The new ethnicity classification in the Labour Force Survey

By Allan Smith, Labour Market Division, Office for National Statistics

Key points

- The recommended classification of ethnic groups from National Statistics data sources changed in 2001 to be broadly in line with the 2001 Census.
- The new ethnicity classification has 15 categories including four which describe people of mixed race.
- The new classification is fundamentally different to the previous one so headline labour market series for spring 1997 to winter 2000/01 have been estimated on the new basis.
- The highest working-age employment rates were for the White British population and the lowest were for the Bangladeshi population.
- Analysis of ethnicity data from the LFS demonstrates great diversity of different groups.

From spring 2001 the Labour Force Survey adopted new ethnicity questions and a new interim output classification for the presentation of ethnicity data.

Introduction

THIS ARTICLE describes the new ethnicity questions and output classifications used by the Labour Force Survey (LFS) from spring 2001. It describes the background to these changes, explains the implications for continuity, and presents some illustrative data using the new classification.

New LFS ethnicity output classification

A new interim output classification of ethnic groups for National Statistics data sources was introduced in 2001. The new output categories support varying degrees of comparability with the 2001 population censuses of the different countries of the UK (which differ in the categories used), allowing commensurate comparability at the Great Britain and UK levels. This change is described in more detail on the National Statistics website www.statistics.gov.uk/about/classifications/ns_ethnic_classification.asp.

The production of ethnicity data from the LFS, as with other sources, can be seen as a process requiring inputs (in this case survey questions) which feed into outputs; here the ethnicity output classification. The new ethnicity output classification as applied to the LFS
The new ethnicity classification in the Labour Force Survey operates on two levels: level 1 is a broad classification into six main groups; level 2 nests within level 1, and provides a finer 15-point classification (see Table 1). Users of LFS individual record databases should note that the variables cov- ering these two levels of classification are derived from raw responses collected in the LFS questionnaire, and are described in more detail in the technical note. Table 1 shows the two levels of the new classification, the relationship between them, and compares these levels with the old classification structure previously used by the LFS.

Table 1: Old and new Labour Force Survey ethnicity classifications

<table>
<thead>
<tr>
<th>Old output classification</th>
<th>New output classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Level 1</td>
</tr>
<tr>
<td>White</td>
<td>British¹</td>
</tr>
<tr>
<td>Black</td>
<td>Black Caribbean</td>
</tr>
<tr>
<td>Pakistani/Bangladesi</td>
<td>Asian or Asian British</td>
</tr>
<tr>
<td>Mixed/Other origins</td>
<td>Black or Black British</td>
</tr>
<tr>
<td>Indian</td>
<td>Indian</td>
</tr>
<tr>
<td></td>
<td>Pakistani</td>
</tr>
<tr>
<td></td>
<td>Bangladeshi</td>
</tr>
<tr>
<td></td>
<td>Black</td>
</tr>
<tr>
<td></td>
<td>Mixed</td>
</tr>
<tr>
<td></td>
<td>Chinese</td>
</tr>
<tr>
<td></td>
<td>Other ethnic group</td>
</tr>
</tbody>
</table>

¹ The questions which allow these categories to be derived are not asked in Northern Ireland. Analysis of the level 2 classification variable will represent Great Britain only for these two categories.

Quality issues for spring 2001 data

LFS respondents are interviewed in five successive quarters. In normal circumstances, where information about the respondent does not change between quarters, for example date of birth, or in situations where respondents could not be contacted in a later quarter, information from the previous quarter is rolled forward. This is referred to as imputation. With the introduction of new ethnicity questions to the LFS, there were no data to roll forward for respondents who could not be contacted.

An analysis of non-respondents showed that they represented 6 per cent of the total population aged 16 and over. An examination of their known characteristics (using data from the winter 2000/01 quarter) showed that, in comparison with respondents that quarter, they contained a smaller proportion of White people, a higher proportion of men and a higher proportion of people aged under 25. They were also more likely to be employed and less likely to be economically inactive than respondents.

Without treating missing values, analysis by ethnic group for this quarter would be misleading. For this reason, additional imputation procedures were adopted to ensure the greatest possible number of cases had the new ethnicity information present for the spring 2001 quarter. This imputation process is described in the technical note.

Continuity

Although the key messages regarding differences between and within ethnic groups remains the same for broadly comparable groups under the old and new classification, it is no longer possible to produce directly comparable analysis over time directly from the LFS individual record data. Nor is it possible to compare tables of aggregates on the new basis with those produced on the old basis. Discontinuity exists even for analysis comparing the White and non-White groups. However, it was clear that users required consistent time series information on the labour market behaviour of people from ethnic minority groups, not least for the monitoring and assessment of government policy.

For this reason, historical or ‘back-cast’ estimates were produced for headline labour market series of levels and rates at both level 1 and level 2 of the new classification. The process adopted is described in the technical note. Some backcast data are used in the second section of this article, while the full backcast data tables are available on the National Statistics website (www.statistics.gov.uk/statbase/product.asp?vlnk=9670).
The new ethnicity classification in the Labour Force Survey

For two main reasons, quarterly backcast estimates have only been produced for periods from spring 1997 onwards. Firstly, an amendment was made to the answer categories of the ethnic origin questions in winter 1996, which means that the backcasting methodology could not be easily applied to periods before this. Secondly, and more importantly, ethnicity as a concept and the terms used to describe it change over time. This is reflected in the need to update ethnicity classifications to keep them current. Using more recent concepts to describe the past runs the risk of providing an inaccurate historical picture which increases the further back in time they are applied.

Results

Since the release of the 2001 Census estimate for the UK population, LFS estimates have needed to be reweighted to the new population figures. Estimates of employment and unemployment levels from the LFS released before 30 October 2002 are too high and rates are also affected. ONS has published interim reweighted LFS estimates for the UK all available on the National Statistics website.

The reweighted figures only cover top level series published in the labour market statistics First Release. The figures included in this article are produced from unrevised microdata. This means that some of the figures may be inaccurate as they are based on old population figures. The impacts of this are significantly reduced when looking at sex-specific rates and at the working-age population. A full reweighting of the microdata should be complete by summer 2003 but until then these figures should be treated with caution.

Given the issues mentioned above, no levels are given in the following sections. Estimates for small groups have relatively high sampling variability so

Table 2

<table>
<thead>
<tr>
<th>Ethnic Group</th>
<th>0-15</th>
<th>16-34</th>
<th>35-59/64</th>
<th>60/65+</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>19</td>
<td>25</td>
<td>38</td>
<td>18</td>
</tr>
<tr>
<td>British</td>
<td>19</td>
<td>24</td>
<td>38</td>
<td>19</td>
</tr>
<tr>
<td>Other White</td>
<td>12</td>
<td>34</td>
<td>39</td>
<td>15</td>
</tr>
<tr>
<td>Mixed</td>
<td>56</td>
<td>27</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>White and Black Caribbean</td>
<td>61</td>
<td>26</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>White and Black African</td>
<td>52</td>
<td>30</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>White and Asian</td>
<td>53</td>
<td>28</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Other Mixed</td>
<td>48</td>
<td>24</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>Asian or Asian British</td>
<td>28</td>
<td>36</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>Indian</td>
<td>21</td>
<td>34</td>
<td>36</td>
<td>8</td>
</tr>
<tr>
<td>Pakistani</td>
<td>36</td>
<td>36</td>
<td>24</td>
<td>5</td>
</tr>
<tr>
<td>Bangladeshi</td>
<td>39</td>
<td>39</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Other Asian</td>
<td>23</td>
<td>38</td>
<td>34</td>
<td>4</td>
</tr>
<tr>
<td>Black or Black British</td>
<td>28</td>
<td>30</td>
<td>35</td>
<td>7</td>
</tr>
<tr>
<td>Black Caribbean</td>
<td>24</td>
<td>26</td>
<td>39</td>
<td>11</td>
</tr>
<tr>
<td>Black African</td>
<td>32</td>
<td>34</td>
<td>31</td>
<td>3</td>
</tr>
<tr>
<td>Other Black</td>
<td>36</td>
<td>34</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>Chinese</td>
<td>18</td>
<td>40</td>
<td>36</td>
<td>5</td>
</tr>
<tr>
<td>Other ethnic group</td>
<td>24</td>
<td>35</td>
<td>36</td>
<td>5</td>
</tr>
</tbody>
</table>

Source: Labour Force Survey

a Excludes people whose ethnic group is not known.
b These data are presented for Great Britain only and exclude Northern Ireland. Detailed level ethnicity questions are not asked of the White group in Northern Ireland.

* Sample size too small for reliable estimates.
the estimates shown here should be regarded as illustrative only. See technical note.

**Ethnic minority population**

LFS data for summer 2002 show that of the total population living in households, 8 per cent identified themselves as members of an ethnic minority group. Figure 1 demonstrates that the broad age-distribution of those from ethnic minority groups differs significantly for ethnic minority groups overall, relative to the White group. The proportion of the ethnic minority group who are of working age, at 64 per cent compares with 63 per cent for the White group. However, the most significant difference is among those aged under 16, and those over working age. The ethnic minority population overall is younger, with just 6 per cent of the population being of state retirement age or older, compared with 18 per cent for the White population.

Table 2 demonstrates this age distribution in greater detail to show how this differs for ethnic minority groups at level 1 and level 2. The White population is older than each of the ethnic minority groups. Perhaps the most marked difference in ages relative to the White group is the structure of the population identifying themselves as Mixed. Overall, of those identifying themselves as of Mixed ethnicity well over half were aged under 16 (56 per cent) with only 2 per cent aged 60/65 and over. There are also considerable differences in the age structure beneath the broad six category classification. For example, within the Asian or Asian British group, those identifying themselves as Bangladeshi have a much younger age structure than any of the other Asian groups, with almost two-thirds aged under 16 and only 4 per cent aged over 60/65, compared with 28 per cent and 6 per cent respectively for the Asian or Asian British group as a whole.

**Labour market experiences**

When considering the labour market experiences of ethnic minority groups,
it is important to appreciate their diversity. Not only are there considerable differences between the groups at level 1, but also within groups at the detailed level 2 classification and between sexes. Both supply and demand side factors are likely to explain these differences. Among the supply side factors to bear in mind are the age structures of the ethnic minority populations, the relationship between age structures and participation in education, and the likely influence of cultural factors in explaining the differences. On the demand side are factors such as education and skills, and discrimination. These factors are not pursued in this article, but a more detailed examination of them can be found in Ethnic Minorities in Britain, Diversity and Disadvantage.1

Economic activity

Economic activity rates vary considerably between ethnic groups (see Table 3). For women, the highest working-age activity rates in summer 2002 at level 1 were for White women (75 per cent), followed by women from the Mixed group (66 per cent), and then by Black or Black British women (64 per cent). Asian or Asian British women had the lowest overall activity rate at 52 per cent. However, this hides the diversity of experience for women from different Asian backgrounds, with Indian women having the highest activity rate of 69 per cent and Bangladeshi women having the lowest of 22 per cent.

The situation for men is different, although still with great diversity for men from different ethnic groups. Again, at level 1, the activity rate is highest for men from the White group (85 per cent). However, the lowest activity rates are for men from the Other ethnic group (69 per cent) followed by the Black or Black British group (77 per cent). At the more detailed level 2, male economic activity rates are much less widely distributed than those for women. For example, within the Asian or Asian British group, a large difference between activity rates for Indian and Bangladeshi men exists, as it does for women, but this is much narrower in percentage point terms, 80 per cent and 75 per cent respectively.

In interpreting the differences between the activity rates of different ethnic groups relative to the White groups, the importance of age structure should be borne in mind. The younger age profile of ethnic minority groups is one reason why they tend to have lower activity rates overall than the White population. Young people are much more likely to be in full-time education and therefore less likely to be economically active than those in older age groups. In addition, ethnic minority groups tend to have higher participation in full-time education than those in the White group. Table 4 shows the distribution of activity rates by age and sex for different ethnic groups. This demonstrates the point that, in general, the economic activity rates of the White group are higher than those of ethnic minority groups for both sexes and in each of the age groups. It also highlights the point previously made regarding activity rates being relatively low for those in the 16-24 age group. Also worth noting is that, while the gap between the activity rates of the White population and ethnic minority groups generally tends to narrow at older working ages, for Asian or Asian British women it remains at between 20 and 30 percentage points in each of the age groups.

Employment

Employment rates follow a very similar pattern to economic activity rates. The highest working-age employment rates were for the White British population and the lowest were for the Bangladeshi population (see Figure 2), and in the case of men, for the Other Mixed ethnicity group. As with activity rates, there were some very different employment rates between ethnic minority groups, in par-
ticular for women, and also notable differences between men and women of the same ethnic minority group. The biggest differences in rates between the sexes in the same ethnic group were for Bangladeshi men and women at 44 percentage points, followed by Pakistani men and women (34 percentage points). The picture is very different for women in the Other Black group, where the employment rate for women was 8 percentage points higher than for men in summer 2002.

Unemployment

Table 3 also shows the unemployment rates for the different ethnic minority groups using the new classification. Here, the commentary is largely limited to the level 1 classification due to the small number of observations of unemployment in the summer 2002 sample for ethnic groups at level 2. For men, the lowest unemployment level was found for the White population at 5 per cent in summer 2002. The rates for men from ethnic minority groups were almost all double this rate or more, with the highest rate being for Bangladeshi men at 21 per cent. For women also, the lowest unemployment rate was in the White population; the highest rate was among the Mixed ethnicity population.

Time series

Figure 3 to Figure 5 present the most recent data for activity, employment and unemployment using the backcast time series to show how the rates have developed over time. In interpreting changes over time, users should bear in mind that the results for ethnic minority groups tend to be more volatile than for the White group. Being based on a smaller number of observations, they tend to have higher sampling variability than for the White group. Users should also note that the changes here are shown only for summer quarters, as the data are not seasonally adjusted.

Figure 3 shows the changes to activity rates since summer 1997. For men activity rates for the White population declined by 1 percentage point to 85 per cent in summer 2002. Over the same period the largest falls in activity rates were for men in the Black or Black British group (3 percentage points) and the Other ethnic groups category (9 percentage points). For women, activity rates in the White group increased over the same period by 2 percentage points to 75 per cent. As for men, the Black or Black British group and the Other ethnic groups category have both seen decreases in their activity rates over the period, while in the remaining three categories the activity rates have increased and narrowed the gap with White women.

Figure 4 shows the changes in employment rates over the same period. For White men, the working-age employment rate has increased over the five-year period by 1 percentage point to 81 per cent. Chinese men, whose employment rate increased by 9 percentage points to 75 per cent, experienced the biggest increase in percentage point terms. For women, the largest increases in employment rate were in the Chinese group (6 percentage points to 58 per cent) and for Asian and Asian British women (by 5 percentage points to 47 per cent).

In the case of unemployment rates, the sample sizes are often too small to present the full time series for each of the six ethnic categories, so Figure 5 is presented comparing only the White...
Figure 3: Economic activity rates by ethnic group and sex; United Kingdom; summer 1997 to summer 2002, not seasonally adjusted

Per cent

Men

Women

Source: Labour Force Survey

Figure 4: Employment rates by ethnic group and sex; United Kingdom; summer 1997 to summer 2002

Per cent

Men

Women

Source: Labour Force Survey

Note: Figures are not seasonally adjusted.

Data for the period 1997 to 2000 are backcast.
group with all ethnic minorities combined. Users should note, as already identified, this type of presentation clearly disguises a range of very diverse experiences for different ethnic groups. Overall for both men and women, unemployment rates have been decreasing over the five-year period. At the same time, the gap has narrowed in percentage-point terms between the rates for the ethnic minority population and those for the White population. In the case of men, the unemployment rate has fallen by 2 percentage points to 5 per cent, while for all ethnic minority groups it has fallen by 3 percentage points to 12 per cent. For women, the unemployment rate has fallen by almost 2 percentage points to 4 per cent, while for all ethnic minority groups it has fallen by 3 percentage points to 12 per cent.

**Conclusion**

Analysis of ethnicity data from the LFS demonstrates great diversity in the experiences of different groups, not only between the level 1 classifications but also within these broad classifications at the more detailed level 2 classification. These differences will be caused by a combination of both labour demand and supply side factors.

The basis of the new ethnicity classification is fundamentally different from the classification previously used, which means that data presented on the new basis should not be compared directly with data produced on the old basis. Users should bear in mind, however, that the broad messages regarding the experiences of approximately equivalent groups are not changed greatly.

**Note**

Ethnicity questions in the LFS

The following are the questions used by the LFS since spring 2001. In the presentation below the question is followed by the geographical coverage. The way in which these questions are converted from responses into the two main output classification variables cannot easily be presented as part of this article. However, the derivations can be provided as flow charts on request and will be available in the next edition of volume 4 of the LFS User Guide.

All people are asked at first interview: To which of these ethnic groups do you consider you belong? (UK)
1 White
2 Mixed
3 Asian or Asian British
4 Black or Black British
5 Chinese
6 Other ethnic group

If White: And to which of these ethnic groups do you consider you belong? (GB)
1 British
2 Another White background?

If Mixed: And to which of these ethnic groups do you consider you belong? (UK)
1 White + Black Caribbean
2 White + Black African
3 White + Asian, or
4 Another Mixed background?

If Asian or Asian British: And to which of these ethnic groups do you consider you belong? (UK)
1 Indian
2 Pakistani
3 Bangladeshi, or
4 Another Asian background?

If Black or Black British: And to which of these ethnic groups do you consider you belong? (UK)
1 Caribbean
2 African, or
3 Another Black background?

If Other: Please can you describe your ethnic group? (UK)
INTERVIEWER ENTERS DESCRIPTION OF ETHNIC ORIGIN
Another White background
Another Mixed background
Another Asian background
Another Black background

Imputation methodology
The work to correct for the quality issues in the spring 2001 files can be separated into four stages: augmentation, recoding, model development and imputation.

Augmentation
Spring 2001 ethnicity data were augmented with data collected in summer 2001. That is, in cases where ethnicity was missing in spring, but for which a response was recorded in summer, data was fed back to repopulate the spring 2001 dataset.

Recoding
‘Other’ type responses recorded verbatim were recoded according to a provisional census coding schema. This code was used in conjunction with the response at the first question to derive a new six-point classification for each case. Some adjustment was needed to the outcomes to reflect the differences in questionnaire design between the Census and the LFS.

Modelling
Using adult cases where both new (spring 2001) and old (winter 2000/01) ethnicity was present, a predictive model for new ethnicity was devised. Taking old ethnicity as the best predictor of new ethnicity, an exhaustive ‘chaid’ analysis (using AnswerTree® software) further identified tenure, age and number of children in the family unit as variables to be included in the model for some of the old ethnic groups. These breakdowns determined the imputation classes to be used in the imputation process.

Imputation
The remaining cases of missing new adult ethnicity were imputed using the computer package Stata®. A method of hot-decking imputation which randomly selects a donor case from within an imputation class was employed to populate the missing ethnic group values. This process was repeated five times to produce five replicate datasets to investigate the amount the final distribution of ethnic groups varies according to the imputation process.

Table 5 shows ethnicity of respondents (numbering 75,118) and imputed cases (numbering 3,129) for each imputation. It demonstrates that the amount of variation due to the imputation process was very small.

Backcasting methodology
LFS respondents are interviewed in five successive quarters. Certain information, such as ethnicity, which does not change from quarter to quarter, is generally collected only once at first interview. With the introduction of the new ethnicity classification in spring 2001, the information was collected again on the new basis. Once missing values for spring 2001 had been imputed (described above) a matrix was constructed using data from respondents whose ethnicity was available on both bases, which described the relationship between the old and new classifications. Tables of aggregates for the series which were to be backcast were then produced for periods from spring 1997 to winter 2000/01 using the old ethnicity classification, and the relationship matrix was used to estimate historical series on the new basis.
Analysis of the backcast results demonstrates that the method produces stable distributions of ethnicity over time, while sampling errors produced to quality assure the backcast estimates were no larger than those calculated for direct estimates.

### Table 5: Ethnicity of respondents and imputed cases for each imputation

<table>
<thead>
<tr>
<th>Ethnic group</th>
<th>First replicate</th>
<th>Second replicate</th>
<th>Third replicate</th>
<th>Fourth replicate</th>
<th>Fifth replicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>94.82</td>
<td>94.81</td>
<td>94.81</td>
<td>94.82</td>
<td>94.82</td>
</tr>
<tr>
<td>Mixed</td>
<td>0.4</td>
<td>0.41</td>
<td>0.41</td>
<td>0.42</td>
<td>0.4</td>
</tr>
<tr>
<td>Asian</td>
<td>2.71</td>
<td>2.71</td>
<td>2.71</td>
<td>2.7</td>
<td>2.71</td>
</tr>
<tr>
<td>Black</td>
<td>1.5</td>
<td>1.5</td>
<td>1.49</td>
<td>1.49</td>
<td>1.5</td>
</tr>
<tr>
<td>Chinese</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Other</td>
<td>0.32</td>
<td>0.32</td>
<td>0.33</td>
<td>0.33</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Source: Office for National Statistics

### Sampling variability

The use of annual averages provides estimates that are more reliable than those based on quarterly data, particularly for smaller groups. This method was not used in this article as the primary focus is the changing ethnicity classification. For further analysis of ethnic groups the use of annual averages is recommended. Further information on this can be found in previous articles (see pp29-42, Labour Market Trends, January 2001 and pp17-22, Labour Market Trends, January 1998). An alternative is to use the annual LFS databases.

The analysis is based on quarterly LFS data, especially for summer (June-August 2002). This reflects the facts that key labour market indicators for ethnic groups are published quarterly, and that the backcast data have been produced for quarters back to spring 1997. However, in general it is recommended that the (annual) local area LFS database is used. The sample size is larger and oversampling occurs in urban areas where the highest concentrations of minority ethnic groups occur. Thus estimates relating to ethnic groups are likely to have lower sampling errors and therefore be more precise.