
| ScienceTA | k3sci ta |
| WelshTA | k3wel ta |
| EnglishFinalTestLevel | k31ev e |
| MathsFinalTestLevel | k31ev m |
| ScienceFinalTestLevel | k31ev s |
| WelshFinalTestLevel | k31ev we |
| EnglishPaper 1 Mark | papl e |
| EnglishPaper2Mark | pap2_e |
| EnglishTotalMark | tot 3e |
| EnglishTier | tier 2e |
| EnglishMainTestLevel | main 2e |
| EnglishExtensionTestMark | ext 2e |
| EnglishExtensionTestLevel | extlev2e |
| MathsPaper 1 Mark | papl 3m |
| MathsPaper2Mark | pap2 3m |
| MentalArithmaticMark | mastat2 |
| MathsTotalMark | tot 2m |
| MathsTier | tier 2m |
| MathsMainTestLevel | main 2m |
| MathsExtensionTestMark | ext 2m |
| MathsExtensionTestLevel | extlev2m |
| ScienceTestAMark | papa_2s |
| ScienceTestBMark | papb_2s |
| ScienceTotalMark | tot 2s |
| ScienceTier | tier 2s |
| ScienceMainTestLevel | main 2s |
| ScienceExtensionTestMark | ext 2s |
| ScienceExtensionTestLevel | extlev2s |

**Table A2. National Pupil Dataset extract, 2002 GLA file descriptions, continued**

| KS4Qual.txt variables | |
|---|---|
| QualificationCode | qualcode |
| Description | k4 desc |

| KS4Subs.txt variables | |
|---|---|
| LEAPCode | lea code |
| Subject | subject |

| KS4C.txt variables | |
|---|---|
| AcademicYear | k4c ac |
| PupilMatchingRef | k4c_pmr |
| Gender | k4c sex |
| YearGroup | yr group |
| PrivateExternalMarker | k4c_pem |
| LatestCentre | lat cen |
| LatestDFES | lat dfes |
| SourceCountry | k4 land |
| SchoolID | k4 schid |
| LEA | k4i lea |
| Estab | k4 estab |
| GCSEEntriesFull | k4 gcse f |
| GCSEEntriesShort | gcse_sho |
| GNVQEntriesFullInt | gnvq_full |
| GNVQEntriesFullFound | gnvq_f_f |
| GNVQEntriesPart 1 Found | gnvq pi |
| GNVQEntriesLangUnitFound | gnvq_lanf |
| GNVQEntriesLangUnitInt | gnvq_lani |
| NationalCurriculumCertEntries | nc cert |
| PointsScoreGCSE_GNVQ | points_s |
| PointsScoreNationalCurriculumCert | points nc |
| PointsScoreCappedBestS | points c |
| GCSEPasses A* | gcse_pa* |
| GCSEPasses A | gcse_pa |
| GCSEPasses B | gcse_pb |
| GCSEPasses C | gcse_pc |
| GCSEPasses D | gcse_pd |
| GCSEPasses E | gcse_pe |
| GCSEPasses F | gcse_pf |
| GCSEPasses G | gcse g |
| GNVQPasses A*toC_Equiv | gnvq_a*c |
| GNVQPasses DtoG Equiv | gnvq_d*g |
| HighestGradeEng | high_Eng |
| HighestGradeMaths | high_mat |

**Table A2. National Pupil Dataset extract, 2002 GLA file descriptions, continued**

| | |
|---|---|
| Total A*toC Passes | tot a*c |
| Total A*toC Passes ShortCourse | tot a*cs |
| SOrMore A*toC Passes | 5 a*c |
| TotalGCSE GNVQ A*toG Passes | tot a*g |
| TotalGCSE_GNVQ_A*toG_ Passes ShortCourse | tot_a*gs |
| 5OrMore A*toG Passes | 5 a*g |
| TotalGCSE_GNVQ_A*toG_Or NationalCurriculumCertPasses | gnvq_a*g |
| 5OrMoreGCSE_GNVQJ3r NationalCurriculumCertPasses | 5_a*g_cp |
| SOrMore A*toG PassesIncEngMaths | 5 a*g em |
| TotalGCSEEntries | tot_gcse |
| MeanGCSEScore | mea gcse |
| NumberOfGNVQEquiv_AorA*_Passes | gnvq_eaa |
| NumberOfGNVQEquiv_B Passes | gnvq_eb |
| NumberOfGNVQEquiv_C Passes | gnvq_ec |
| NumberOfGNVQEquiv D Passes | gnvq_ed |
| NumberOfGNVQEquiv E Passes | gnvq ee |
| NumberOfGNVQEquiv ForG Passes | gnvq_efg |

| KS4R.txt variables | |
|---|---|
| Academic Year | k4r ac |
| PupilMatchingRef | k4r_pmr |
| LatestDFESNumber | lat dfes |
| QualificationAssessmentCode | qual_ac |
| LEAPSubjectMapping | lea_smap |
| ExamYear | exam_yr |
| FinalGrade | final_gr |

**Technical Description**

The eight files represent one MS SQL Server table per file, with one MS SQL Server table row per file record. The format of the file is:

- File type - ANSI
- Row delimiter - carriage return and line feed
- Column delimiter - tab
- The first row in each file contains column names

## A3. Sample sections from January 2000 DfEE Form 7 for secondary schools

- Only show those pupils who are entered on an Admission Register in accordance with Regulation 6 of the Education *(Pupil Registration)* Regulations 1995.
- A working sheet is attached to help you complete this Section.

**2.1 All full-time pupils on the register on Thursday 20th January 2000**

- Do not enter zeros in the boxes where no entry is required.

| Age on 31.8.99 | Date of birth | Full-time boys | Full-time girls |
|---|---|---|---|
| 19 and over | 31.8.80 and earlier | 20 | 31 |
| 18 | 1.9.80 to 31.8.81 | 21 | 32 |
| 17 | 1.9.81 to 31.8.82 | 22 | 33 |
| 16 | 1.9.82 to 31.8.83 | 23 | 34 |
| 15 | 1.9.83 to 31.8.84 | 24 | 35 |
| 14 | 1.9.84 to 31.8.85 | 25 | 36 |
| 13 | 1.9.85 to 31.8.86 | 26 | 37 |
| 12 | 1.9.86 to 31.8.87 | 27 | 38 |
| 11 | 1.9.87 to 31.8.88 | 28 | 39 |
| Under 11 | 1.9.88 and later | 29 | 40 |
| | TOTAL | 30 | 41 |

**2.2 Total number of pupils on the register on Thursday 20th January 2000**

- This figure must agree with the sum of the TOTAL boxes in **2.1**.

**2.3 New admissions**

- Enter the number of pupils of the school's basic entry age who started school in September 1999. 42

**2.4 Boarding pupils on the register on Thursday 20th January 2000** *(included in 2.1)*

- Do not enter zeros in the boxes where no entry is required.

Number of boys 43 Number of girls 44

**2.5 School meal arrangements on Thursday 20th January 2000**

- Enter the number of **day** pupils on the register who:

| | Full-time Compulsory school age | Full-time Above compulsory school age |
|---|---|---|
| • took a **free** school meal. | 45 | 498 |
| • are known to be eligible for a **free** school meal. | 46 | 499 |

2

Note: DFEE, Department for Education and Employment. This was the precursor of the present DfES.

## A3. Sample sections from January 2000 DfEE* Form 7 for secondary schools, continued

**Section 3: Pupils on the register with special educational needs *(SEN)***

**3.1 Number of pupils for whom a statement of SEN is maintained by a LEA** 47

**3.2 Number of pupils with special educational needs *but no statement*** 48

**Section 4: Pupils' ethnicity**

- **For Sections 4.1 and 4.2, enter the numbers of pupils of compulsory school age and above as at 31st August 1999.**

**4.1 Ethnic origin**

| | White - UK heritage | White - European | White - other | White - TOTAL |
|---|---|---|---|---|
| | 49 | 50 | 51 | 52 |
| **Black - Caribbean heritage** | | | | 53 |
| **Black - African heritage** | | | | 54 |
| **Black - other** | | | | 55 |
| **Indian** | | | | 56 |
| **Pakistani** | | | | 57 |
| **Bangladeshi** | | | | 58 |
| **Chinese** | | | | 59 |
| **Any other minority ethnic group** | | | | 60 |

**IMPORTANT**

The sum of boxes 52 to 60 **should but may not be equal to** the sum of boxes 30 and 41 in **Section 2**. Please refer to Section 4 of the Completion Notes.

**4.2 Number of pupils for whom English is an additional language** 61

**4.3 Fifteen year old *(as at 31.8.99)* pupils' ethnic origin**

- Of the pupils entered in **4.1**, enter the number of pupils aged fifteen on 31st August 1999 in each of the categories shown.

| | White - UK heritage | White - European | White - other | White - TOTAL |
|---|---|---|---|---|
| | 62 | 63 | 64 | 65 |
| **Black - Caribbean heritage** | | | | 66 |
| **Black - African heritage** | | | | 67 |
| **Black - other** | | | | 68 |
| **Indian** | | | | 69 |
| **Pakistani** | | | | 70 |
| **Bangladeshi** | | | | 71 |
| **Chinese** | | | | 72 |
| **Any other minority ethnic group** | | | | 73 |

**IMPORTANT**

The sum of boxes 65 to 73 **should but may not be equal to** the sum of boxes 24 and 35 in **Section 2**. Please refer to Section 4 of the Completion Notes.

3 over

Note: DFEE, Department for Education and Employment. This was the precursor of the present DfES.

## A4. Ethnic categories available for 2003 LPD.

| Main categories. These are derived from DfES codes. | Group total | Subcategories | Subcategory code | Group total, including figures for subcategories | Negotiated extended categories | Extended category code | Totals, including figures for subcategory codes where used |
|---|---|---|---|---|---|---|---|
| **White** | **566,195** | | | | | | |
| | | **White British (mc)** | WBRI | **475,542** | | | ***371,334*** |
| | | | | | White English | WENG | 101,323 |
| | | | | | White Scottish | WSCO | 509 |
| | | | | | White Welsh | WWEL | 256 |
| | | | | | Other British | WOMB | 2,120 |
| | | **Irish** | WIRI | **13,056** | | | 13,056 |
| | | **Traveller of Irish heritage** | WIRT | **1,316** | | | 1,316 |
| | | **Any other White background (mc)** | WOTH | **75,424** | | | 13,511 |
| | | | | | Albanian other than Kosovan | WALB | 905 |
| | | | | | Bosnian-Herzegovinian | WBOS | 129 |
| | | | | | Croatian | WBOS | 110 |
| | | | | | Greek or Greek Cypriot (s) | WGRE | 2,497 |
| | | | | | Greek | WGRK | 642 |
| | | | | | Greek Cypriot | WGRC | 3,318 |
| | | | | | Italian | WITA | 875 |
| | | | | | Kosovan | WKOS | 2,391 |
| | | | | | Portuguese | WPOR | 1,968 |
| | | | | | Serbian | WSER | 99 |
| | | | | | Turkish or Turkish Cypriot (s) | WTUR | 5,372 |
| | | | | | Turkish | WTUK | 7,074 |
| | | | | | Turkish Cypriot | WTUC | 2,978 |
| | | | | | White European (s) | WEUR | 6,711 |
| | | | | | White Eastern European | WEEU | 5,201 |
| | | | | | White Western European | WWEU | 6,287 |
| | | | | | White other (s) | WOTW | 15,356 |
| | | **Gypsy/Roma** | WROM | **843** | | | 843 |

Continued

The suffix (mc) indicates a subcategory for which subdivisions exist. The suffix (s) indicates an extended category for which subdivisions exist. Figures in italics in the totals column give the number of records where a subcategory has been used which could have been subdivided into extended categories.

## A4 continued. Ethnic categories available for 2003 LPD.

| Main categories. These are derived from DfES codes. | Group total | Subcategories | Subcategory code | Group total, including figures for subcategories | Negotiated extended categories | Extended category code | Totals, including figures for subcategory codes where used |
|---|---|---|---|---|---|---|---|
| **Mixed/ Dual** | **66,904** | | | | | | |
| | | **White & Black Caribbean** | MWBC | **22,349** | | | **22,349** |
| | | **White & Black African** | MWBA | **7,278** | | | **7,278** |
| | | **White & Asian (mc)** | MWAS | **10,275** | | | ***9,455*** |
| | | | | | White & Pakistani | MWAP | 92 |
| | | | | | White & Indian | MWAI | 272 |
| | | | | | White & any other Asian | MWAO | 456 |
| | | **Any other mixed background (mc)** | MOTH | **27,001** | | | 20,864 |
| | | | | | Asian & any other ethnic group | MAOE | 542 |
| | | | | | Asian & Black | MABL | 218 |
| | | | | | Asian & Chinese | MACH | 20 |
| | | | | | Black & any other ethnic group | MBOE | 815 |
| | | | | | Black & Chinese | MBCH | 19 |
| | | | | | Chinese & any other ethnic group | MCOE | 93 |
| | | | | | White & any other ethnic group | MWOE | 2,021 |
| | | | | | White & Chinese | MWCH | 149 |
| | | | | | Other mixed background (s) | MOTM | 2,260 |
| **Asian or Asian British** | **177,064** | | | | | | |
| | | **Indian** | AIND | **68,163** | | | **68,163** |
| | | **Pakistani (mc)** | APKN | **36,664** | | | ***32,296*** |
| | | | | | Pakistani - Mirpuri Pakistani | AMPK | 285 |
| | | | | | Other Pakistani | AOPK | 3,723 |
| | | | | | Kashmiri Pakistani | AKPA | 360 |
| | | **Bangladeshi** | ABAN | **45,666** | | | **45,666** |
| | | **Any other Asian background (mc)** | AOTH | **26,571** | | | ***13,577*** |
| | | | | | African Asian | AAFR | 1,039 |
| | | | | | Kashmiri other | AKAO | 5 |
| | | | | | Nepali | ANEP | 127 |
| | | | | | Sinhalese | ASNL | 357 |
| | | | | | Sri Lankan Tamil | ASLT | 6,014 |
| | | | | | Other Asian (s) | AOTA | 5,452 |

Continued

The suffix (mc) indicates a subcategory for which subdivisions exist. The suffix (s) indicates an extended category for which subdivisions exist. Figures in italics in the totals column give the number of records where a subcategory has been used which could have been subdivided into extended categories.

## A4 continued. Ethnic categories available for 2003 LPD.

| Main categories. These are derived from DfES codes. | Group total | Subcategories | Subcategory code | Group total, including figures for subcategories | Negotiated extended categories | Extended category code | Totals, including figures for subcategory codes where used |
|---|---|---|---|---|---|---|---|
| **Black or Black British** | **196,923** | | | | | | |
| | | **Black Caribbean background** | BCRB | **74,061** | | | **74,061** |
| | | **African (mc)** | BAFR | **103,181** | | | ***32,206*** |
| | | | | | Angolan | BANN | 247 |
| | | | | | Congolese | BCON | 1,038 |
| | | | | | Ghanaian | BGHA | 7,439 |
| | | | | | Nigerian | BNGN | 18,033 |
| | | | | | Sierra Leonian | BSLN | 1,181 |
| | | | | | Somali | BSOM | 14,730 |
| | | | | | Sudanese | BSUD | 117 |
| | | | | | Other Black African | BAOF | 28,190 |
| | | **Any other Black background (mc)** | BOTH | **19,681** | | | ***16,906*** |
| | | | | | Black European | BEUR | 346 |
| | | | | | Black North American | BNAM | 37 |
| | | | | | Other Black (s) | BOTB | 2,392 |
| **Chinese** | **8,507** | | | | | | |
| | | **Chinese (mc)** | CHNE | **8,507** | | | ***7,448*** |
| | | | | | Hong Kong Chinese | CHKC | 524 |
| | | | | | Malaysian Chinese | CMAL | 42 |
| | | | | | Singaporean Chinese | CSNG | 4 |
| | | | | | Taiwanese | CTWN | 7 |
| | | | | | Other Chinese | COCH | 482 |

continued

The suffix (mc) indicates a subcategory for which subdivisions exist. The suffix (s) indicates an extended category for which subdivisions exist. Figures in italics in the totals column give the number of records where a subcategory has been used which could have been subdivided into extended categories.

## A4 continued. Ethnic categories available for 2003 LPD.

| Main categories. These are derived from DfES codes. | Group total | Subcategories | Subcategory code | Group total, including figures for subcategories | Negotiated extended categories | Extended category code | Totals, including figures for subcategory codes where used |
|---|---|---|---|---|---|---|---|
| **Any other** | **39,643** | | | | | | |
| | | **Any other ethnic group (mc)** | OAFG | **39,643** | | | ***11,901*** |
| | | | | | Afghan | OAFG | 2,945 |
| | | | | | Arab other | OARA | 2,315 |
| | | | | | Egyptian | OEGY | 303 |
| | | | | | Filipino | OFIL | 1,362 |
| | | | | | Iranian | OIRN | 1,849 |
| | | | | | Iraqi | OIRQ | 1,573 |
| | | | | | Japanese | OJPN | 429 |
| | | | | | Korean | OKOR | 672 |
| | | | | | Kurdish | OKRD | 2,977 |
| | | | | | Latin/South/Central American | OLAM | 2,787 |
| | | | | | Lebanese | OLEB | 367 |
| | | | | | Libyan | OLIB | 21 |
| | | | | | Malay | OMAL | 25 |
| | | | | | Moroccan | OMRC | 739 |
| | | | | | Polynesian | OPOL | 7 |
| | | | | | Thai | OTHA | 36 |
| | | | | | Vietnamese | OVIE | 2,733 |
| | | | | | Yemeni | OYEM | 11 |
| | | | | | Any other ethnic group (s) | OOEG | 6,591 |
| **No ethnic category** | **34,605** | Refused | REFU | **15,820** | | | **15,820** |
| | | Information not yet obtained | NOBT | **17,583** | | | **17,584** |
| | | Missing data | | **1,201** | | | **1,201** |
| **Totals** | **1,089,843** | | | **1,089,843** | | | **1,089,825** |

The suffix (mc) indicates a subcategory for which subdivisions exist. The suffix (s) indicates an extended category for which subdivisions exist. Figures in italics in the totals column give the number of records where a subcategory has been used which could have been subdivided into extended categories.

## A5. Variations in coding 2003 pupil ethnicity records. Divisible and 'other' categories used in London LEAs. Number.

| Maintaining LEA of school attended in London | Divisible or 'other ' | Not divisible or other | Missing | Total |
|---|---|---|---|---|
| Camden | 11,709 | 9,583 | 21 | 21,313 |
| Hackney | 4,044 | 22,298 | 26 | 26,368 |
| Hammersmith and Fulham | 11,198 | 6,089 | | 17,287 |
| Haringey | 14,629 | 19,387 | 140 | 34,156 |
| Islington | 4,349 | 18,812 | 2 | 23,163 |
| Kensington and Chelsea | 6,278 | 4,358 | 1 | 10,637 |
| Lambeth | 16,910 | 11,026 | 2 | 27,938 |
| Lewisham | 21,267 | 14,767 | 86 | 36,120 |
| Newham | 24,529 | 25,194 | 3 | 49,726 |
| Southwark | 16,543 | 19,574 | 68 | 36,185 |
| Tower Hamlets | 11,130 | 25,300 | 356 | 36,786 |
| Wandsworth | 18,162 | 11,134 | 2 | 29,298 |
| Westminster | 12,655 | 6,504 | | 19,159 |
| Barking and Dagenham | 27,573 | 2,939 | 7 | 30,519 |
| Barnet | 18,551 | 28,695 | 152 | 47,398 |
| Bexley | 36,182 | 4,387 | | 40,569 |
| Brent | 14,428 | 25,120 | | 39,548 |
| Bromley | 40,355 | 6,460 | 28 | 46,843 |
| Croydon | 39,169 | 12,466 | 23 | 51,658 |
| Ealing | 21,391 | 21,043 | | 42,434 |
| Enfield | 8,048 | 40,554 | 105 | 48,707 |
| Greenwich | 25,372 | 10,469 | | 35,841 |
| Harrow | 18,523 | 10,314 | | 28,837 |
| Havering | 29,643 | 7,466 | | 37,109 |
| Hillingdon | 3,989 | 38,638 | | 42,627 |
| Hounslow | 8,565 | 27,480 | 30 | 36,075 |
| Kingston upon Thames | 3,248 | 17,807 | | 21,055 |
| Merton | 19,647 | 3,671 | | 23,318 |
| Redbridge | 23,227 | 20,719 | 29 | 43,975 |
| Richmond upon Thames | 16,467 | 3,483 | 32 | 19,982 |
| Sutton | 27,385 | 3,691 | 88 | 31,164 |
| Waltham Forest | 23,625 | 12,064 | | 35,689 |
| Inner London | 173,484 | 194,167 | 707 | 368,358 |
| Outer London | 405,388 | 297,466 | 494 | 703,348 |
| Greater London | 578,872 | 491,633 | 1,201 | 1,071,706 |

Figures from the one school maintained by the Corporation of London are suppressed to avoid disclosure, but have been included in calculations in this table and in table A6.
Pupils living in London and attending maintained schools outside London are not included in this table. Totals therefore differ from those in table A4, which includes all pupils in the LPD.
Source: version 1 2003 London Pupil Dataset

## A6. Variations in coding 2003 pupil ethnicity records. Divisible and 'other categories used in London LEAs. Percentage.

| Maintaining LEA of school Attended in London | Divisible' or 'other' records as % of all records | Highest % of school records with divisible or 'other ethnic labels | Lowest % of school records with divisible or 'other' ethnic labels | Difference – highest and lowest % use of divisible or 'other' categories by schools | Standard deviation - percentage of records with divisible or 'other' labels |
|---|---|---|---|---|---|
| Camden | 54.9 | 80.9 | 18.0 | 62.9 | 13.9 |
| Hackney | 15.3 | 99.3 | 7.2 | 92.1 | 12.1 |
| Hammersmith and Fulham | 64.8 | 87.8 | 40.6 | 47.1 | 9.6 |
| Haringey | 42.8 | 91.1 | 11.3 | 79.8 | 19.1 |
| Islington | 18.8 | 63.5 | 6.3 | 57.2 | 11.7 |
| Kensington and Chelsea | 59.0 | 89.6 | 19.3 | 70.3 | 15.9 |
| Lambeth | 60.5 | 81.6 | 41.2 | 40.4 | 9.7 |
| Lewisham | 58.9 | 82.6 | 0.1 | 82.5 | 12.5 |
| Newham | 49.3 | 80.8 | 28.3 | 52.5 | 13.6 |
| Southwark | 45.7 | 85.3 | 0.0 | 85.3 | 18.6 |
| Tower Hamlets | 30.3 | 84.8 | 0.0 | 84.8 | 26.1 |
| Wandsworth | 62.0 | 90.4 | 29.2 | 61.2 | 14.0 |
| Westminster | 66.1 | 82.5 | 47.1 | 35.3 | 8.8 |
| Barking and Dagenham | 90.3 | 98.0 | 74.4 | 23.6 | 5.0 |
| Barnet | 39.1 | 100.0 | 6.8 | 93.2 | 23.6 |
| Bexley | 89.2 | 100.0 | 38.2 | 61.8 | 8.8 |
| Brent | 36.5 | 99.5 | 0.0 | 99.5 | 18.0 |
| Bromley | 86.1 | 100.0 | 4.7 | 95.3 | 18.4 |
| Croydon | 75.8 | 97.6 | 45.2 | 52.4 | 13.0 |
| Ealing | 50.4 | 77.9 | 9.1 | 68.8 | 16.5 |
| Enfield | 16.5 | 43.4 | 4.8 | 38.6 | 7.9 |
| Greenwich | 70.8 | 94.9 | 35.7 | 59.1 | 13.0 |
| Harrow | 64.2 | 100.0 | 33.8 | 66.2 | 12.4 |
| Havering | 79.9 | 98.6 | 0.0 | 98.6 | 17.0 |
| Hillingdon | 9.4 | 43.4 | 0.0 | 43.4 | 5.8 |
| Hounslow | 23.7 | 53.6 | 4.9 | 48.7 | 9.3 |
| Kingston upon Thames | 15.4 | 33.3 | 4.1 | 29.2 | 6.9 |
| Merton | 84.3 | 95.3 | 57.3 | 38.0 | 7.3 |
| Redbridge | 52.8 | 100.0 | 10.8 | 89.2 | 22.6 |
| Richmond upon Thames | 82.4 | 96.9 | 9.6 | 87.3 | 14.0 |
| Sutton | 87.9 | 97.8 | 8.0 | 89.9 | 13.9 |
| Waltham Forest | 66.2 | 89.7 | 27.3 | 62.5 | 13.4 |
| Inner London | 47.1 | 99.3 | 0.0 | 99.3 | 22.0 |
| Outer London | 57.6 | 100.0 | 0.0 | 100.0 | 30.5 |
| Greater London | 54.0 | 100.0 | 0.0 | 100.0 | 28.1 |

Figures from the one school maintained by the Corporation of London are suppressed to avoid disclosure but have been included in the calculations in this table and in table A5.
Pupils living in London and attending maintained schools outside London are not included in this table. Percentages therefore differ marginal from those which would be arrived at using figures from table A4, which include all pupils in the LPD.
Source: version 1 2003 London Pupil Dataset

## Regular Briefings from the GLA Data Management and Analysis Group

Latest DMAG Briefings:

| | | |
|---|---|---|
| DMAG 2005/1 | County of Birth and Labour Market Outcomes | Lorna Spence |
| DMAG 2005/2 | 2001 Census: London Country of Birth Profiles | Giorgio Finella |
| DMAG 2005/3 | 2001 Census: Economic Activity in London | Giorgio Finella |
| DMAG 2005/4 | 2001 Census Profiles: Pakistanis in London | Gareth Piggott |
| DMAG 2005/5 | Indices of Deprivation 2004: Ward analysis | Lovedeep Vaid |
| DMAG 2005/6 | London – The World in a City | Marian Mackintosh |
| DMAG 2005/7 | Claimant Count Model: Technical Note | Lorna Spence/ Georgia Hay |

Recent *DMAG Briefings*:

| | | |
|---|---|---|
| DMAG 2004/1 | 2003 Round Ward Population Projections | Georgia Hay |
| DMAG 2004/2 | London's Workers Part 1: 2001 Census results | Eileen Howes |
| DMAG 2004/3 | Borough Demographic Profiles | Georgia Hay |
| DMAG 2004/4 | DMAG Annual Review 2003 | Jackie Maguire/ Rob Lewis |
| DMAG 2004/5 | 2003 Round Demographic Projections | John Hollis |
| DMAG 2004/6 | Londoners' Qualifications: Analysis of 2001 Census data | Gareth Piggott |
| DMAG 2004/7 | The Parliamentary General Election, 2001 | Michael Minors/ Dennis Grenham/ Jackie Maguire |
| DMAG 2004/8 | Constituency Demographic Profiles | Georgia Hay |
| DMAG 2004/9 | Measuring Unemployment | Lorna Spence |
| DMAG 2004/10 | Measures of Income | Rachel Leeser/ Lorna Spence |
| DMAG 2004/11 | 2001 Census Profiles: The Irish in London | Eileen Howes |
| DMAG 2004/12 | Health in London Age-standardised Rates | Gareth Piggott |
| DMAG 2004/13 | Data Sources on Refugees and Asylum Seekers | Ed Klodawski |
| DMAG 2004/14 | Census Information Note 2004-1 | Eileen Howes |
| DMAG 2004/15 | 2001 Census: Local Authority Population Studies | John Hollis |
| DMAG 2004/16 | 2001 Census Profiles: Bangladeshis in London | Gareth Piggott |
| DMAG 2004/17 | London Plan Sub-Regional Demographic Profiles | Georgia Hay |
| DMAG 2004/18 | Indices of Deprivation 2004 | Rachel Leeser |
| DMAG 2004/19 | 2001 Census Profiles: Muslims in London | Eileen Howes |
| DMAG 2004/20 | ONS Mid-Year Population Estimates: 2001 and 2002 (revised) and 2003 | John Hollis/ Georgia Hay |
| DMAG 2004/21 | School Register | Shen Cheng |
| DMAG 2004/22 | Census Information Note 2004-2 | Gareth Piggott |
| DMAG 2004/23 | ONS Mid-year Population Estimates: 1991 to 2003 | John Hollis/ Georgia Hay |
| DMAG 2004/24 | Fertility of Ethnic Groups in London, 2002/03 | Ed Klodawski |
| DMAG 2004/25 | ONS 2003-based Subnational Projections | John Hollis/ Georgia Hay |
| DMAG 2004/26 | Census Information Note 2004-3 | Eileen Howes |
| DMAG 2004/27 | PayCheck 2004 | Lovedeep Vaid |

If you would like copies of previous briefings, please contact Jackie Maguire at the GLA, jackie.maguire@london.gov.uk

## Contact details for the Data Management and Analysis Group are as follows:

**Rob Lewis (020 7983 4652)** is **Head of the Data Management and Analysis Group**. rob.lewis@london.gov.uk

**Bill Armstrong (020 7983 4653)** works in the **Census Team** with particular responsibilities for **commissioned tables**, **workplace data** and **mapping**. bill.armstrong@london.gov.uk

**Baljit Bains (020 7983 4613)** works in the **Demography Team** and is responsible for **ethnic demography**, including **ethnic group projections**. baljit.bains@london.gov.uk

**Gareth Baker (020 7983 4965**) works in the **GIS Team**. gareth.baker@london.gov.uk

**Shen Cheng (020 7983 4889)** works on **Education data** and is responsible for **school roll projections**. shen.cheng@london.gov.uk

**Hywel Davies (020 7983 4696)** is responsible for the **SASPAC** project and for the development of **GIS** work. hywel.davies@london.gov.uk

**David Ewens (020 7983 4656)** is responsible for **Education research and data analysis**. david.ewens@london.gov.uk

**Giorgio Finella (020 7983 4328**) works in the **Census Team.** giorgio.finella@london.gov.uk

**Dennis Grenham (020 7983 4532)** works mostly on **statistical compendia**, **election statistics** and **special publications**. dennis.grenham@london.gov.uk

**Georgia Hay (020 7983 4347**) works in the **Demography Team** and is responsible for **ward level projections,** the **Demography Extranet** and **borough liaison**. georgia.hay@london.gov.uk

**John Hollis (020 7983 4604)** is responsible for the work of the **Demography Team** and the **Social Exclusion Team**, and particularly for **demographic modelling.** john.hollis@london.gov.uk

**Eileen Howes (020 7983 4657)** is responsible for the work of the **Census Team**. eileen.howes@london.gov.uk

**Ed Klodawski (020 7983 4694)** works in the **Demography Team.** His post is joint with the **London Health Observatory** and specialises in **ethnic** and **health** issues**.** edmund.klodawski@london.gov.uk

**Rachel Leeser (020 7983 4699)** works in the **Social Exclusion Team** with particular responsibilities for **indicators** and **income data**. rachel.leeser@london.gov.uk (maternity leave until late 2005)

**Alan Lewis (020 7983 4348)** is a member of the **SASPAC Team**. alan.lewis@london.gov.uk

**Jackie Maguire (020 7983 4655)** is responsible to the Group Head and co-ordinates the **administrative and financial** work of the Group. jackie.maguire@london.gov.uk

**Michael Minors (020 7983 4654)** is responsible for the work of the **General Statistics and Education Team**. michael.minors@london.gov.uk

**Gareth Piggott (020 7983 4327)** works in the **Census Team**. gareth.piggott@london.gov.uk

**Lorna Spence (020 7983 4658)** is a member of the **Social Exclusion Team**, with particular responsibilities for the **Labour Force Survey** and **benefits data**. lorna.spence@london.gov.uk

**Lovedeep Vaid (020 7983 4699)** works in the **Social Exclusion Team** with particular responsibilities for **indicators** and **income data**. lovedeep.vaid@london.gov.uk (maternity cover until late 2005)

Please use the above descriptions in deciding whom to contact to assist you with your information needs.