



Department
for Business
Innovation & Skills

BIS RESEARCH PAPER NUMBER 81D

2011 Skills for Life Survey: Small Area
Estimation User Guide

DECEMBER 2012

RESEARCH

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Business, Innovation and Skills.

Department for Business, Innovation and Skills

1 Victoria Street

London SW1H 0ET

www.bis.gov.uk

Research paper number 81D

November 2012

Contents

Contents	2
Report Authors and Acknowledgements	4
Data Sets	5
Survey Background	8
Survey introduction	8
Survey methodology	8
The interview.....	8
Fieldwork.....	9
Data weighting and imputation.....	9
Modelling	10
Top line geographical area estimates	10
Local Area Estimates	10
Modelling Local Area Estimates.....	11
Variable Selection and Model Fitting	11
Determining the composition of local populations.....	14
Generating local area predictions	15
Summary of results	16
Relationship between ‘direct’ survey estimates and ‘modelled’ small area estimates	20
Caveats	21
References	23
Annex: Definition of Skills Levels	24

Report Authors and Acknowledgements

The *2011 Skills for Life Survey* was conducted by a partnership of two complementary agencies: TNS-BMRB, a research agency and AlphaPlus Consultancy Ltd. TNS-BMRB was responsible for all data collection and primary data processing, whilst AlphaPlus provided advice on Skills for Life policy and related issues throughout the survey. Both agencies were responsible for the main survey analysis and reporting.

The Small Area Estimation project was undertaken by Alex Gibson, Principal Researcher and Director of RAE Consulting and Innovation and Research Fellow at the University of Plymouth, and Paul Hewson, Associate Professor in Statistics at the University of Plymouth.

Data Sets

Two sets of small area estimates are provided for seven geographies; Middle Layer Super Output Areas (MSOAs), Standard Table Wards, 2005 Statistical Wards, 2011 Council Wards, 2011 Parliamentary Constituencies, Local Authorities, and Local Enterprise Partnership areas.

The first set of estimates concern the number and proportion of adults falling into each of the reported adult literacy, numeracy and ICT skill levels, as well as of the number and proportion of adults for whom English is, or is not, a first language (see Table 1 below). The second set conflates the adult skill estimates into just two categories; namely “Entry Level and below” and “Level 1 and above”. All estimates are accompanied by a measure of uncertainty. These 95% Credible Intervals (CIs) – literally describing the range within which we are 95% certain the true number or rate will lie – have been drawn directly from each estimate’s posterior distribution (see Section 0 below). If estimates are required for other combinations of skill levels it will be necessary to algebraically approximate the CIs as described in Section 6 of the *Technical Report*,¹ although the estimates themselves can be simply summed.

Table 1 Estimates provided in Excel files reporting all skill level estimates

Literacy: (a) counts & (b) proportion	EL1 & below	EL2	EL3	L1	L2 & above	
Numeracy: (a) counts & (b) proportion	EL1 & below	EL2	EL3	L1	L2 & above	
Email: (a) counts & (b) proportion	EL1 & below	EL2	EL3	L1	L2 & above	
Word Processing: (a) counts & (b) proportion	Below EL	EL1	EL2	EL3	L1	L2 & above
Spreadsheets: (a) counts & (b) proportion	EL2 & below	EL3	L1	L2 & above		
ICT Multiple Choice: (a) counts & (b) proportion	Below EL	EL1	EL2	EL3	L1	L2 & above
English not a first language: (a) counts & (b) proportion	Not ESOL	ESOL	ESOL = ‘English Spoken as an other language’			

Each ‘counts’ table also gives the reference population for each local area, which is the number of people aged 16-64 living in households.

In addition to the above seven geographies, regional estimates have also been provided, however, unlike the other geographies, these estimates are based on direct survey estimates and not modelled estimates.

The estimates are made available through a total of 17 Excel files. The file-naming convention describes (a) the geography, (b) the population base and (c) whether the file refers to all reported skill levels (suffix “_A”) or just “Entry Level and below” and “Level 1 and above” (suffix

¹ Gibson, A. and Hewson, P. (forthcoming) *2011 Skills for Life Survey: Small Area Estimation Technical Report*. Department for Business, Innovation and Skills, Research Paper.

“_B”). All estimates assume the socio-economic characteristics of local populations to be as detailed in the 2001 Census, although all except a single set of MSOA-level estimates have been updated to reflect the ONS’s mid-year 2009 age-sex population estimates.

Table 2 Small Area Estimation Data: Excel Files and their contents

Geography	Population Base	Filenames (all have ‘.xlsx’ suffixes)	
		All Skill Level Estimates	EL & below and L1 & above categories
Middle Layer Super Output Areas	2001	<i>MSOA(2001)_A</i>	–
Middle Layer Super Output Areas	2009	<i>MSOA(2009)_A</i>	<i>MSOA(2009)_B</i>
Standard Table wards	2009	<i>STWards(2009)_A</i>	<i>STWards(2009)_B</i>
2005 Statistical wards	2009	<i>StatWards(2009)_A</i>	<i>StatWards(2009)_B</i>
2011 Council wards	2009	<i>CouncilWards(2009)_A</i>	<i>CouncilWards(2009)_B</i>
2011 Parliamentary Constituencies	2009	<i>ParlCons(2009)_A</i>	<i>ParlCons(2009)_B</i>
Local Authorities	2009	<i>LAs(2009)_A</i>	<i>LAs(2009)_B</i>
Local Enterprise Partnership areas	2009	<i>LEPs(2009)_A</i>	<i>LEPs(2009)_B</i>
Region	Survey	<i>Region(survey estimates)_A</i>	<i>Region(survey estimates)_B</i>

Middle Layer Super Output Areas (n = 6,781): MSOAs were devised by the ONS as part of a hierarchy of units specifically designed for the collection and publication of small area statistics. They are of broadly consistent size (of about 7,200 people) and are not subject to boundary changes. They are the *de facto* standard geography for which most ONS and other administrative and statistical data are published. MSOAs have a Government Statistical Service (GSS) code starting E02.

2001 population base: Estimates have been produced by applying the model to population covariates derived from the 2001 Census and 4th Quarter 2010 DWP benefit data (weighted to match the 2001 Census population). These figures have not been adjusted to account for any population change since 2001.

2009 population base: Estimates have been produced by applying the model to population covariates derived from the 2001 Census and 4th Quarter 2010 DWP benefit data, weighted to fit ONS mid-year age-sex population estimates for 2009. These figures have thus been adjusted to account for known demographic population changes since 2001, although it is necessary to assume that MSOA populations have remained static in terms of their socio-economic composition.

Estimates for all other geographies assume a 2009 population base and are derived from the 2009 MSOA estimates, being proportionally attributed on the basis of addresses in the February 2011 *Open National Statistics Postcode Directory (ONSPD)*.

Standard Table (ST) Wards (n = 7,932): ST wards were used to release 2001 Census outputs with a higher disclosure risk than was acceptable using Census Area Statistics (CAS) wards. ST wards were based on the 2003 statistical ward set, though a total of 113 wards (those with fewer

than 1,000 residents or 400 households) were merged into neighbouring wards. ST wards were used to report the small area estimates derived from the *SfL2003*.²

2005 Statistical Wards (n = 7,972): In 2003 a policy was introduced across National Statistics to minimise the statistical impact of frequent electoral ward boundary changes, particularly in England. The wards resulting from this policy, which were fixed annually, were known as 'statistical wards'. A change of policy meant that the last set of statistical wards were for 2005. Statistical wards are now rarely used, although postcodes are still assigned to 2005 statistical wards in the *National Statistics Postcode Directory*.

2011 Council Wards (n = 7,618): Also known as Electoral Wards/Divisions, these are subdivisions of Local Authorities and are the key building block of UK administrative geography, being the spatial units used to elect local government councillors in metropolitan and non-metropolitan districts, unitary authorities and the London boroughs in England. The ward geography used in this report is that current as of the 1st January 2011, and is detailed in the February 2011 *ONSPD*. Council wards have a GSS code starting E05.

2011 Parliamentary Constituencies (n = 533): The English Parliamentary Constituencies used in this report relate to those defined by the Parliamentary Constituencies (England) Order 2007 and the Parliamentary Constituencies (England) (Amendment) Order 2008. They came into effect at the May 2010 General Election. Their composition is detailed in the February 2011 *ONSPD* and no further changes are envisaged until 2014/2015. Parliamentary constituencies have a GSS code starting E14.

Local Authorities (n = 326): The local authorities to which this study refers are those 'district level' (or 'lower tier') authorities current as of the 1st January 2011, thereby including the structural changes effected on 1st April 2009. There are a total of 326 local authorities; comprising 36 metropolitan districts (with a GSS code starting E08), 201 non-metropolitan districts (E07), 56 unitary authorities (E06), 32 London boroughs and the Corporation of the City of London (E09). Their composition is detailed in the February 2011 *ONSPD*.

Local Enterprise Partnership (LEP) areas (n = 37): Each LEP comprises a number of local authorities, and some Local Authorities are in more than one LEP. The composition of LEPs (in terms of their constituent Local Authorities) is detailed in www.bis.gov.uk/policies/economic-development/leps/statistics (accessed 22/04/2012). When the estimates in this report were produced there were 37 LEPs. Shortly thereafter, in late September 2011, a 38th LEP – Northamptonshire – was announced. A further LEP has been announced more recently – Buckinghamshire Thames Valley. These two LEPs have not been included in the analysis. It is likely that additional LEPs will be formed over the coming years, and basic skills estimates for these, as well as for the Northamptonshire and Buckinghamshire Thames valley LEPs, can be approximated using the method described in Section 6 of the *Technical Report*.

² Gibson, A., Bailey, T, and Fraser, D. (2004) *Demographic mapping of the 2003 Skills for Life Survey to local areas*. Technical Report for the Department for Education and Skills, December 2004.

Survey Background

Survey introduction

The first Skills for Life Survey was commissioned by the Department for Education and Skills (DfES) in 2001. The survey aimed to produce a national profile of adult literacy, numeracy and Information and Communication Technology (ICT) skills, and to assess the impact different skills had on people's lives.

In 2010, the Department for Business, Innovation and Skills (BIS) commissioned a follow up to the initial survey, with the key purpose of updating the baseline information collected about adult literacy and numeracy as well as setting a new functional baseline for ICT skills. A key consideration in designing the new (2011) survey was maintaining comparability with the baseline (2003) survey, to enable measurement and analysis of changes in skills levels amongst the population between the respective periods. A further objective of the *2011 Skills for Life Survey (SfL2011)* was to understand the demographic, social and motivational factors related to skills levels in today's population of 16-65 year olds.

The *SfL2011* headline findings paper is available at:

<http://www.bis.gov.uk/assets/biscore/further-education-skills/docs/0-9/11-1367-2011-skills-for-life-survey-findings.pdf> (accessed 27/7/2012).

The full *SfL2011* survey report will be published in Summer 2012.

Survey methodology

The interview

The interview comprised a background questionnaire followed by a pre-assigned random combination of two of the three skills assessments: literacy, numeracy and ICT. The background questionnaire was designed to collect a broad set of relevant demographic and behavioural data. A refined and updated version of the 2003 background questionnaire was used; redundant items were removed and some new questions were added.

The literacy and numeracy assessment tools were the same as those used in the 2003 Skills for Life survey (*SfL2003*) to ensure absolute comparability with the 2003 survey. Respondents who completed the questions allocated to the literacy and numeracy assessments were assigned to one of the five lowest levels of the National Qualifications Framework: Entry Level 1 or below; Entry Level 2; Entry Level 3; Level 1; or Level 2 or above.

A new ICT assessment tool was, however, included. Major advances in software had made possible the development of a more valid and reliable assessment tool for ICT, using real office desktop applications. The ICT assessment consisted of four components: word processing, emailing, spreadsheets and a set of multiple choice questions assessing ICT skills such as internet use. Respondents allocated to the ICT assessment were assigned to one of five skills levels mentioned above for each of the four components (a 'total' aggregated ICT level is not calculated).

Level 1 is broadly comparable with a low level GCSE (D-G), while Level 2 is comparable with a high level GCSE (A*-C). Entry Level is therefore below a low level GCSE. The Annex to this report offers a brief definition of these levels.

Fieldwork

The survey population was all adults aged 16-65, normally resident in England. Residents of institutions were excluded for practical reasons. The interview was carried out via Computer-Assisted Personal Interviewing (CAPI) and on average lasted 70 minutes. Fieldwork was carried out between May 2010 and February 2011.

In total, 7,230 interviews were conducted, with 57 per cent of assumed eligible households taking part in the survey. Of the 7,230 completed interviews 6,049 respondents were assigned to the literacy assessment, 6,053 were assigned to the numeracy assessment and 2,358 to the ICT.

Data weighting and imputation

The weighting strategy used was matched to that used for the *SfL2003* survey. A sampling weight was calculated for each respondent, and this sampling weight was then used as the base weight - or starting position – for an iterative sequence of ‘rim weighting’ to population totals. In keeping with the 2003 survey, these population totals were (a) the cross-classification of sex and age, and (b) region, and were drawn from the most recent Office for National Statistics (ONS) mid-year population estimates (2009).

In line with *SfL2003*, the assessment scores for some respondents were ‘imputed’ where they did not complete the assessment. For the literacy and numeracy assessments, the approach taken was identical to that in 2003 to maintain comparability: respondents who either (a) said they could not read English or (b) assessed their reading level as ‘poor’ and chose not to attempt the assessments are given Entry Level 1 or below values for all assessments.

In addition, respondents who had never used a computer are given the lowest value for all the ICT assessments (below Entry Level 1 for multiple-choice and word processing, below Entry Level 2 for email, below Entry Level 3 for spreadsheets).

Any other cases missing a skill Level are excluded from analysis that covers that skill Level.

Modelling

Top line geographical area estimates

The weighted *SfL2011* survey provides a sound basis upon which to estimate levels of literacy, numeracy and ICT skills in England as a whole and for individual regions. Some of the key findings for literacy and numeracy are:

- 29.0 million 16-65 year olds (85 per cent) were at Level 1 or above in literacy, with 5.1 million (15 per cent) at Entry Level 3 or below.
- 26.0 million 16-65 year olds (76 per cent) were at Entry Level 3 or above in numeracy, with 8.1 million (24 per cent) at Entry Level 2 or below.
- ICT overall distributions are shown in Table 3.

Table 3 ICT Levels in England

Level	Word Processing (n=2,253)		Email (n=2,247)		Spreadsheet (n=2,228)		Multiple Choice (n=2,274)	
	%	Adult Population Estimate (millions)	%	Adult Population Estimate (millions)	%	Adult Population Estimate (millions)	%	Adult Population Estimate (millions)
Below Entry Level 1	15	5.0	30	10.4	39	13.2	8	2.6
Entry Level 1	12	4.0					*	0.2
Entry Level 2	17	5.8	1	0.3			1	0.4
Entry Level 3	16	5.6	9	2.9	27	9.3	12	4.2
Level 1	15	5.2	8	2.6	17	5.8	26	8.8
Level 2 or above	25	8.6	52	17.9	17	5.8	53	17.9
Unweighted Total		34.1		34.1		34.1		34.1

Base: *SfL2011* All aged 16-65 with word processing scores / emailing scores / spreadsheet scores / multiple choice scores

Estimates presented at national and regional level are direct survey estimates.

Local Area Estimates

The *SfL2011* is, however, clearly unable to support direct estimation at the sub-regional level. It is not just that the sample size is too small, but that many of the wards and MSOAs for which estimates are required are not even represented in the national survey. A model-based approach has thus been adopted.

The idea here is that survey data, which includes information on the socio-demographic characteristics of individuals, are interrogated in order to derive models which best describe how a dependent variable (for instance numeracy skills) responds to individual and area-level predictor variables (e.g. age, sex, educational qualifications and, at an area level, the proportion of people with low incomes in the MSOA in which the survey respondent lives).

Local area estimates for the dependent variable are then calculated by applying the model's parameter estimates to the corresponding covariate values for the local areas. In effect, the goal is to 'pool' evidence from across the wider sample in order to generate local estimates.

Modelling Local Area Estimates

Multilevel models have been specified in order to capture the hierarchical structure within which processes are presumed to operate (whereby, for instance, it is assumed that individuals' literacy skills can best be predicted with reference to both their individual social-demographic characteristics and aspects of the group of which they are a part).

The dependent variables modelled were (a) whether or not individuals speak English as a first language, for which a *multilevel binary logistic regression model* was specified, and (b) a series of skills levels for literacy, numeracy and the four ICT skill domains, for which *multilevel ordinal logistic regression models* were specified. The latter were constructed as cumulative link models relating an assumed continuous latent variable to specific skills levels using modelled 'cut points' between skill Levels.

As detailed in the *Technical Report*, a Bayesian approach has been adopted. Arguably more straightforward and flexible than classical, or frequentist, approaches, this provides comprehensive information about the estimates generated. In other words, by modelling the full posterior predictive distribution of estimates (in effect a large number of independent estimates of, for instance, the number of adults with Level 1 literacy skills) it is possible to derive empirically both a 'point estimate' of the number of adults with Level 1 literacy skills (the mean of the posterior distribution) and a 95 per cent 'credible interval' around that point estimate (the range within which 95 per cent of those posterior estimates lie). This defines the range within which we are 95 per cent certain the true value lies.³

Variable Selection and Model Fitting

A key constraint in small area estimation is that individual-level variables must be identically defined in both the survey from which the model was derived and the census, or other source, from which area-level covariate data are drawn. Fortunately, it is now standard practice for most surveys and censuses to elicit a well-defined set of socio-demographic characteristics and many of the socio-demographic questions asked in the *SfL2011* mirror precisely those asked in the *2001 Census*. Questions regarding benefit status, meanwhile, can be related to Lower Layer Super Output Area (LSOA)-level data on the number of people claiming benefits in the 4th Quarter of 2010 as recorded in the DWP's *Work and Pensions Longitudinal Study (WPLS)* dataset (available via NOMIS, <http://www.nomisweb.co.uk/>, accessed 27/7/12).

Parameter selection (including a systematic search for possible interaction effects) was undertaken using standard automatic and manual selection procedures, yielding a parsimonious

³ This approach is described in more detail in the Technical Report, which also includes a glossary defining terms such as 'posterior distribution'.

set of candidate predictor variables. These were fitted using ‘approximate Bayesian’ methods to ensure that each model was suitable and, simultaneously, to consider a variety of potential Middle Layer Super Output Area (MSOA)-level predictor variables. On this basis ‘low income’ (i.e. the proportion individuals in each MSOA deemed to be income deprived according to the authors of the *2010 Index of Multiple Deprivation*⁴) was identified as being a sufficient proxy for underlying differences between MSOA populations.

Having obtained a parsimonious set of candidate variables for each model, as listed in Table 4 below, the public domain and widely used *rjags* software⁵ – a program for Bayesian analysis of complex statistical models using Markov chain Monte Carlo (MCMC) techniques – was used to obtain the posterior distributions of all parameters.

⁴ McLennan, D., *et al.* (2011) *The English Indices of Deprivation 2010*, Department for Communities and Local Government. (Available at <http://www.communities.gov.uk/publications/corporate/statistics/indices2010>.) [Accessed 12/1/2012.]

⁵ Martyn Plummer (2011). *rjags: Bayesian graphical models using MCMC. R package version 3-3.* (Available at <http://CRAN.R-project.org/package=rjags>.) [Accessed on 12/2/2012.]

Table 4 Individual-level variables used in each model

Variable (Factors)	Literacy Model	Numeracy Model	Email Model	Word Processing Model	Spreadsheets Model	ICT Multiple Choice Model	English 'not as first language' Model
Sex (2 factors)	√	√		√	√	√	√
Ageband (4)	√		√	√	√	√	
Ethnicity (6)		√	√	√	√	√	√
Birth Place (2)						√	√
Highest Qualification (6)		√	√	√	√	√	
Limiting Long-term Illness (2)	√	√	√		√		
Tenure (3)	√	√	√	√	√	√	√
Occupation (SOC2000) (10)	√	√	√	√	√	√	√
Economic Activity Status (4)	√	√	√	√	√	√	
Benefit Status: IB or SDA (2)	√	√	√	√	√	√	√

Note: These models are predictive rather than interpretative. Thus whilst 'highest qualification' is highly correlated with literacy, it does not form part of the final model because an alternative set of predictor variables offer a better fit with the data.

Having derived the posterior distributions (comprising 1,000 simulated values) for each model parameter, the next stage is to apply them to *individuals* and to then aggregate and summarize the resulting responses at MSOA-level. This is more challenging than the approach used to generate local area estimates from the *SfL2003* (where model parameter posterior distributions were applied to aggregated *area-level* data) because it requires information on the composition of local populations. The idea, though, is that this new approach should significantly improve estimate accuracy as valuable information about local areas will not have been lost through aggregation.

Determining the composition of local populations

Unfortunately, direct information concerning the detailed composition of local populations does not exist. It is, however, possible to use what is known about the aggregate characteristics of any given population (i.e. how many males and females, how many people with or without a limiting long-term illness, how many people in each age band, etc.) in order to 'microsimulate' the likely number of people with each unique combination of characteristics (for instance the number of 16-34 year-old males with a limiting long-term illness; the number of 16-34 year-old females with a limiting long-term illness; and so on). The defining characteristic of a successfully microsimulated population is that, when aggregated, it will match in all respects what is known about the overall characteristics of that population.

The data used to generate microsimulated populations for all 6,781 MSOAs in England have been drawn from a series of univariate and multivariate 2001 Census tables (listed in Table 5 below) as well as from the *WPLS* dataset (regarding 4th Quarter 2010 claimants of Incapacity Benefit and Severe Disablement Allowance). As detailed in the *Technical Report*, the goal is to iteratively assign people to 'cells' in the simulated populations until the resulting marginal totals match the 'known' marginal totals derived from the 2001 Census and DWP *WPLS* dataset. Disclosure control measures used to ensure confidentiality of both census and DWP data mean that microsimulation cannot precisely reconstitute the socio-demographic composition of individual MSOA populations, but there is no question that it provides a close approximation.

More problematic is the fact that the 2001 Census was undertaken the best part of a decade before the *SfL2011*. **Demographic** changes since the 2001 Census have been accounted for by adjusting the estimates so that they align with the most recent (2009) ONS mid-year age-sex population estimates available at the time of analysis, but it is inevitable that in some areas there will have been significant changes to the **socio-economic** composition of populations which cannot be captured until the results of the 2011 Census are published.

Table 5 2001 Census tables used to ‘microsimulate’ MSOA populations

Census Area Statistics (CAS) Tables	
CAS001	Age by sex and whether living in a household or communal establishment
CAS016	Sex and age by general health and limiting long-term illness
CAS017	Tenure and age by general health and limiting long-term illness
CAS021	Economic activity by sex and limiting long-term illness
CAS026	Sex and economic activity by general health and provision of unpaid care
CAS032	Sex and age and level of qualifications by economic activity
CAS033	Sex and occupation by age
CAS034	Former occupation by age
CAS061	Tenure and car or van availability by economic activity
CAS105	Age by highest level of qualification
CAS113	Occupation by highest level of qualification
CAS Theme Tables	
CAST03	Theme table: ethnic group cross-tabulated by (a) sex; (b) ageband; (c) birthplace; (d) economic activity; (e) limiting long-term illness; and (f) resident type.
Key Statistics Tables	
KS05	Country of birth
Univariate Tables	
UV03	Sex
UV04	Age
UV09	Ethnic group (England and Wales)
UV22	Limiting long-term illness
UV24	Qualifications (England and Wales)
UV28	Economic activity
UV30	Occupation
UV43	Tenure (people) (England, Wales and Northern Ireland)
UV71	Communal establishment residents

Generating local area predictions

MSOAs are the *de facto* standard geography for which statistical data are published. Up-to-date covariate data thus tend to be available for MSOAs and, as a result, this provided the obvious ‘upper level’ for the multilevel models. MSOA populations were thus microsimulated on the basis of 2001 Census and DWP *WPLS* data, and the posterior model parameter estimates were applied to derive 1,000 independent estimates of the likelihood that each microsimulated individual would have the various literacy, numeracy and ICT skill levels. Aggregating these likelihood estimates to MSOA-level provides a direct estimate of the number and proportion of adults in each of the skill levels (or who do/do not, speak English as a first language). These estimates were then weighted to align with the ONS’s 2009 age-sex population estimates. This accounts for known demographic changes, although the estimates still assume socio-economic characteristics of MSOAs are as reported in the 2001 Census.

Estimates were also required for a number of other geographies for which little or no relevant covariate data are available. For these geographies it has been necessary to attribute MSOA-

level estimates on the basis of the February 2011 *Open National Statistics Postcode Directory (ONSPD)*. This provides a count of residential addresses in each postcode, and lists within which higher geographies – MSOAs, Local Authorities, Parliamentary Constituencies, etc. – these postcodes lie. On this basis it is possible to derive address-weighted lookup tables relating MSOAs to each of the other geographies of interest.

These lookup tables were used to allocate the 2009 MSOA-level posterior estimates as appropriate, and mean estimates and 95% CIs for the new geographies were derived directly from the re-distributed and re-aggregated sets of posterior estimates. In this way full information is retained concerning model uncertainty. Whilst a perfectly adequate approach with respect to Local Authorities, Parliamentary Constituencies, and Local Enterprise Partnership areas (all of which are essentially aggregations of MSOAs), the degree of spatial overlap with the various ward geographies raises concerns. Geographical variations in household size and composition means this use of addresses will provide only an imperfect guide to the distribution of adult residential populations between different spatial units. **The various ward estimates must therefore be considered secondary to those produced for MSOAs.**

Summary of results

As illustrated by Figure 1 below, MSOA-level results exhibit impressive granularity and serve to highlight and quantify the extent to which adult skills vary at the very local level. For instance, whilst Entry Level and below literacy rates vary from 11.3% (Wokingham) to 21.6% (Knowsley) at local authority level, at MSOA-level the proportion ranges from 8.8% (Basingstoke & Deane 021) to 33.6% (Liverpool 022). As described in the *Technical Report*, this pattern is repeated across the numeracy and ICT skills domains, and it is clear that areas of adult skill deficits can be very localised.

It is also apparent that there are marked variations in skill levels within individual local authority areas. This is pronounced for many large urban authorities – for instance Liverpool, with 20.8% of adults with Entry Level or below literacy skills, includes individual MSOAs with rates varying from 11.6% to 33.6% – but rural areas are not exempt. Thus North East Lincolnshire (with an overall rate of 17.3%) contains individual MSOAs with rates varying from 13.7% to 28.3%. This suggests that any policy response to skills deficits will also need to be spatially fine grained.

Figure 1 also draws attention to the fact that adult skills deficits are not a purely urban phenomenon, although the densest concentrations of poor literacy, numeracy and ICT skills do appear to affect England's principal conurbations. As detailed in the *Technical Report*, this rural dimension appears particularly marked with respect to ICT skills, with relatively high rates of Entry Level 1 and below skill levels emerging in many south western and northern counties, as well as around the Wash.⁶

Turning to the distribution of adults who do not speak English as a first language (Figure 2) the pattern is much as might be expected, with high rates across most of London as well as in the centres of many of England's other major towns and cities – notably the old cotton towns of Lancashire. The glaring anomaly on the map concerns an improbably high proportion of adults not speaking English as a first language in a small part of rural East Anglia.

⁶ Choropleth maps illustrating the proportion of adults with low literacy, numeracy and ICT skills at both MSOA and local authority level are included in the *Technical Report*. A local authority level map of the proportion of adults not speaking English as a first language is also included in the *Technical Report*.

Thus the model predicts that over 30% of adults in the Forest Heath 002 MSOA (E02006239) do not speak English as a first language, and that the rate exceeds 10% in another four Forest Heath MSOAs. In fact, Forest Heath is distinctive because of the very high proportion of non-UK born household residents recorded in the *2001 Census* (50.3%). Elsewhere in the country such high percentages are invariably associated with immigrant populations, but here it is due to the presence of two large American airbases at Lakenheath and Mildenhall.

In the *SfL2011*, as in the country as a whole, being non-UK born is strongly associated with not speaking English as a first language. In Forest Heath this relationship will almost certainly not apply, but without sampling a number of individuals from that particular MSOA it is impossible for the model to recognise and reflect Forest Heath's uniqueness. In this respect the methodological 'strength' of small area estimation – the fact that it 'pools' the evidence of modelled relationships derived from an analysis of the dataset as a whole – is also its principal weakness. It cannot isolate and reflect genuine local exceptions such as Forest Heath. This is inevitable when a survey of 7,230 individuals is being used to derive estimates across a total of 6,781 MSOAs.

Figure 1 Entry Level and Below Literacy Skills: English MSOAs (n=6,781)

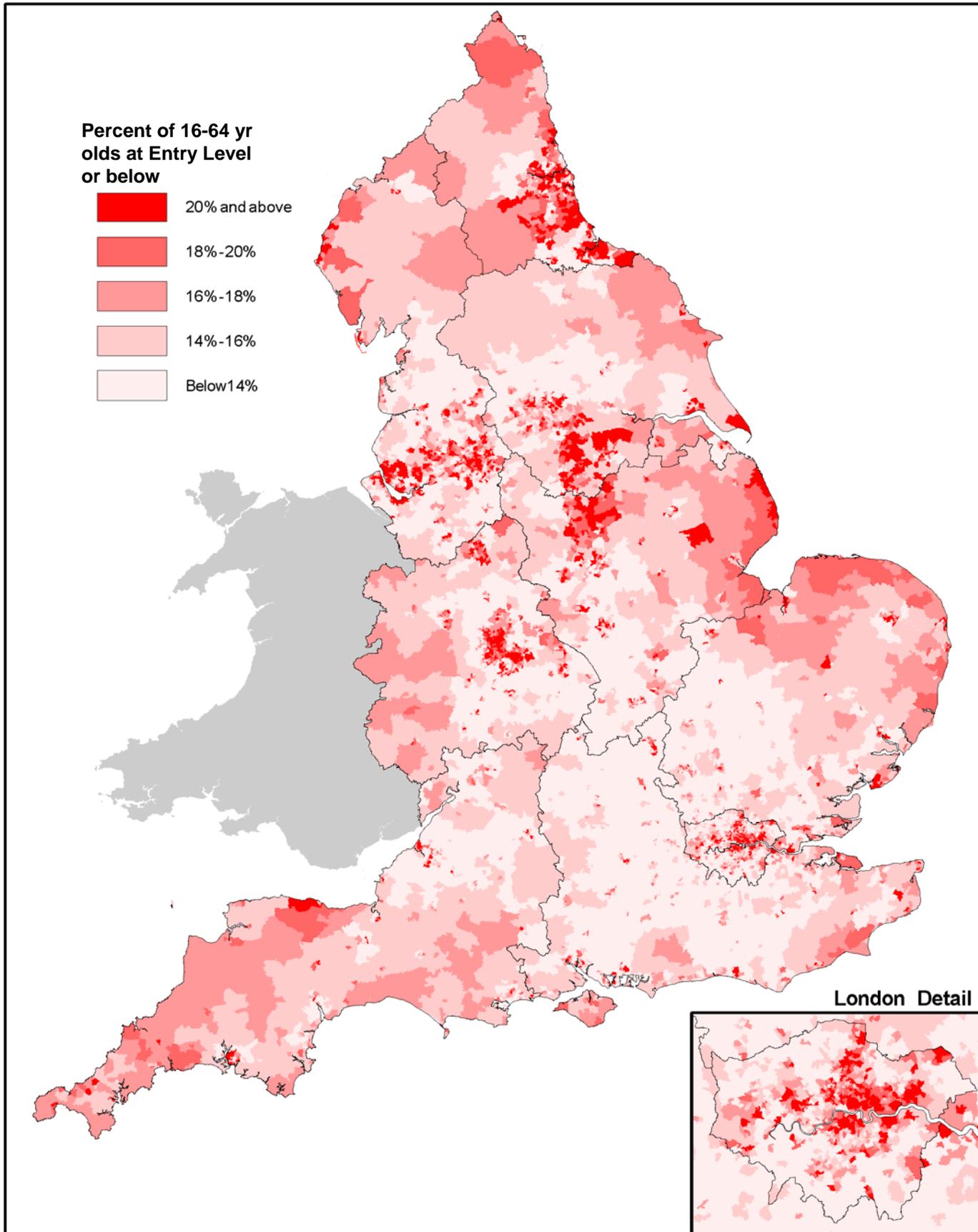
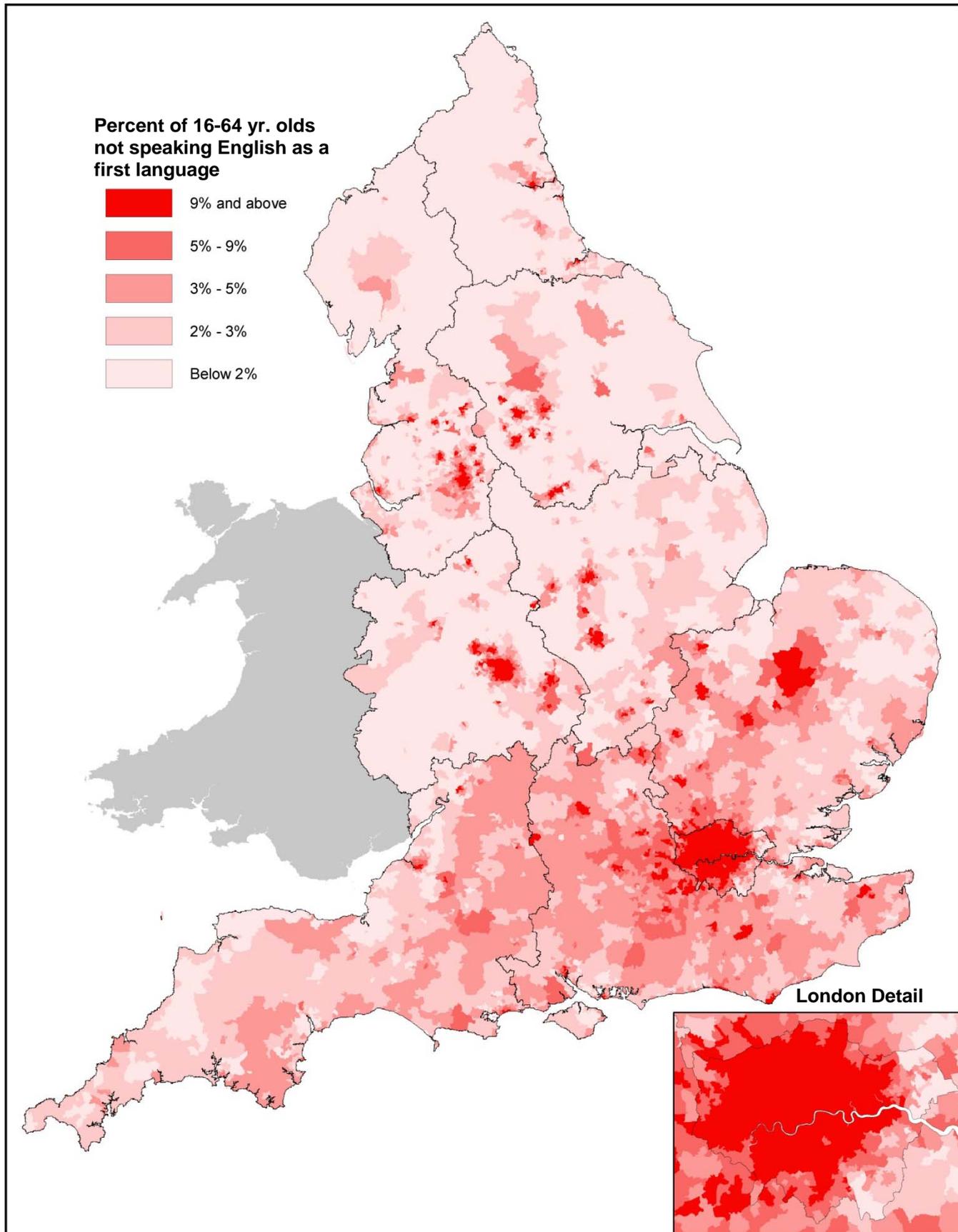


Figure 2 Proportion of adults not speaking English as a first language: English MSOAs (n=6,781)



Relationship between ‘direct’ survey estimates and ‘modelled’ small area estimates

The *SfL2011* dataset has been used to generate estimates of literacy, numeracy and ICT skill levels for every MSOA in England. Aggregation of these estimates to regional or national level will produce alternatives to the ‘direct’ estimates used in this report. For the most part, the differences between the two are slight, but there are several instances in which one of the estimates lies outside of the ‘credible range’ of the other. This prompts the question ‘which is the more accurate?’

It is important to stress that the two methods treat the survey data in a different way. With direct estimation, it is being treated as a sample drawn from the population. With modelled estimation it is being used as a source for identifying and quantifying relationships between demographic characteristics and literacy, numeracy or ICT skill levels.

It is also important to recognise that both sets of estimates hold only in a conditional way.

For the direct estimates used in the *SfL2011* report, the conditions include (a) that the weights employed effectively counter any systematic non-coverage or non-response bias for all population sub-groups, and (b) that the sampling distribution has a ‘normal’ shape with a mean equal to the true population value.

For the modelled estimates, the conditions include (a) that the small area population data – largely drawn from the 2001 Census – are accurate, (b) that the survey data allow unbiased estimates of model parameters to be extracted, (c) that the population and survey data are comparable, and (d) that the processes used to microsimulate the composition of local populations, and to simulate parameter posterior distributions, are appropriate to the task and correctly implemented.

Given the variety of conditions placed on both sets of estimates, some conflicts are to be expected, even if (as here) the majority of estimates are close matches for one another.

Ultimately, it is impossible to answer the question ‘which is the more accurate?’ without full and current population data. The choice between estimation methods depends on the degree of belief about the assumptions that underpin each method.

The Department has chosen to use the modelled estimates only for sub-regional geographies for which direct estimates are unreliable due to sample size limitations; however, direct survey results are used for national and regional-level estimates. There are two reasons for this. First, the direct estimates are simpler to present because they are based on traditional survey methods. Second, because the modelled estimates have been computed at MSOA-level and only then aggregated to regions. Any small systematic biases at MSOA level may be magnified at higher order geographies.

Caveats

When reviewing and interpreting small area estimation data it is essential that the following caveats are borne in mind:

- As noted in Section 0, all estimates necessarily assume that local populations have the same socio-economic composition as they had in 2001. Demographic (age-sex) changes since the 2001 Census have been accounted for by adjusting the estimates to align with the ONS's *experimental* quinary population estimates for mid- 2009. This is likely to be a generally reasonable assumption but, as described in Section 4 the *Technical Report*, in some areas this assumption will not hold. It should also be recognised that these estimates only apply to people living in households – communal establishments were excluded from the *SfL2011* sample and the model parameters derived from our analysis of the *SfL2011* have been applied to microsimulated populations describing the socio-demographic composition of local household populations.
- It is important to view the estimates in the light of local knowledge – partly because of the data limitation noted above, but also because small area estimation cannot capture locally anomalous or unusual patterns in the distribution of adult basic skills. Thus if a particular group of people, nationally, is found to have particularly high skill levels, then it is assumed that this will apply to all local areas. The multilevel nature of the model will reveal whether this relationship is mediated by any MSOA-level effects, but the same principle applies – these MSOA-level effects are also based on modelled relationships derived from an analysis of the dataset as a whole. Local knowledge should be used to decide whether or not an area is likely to be atypical.
- Given the above, it is important to recognise that small area estimation is effective at identifying and highlighting those areas where adult literacy, numeracy and ICT skill levels are *likely* to be relatively poor and where, therefore, it is *likely* that policy initiatives would be best directed. The results cannot and should not be used as the basis for some sort of 'performance league table' – ranking individual MSOAs, LAs or other geographic or policy units on the basis of, say, predicted adult literacy rates.
- As noted in Section 0, small area estimation was undertaken with respect to MSOA populations. These estimates were then attributed to other geographies on the basis of addresses listed in the February 2011 *Open National Statistics Postcode Directory*. For this reason, as well as the fact that MSOAs are today the *de facto* standard geography for which very local spatial administrative and statistical data are published, the various ward-level estimates must be considered secondary to those produced for MSOAs.
- As noted in Section 0, estimates can be summed across skill categories and/or across a number of individual geographic units (e.g. to combine a series of LA-level estimates to produce a new estimate for a composite Local Enterprise partnership area). Credible

Intervals must be treated differently, however, and the upper and lower 95% CIs cannot be summed in the same way. If 95% CIs are required for combined skill levels and/or aggregated geographic units then they will have to be approximated as described in Section 6 of the *Technical Report*.

- As noted in Section 0, aggregation of MSOA-level literacy, numeracy and ICT estimates to the regional or national level will produce alternatives to the 'direct' estimates from the survey itself. The Department has chosen to use direct national and regional estimates, and modelled estimates for all sub-regional geographies.
- Finally, the estimates are necessarily constrained by the data collected as part of the *SfL2011* survey and the extent to which they could be reconciled with local covariate data. Thus a broader range of appropriate predictor variables would, at least in theory, have enabled us to develop better models and thus produce estimates that were more sensitive to the specific characteristics of some local populations.

References

Gibson, A. and Hewson, P. (forthcoming) *2011 Skills for Life Survey: Small Area Estimation Technical Report*. Department for Business, Innovation and Skills, Research Paper.

Gibson, A., Bailey, T, and Fraser, D. (2004) *Demographic mapping of the 2003 Skills for Life Survey to local areas*. Technical Report for the Department for Education and Skills, December 2004.

Harding, C., et al (2011) *2011 Skills for Life Survey: Headline findings*. Department for Business, Innovation and Skills Research Paper Number 57, available online at: <http://www.bis.gov.uk/assets/biscore/further-education-skills/docs/0-9/11-1367-2011-skills-for-life-survey-findings.pdf>, accessed on 18/01/12.

Harding, C., et al (forthcoming) *2011 Skills for Life Survey: a survey of Literacy, Numeracy and ICT Levels in England*. Department for Business, Innovation and Skills Research Paper.

Martyn Plummer (2011). *rjags: Bayesian graphical models using MCMC. R package version 3-3*. Available at <http://CRAN.R-project.org/package=rjags>, accessed on 12/2/2012.

McLennan, D., et al. (2011) *The English Indices of Deprivation 2010*, Department for Communities and Local Government. Available online at <http://www.communities.gov.uk/publications/corporate/statistics/indices2010>, accessed 12/1/2012.

Williams, J., S. Clemens, S. Oleinikova, and K. Tarvin (2003) *The Skills for Life Survey: a National Needs and Impact Survey of Literacy, Numeracy and ICT skills*. Department for Education and Skills Research Report 490, available online at: <https://www.education.gov.uk/publications/eOrderingDownload/RR490.pdf>, accessed on 27/10/11.

Annex: Definition of Skills Levels⁷

Entry Level 1 is the national school curriculum equivalent for attainment at age 5-7. Adults below Entry Level 1 may not be able to write short messages to family or select floor numbers in lifts. Adults with ICT Entry Level 1 skills are able to get information from an ICT-based source and follow recommended safe practices.

Entry Level 2 is the national school curriculum equivalent for attainment at age 7-9. Adults with below Entry Level 2 may not be able to describe a child's symptoms to a doctor or use a cash point to withdraw cash. Adults with ICT Entry Level 2 skills are able to use ICT to communicate, as well as enter and edit small amounts of information in ways that are fit for purpose and audience.

Entry Level 3 is the national school curriculum equivalent for attainment at age 9-11. Adults with skills below Entry Level 3 may not be able to understand price labels on pre-packaged food or pay household bills. Adults with ICT Entry Level 3 skills are able to interact with and use an ICT system to meet needs, as well as present information in ways that are fit for purpose and audience.

Level 1 is equivalent to GCSE grades D-G. Adults with skills below Level 1 may not be able to read bus or train timetables or check the pay and deductions on a wage slip. Adults with ICT Level 1 skills are able to select and use a variety of appropriate sources of information, as well as enter, organise, develop format and bring together information to suit content and purpose.

Level 2 is equivalent to GCSE grades A*-C. Adults with skills below Level 2 may not be able to compare products and services for the best buy, or work out a household budget. Adults with ICT Level 2 skills are able to use a variety of appropriate sources of information and evaluate its fitness for purpose, as well as evaluate and use different methods of organising and presenting information, taking into account fitness for purpose and audience.

⁷ Level definitions adapted from National Audit Office (2008) *Skills for Life: progress in improving adult literacy and numeracy*, available online at: http://www.nao.org.uk/publications/0708/skills_for_life_progress_in_i.aspx, accessed on 17/8/11: p.10. ICT skills levels adapted from the Qualifications and Curriculum Authority (January 2007) *ICT Skills for Life curriculum*: Department for Education and Skills, available online at: <http://archive.niace.org.uk/Research/ICT/ICT-Skill-for-Life-curriculum-Jan07.pdf>, accessed on 17/8/11: p.19.

© Crown copyright 2012

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. Visit www.nationalarchives.gov.uk/doc/open-government-licence, write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

This publication is also available on our website at www.bis.gov.uk

Any enquiries regarding this publication should be sent to:

Department for Business, Innovation and Skills
1 Victoria Street
London SW1H 0ET
Tel: 020 7215 5000

If you require this publication in an alternative format, email enquiries@bis.gsi.gov.uk, or call 020 7215 5000.

URN 12/1316