

Summary of methodology used to calculate estimates of adult educational attainment within the population using Labour Force Survey data

1. Introduction

1.1. This note describes recent changes to the method used by BIS analysts to measure adult educational attainment within the population using Labour Force Survey (LFS) data.

1.2. The note covers the following areas.

Section 1	Introduction
Section 2	Background
Section 3	Overview of the statistical research leading to the methodological change
Section 4	Summary of new methodology
Section 5	Comparison of estimates from the new and old methodology
Annex A:	Relevant historical developments of the LFS questionnaire and survey method
Annex B:	Details of LFS qualifications question and classification
Annex C:	Details of new methodology
Annex D:	Details of old methodology

2. Background

2.1. As a result of statistical research published in February 2010, the Office for National Statistics (ONS) and BIS Heads of Profession for statistics decided to improve the method for calculating estimates of adult educational attainment.

2.2. Full details of the research that led to the methodological change, and BIS/ONS plans to take forward the recommendations, are available at: <http://www.thedataservice.org.uk/statistics/lfsmethodology>. A summary of the key aspects is included in this paper.

2.3. Changes to these estimates affect measurement of progress against the BIS PSA targets detailed below:

PSA Target: Improve the skills of the population, on the way to ensuring a world class skills base by 2020.

- 79 per cent of working age adults are qualified to at least Level 2 by quarter 4 of 2011;
- 56 per cent of working age adults are qualified to at least Level 3 by quarter 4 of 2011;
- 36 per cent of working age adults are qualified to Level 4 and above by quarter 4 of 2014, with an interim milestone of 34 per cent by quarter 4 of 2011.

PSA Target: Extending opportunities: Opening up opportunities for adults

- to reduce by at least 40 per cent the number of adults in the workforce who

lack Level 2 or equivalent qualifications between 2001 and 2010; working towards this, one million adults in the workforce to achieve level 2 between 2003 and 2006. This requires an additional 3.6 million adults in the workforce to be qualified to level 2 or higher by 2010, compared to 2001.

3. Overview of the statistical research leading to the methodological change

- 3.1. In July 2009, BIS and the ONS jointly commissioned a statistical research project to understand the strengths and weaknesses of the existing method of measuring adult educational attainment using the Labour Force Survey (LFS) and to identify the most optimal measure available for use at the present time. This project is part of wider programme of work, the LFS Education and Training Review (ETR), which is also seeking to improve the survey instrument.
- 3.2. The research was subject to external quality assurance from Denis Allnutt, who reported on his views on the quality and robustness of the research and its findings, alongside the [response from the Government Statistical Service](#) (GSS). Members of the GSS from BIS and ONS worked with the researchers and the Quality Assurer throughout the project to provide peer review for the research as it progressed.
- 3.3. An [interim report](#) on the statistical research conducted by RM Data Solutions was published on 22 October 2009, setting out the issues that the first phase of research had identified and what further research should be undertaken to fully understand these issues. BIS and ONS presented these findings to LFS user groups during November and December.
- 3.4. The [final research report](#) from RM Data Solutions was published on 24 February 2010 and included recommendations for the next steps and further work that BIS and ONS should consider as a result of the research findings. The key recommendation was a detailed proposal for amending the methodology for production of attainment estimates; this was considered necessary to compensate for weaknesses that had been identified in the old method.
- 3.5. The recommendations of RM Data Solutions were commented on by the Quality Assurer and reviewed and accepted by the GSS professionals in BIS and ONS. As a result, additional research will be taking place during summer 2010 to clarify what is the most suitable method for reporting on the PSA Target on Extending opportunities, referred to in paragraph 2.3 above.

4. Summary of new methodology

- 4.1. Under the new method, there are five key changes to the estimation process. These are listed below, with additional details set out in the remainder of this section. Full details of the rationale and process for each of these steps are set out in Section 4 of the [final research report](#).
 - A. Replacement of LFS data by NPD information for 19-21 year olds

- B. The sample in the reference quarter is “boosted” by the inclusion of cases where a response was given in another wave but not in the reference quarter (thus maximising use of information collected during the 5 waves of LFS)
 - C. Use of personal response data (from other waves) to replace information given by proxy, and preferring the final response given by a subject where more than one response was given in person
 - D. Adjustment of responses where matched ILR data from the previous 4 years definitively indicate a higher NQF level than reported
 - E. Imputation, based on personal responses from similar groups, to replace proxy responses not already adjusted, and responses where ‘Other’ is the highest stated qualification.
- 4.2. Step A, the replacement of LFS data by NPD information for 19-21 year olds requires application of qualifications estimates from DCSF’s [“Level 2 and 3 Attainment by Young People in England Measured Using Matched Administrative Data” Statistical First Release](#) to the LFS sub-population aged 19 to 21 who have been resident in England since academic age 14. This is considered a key step to address the difficulty in collecting attainment data from this age group in a social survey. Therefore, the new method makes full use of the authoritative administrative records held for this age group.
- 4.3. Step B is designed to ensure that LFS data are maximised where some respondents do not provide a response in quarter 4, but respond in another wave of the LFS, either prior to or after the reference quarter 4. It is possible to create a “boosted” sample by including those responses; this results in an effective response rate of around 75% in each quarter 4. Rules used in creating the boosted dataset are:
- Only including one household per single household address, giving precedence to households which responded in the reference quarter, or the quarter closest to it.
 - Removing any individuals not resident in a household which responded in the reference quarter but who provided a response in another quarter
- 4.4. Step C is designed to take account of the fact that, for reporting of qualifications, personal responses have been shown by previous research to be more accurate than proxy responses. It has also been demonstrated that personal responses to later waves of the LFS are more likely to be accurate than those in earlier waves. The decision rules for creating an amended micro-dataset are set out in Annex C; these span steps B, C, and D.
- 4.5. Step D requires ONS, under strict data protection controls, to match administrative ILR data cases to LFS data cases. ILR fields can subsequently be added to the anonymised LFS dataset for analytical use; where the respondent has not reported at an NQF level at which we can observe (from the ILR) they have achieved a qualification, the master micro-data set can be amended to show they hold a qualification at the correct level. The process for making these amendments is covered by the decision rules set out in Annex C of this document.

- 4.6. Step E is designed to handle a) the fact that NQF levels derived from responses given by proxy have been demonstrated to be lower than those derived from responses given in person by people with the same characteristics, and b) previous research has shown that responses recording 'other' qualifications can be shown to contain a similar range of qualifications as reported by those providing exact details of the qualifications they hold. Full details of the modelling are provided in sections 4 and 6 and Annex C of the [final research report](#). Some additional details are also provided in Annex C of this document.
- 4.7. **Revisions:** it should be noted that, because the sample boosting requires responses from survey waves occurring subsequent to the reference period, estimates for 2009 are provisional and will be finalised in March 2011. As set out in section 6 of the final research report, the provisional figures include an adjustment to account for the observed difference in previous years between estimates calculated at the reference quarter and those including all of the required data.
- 4.8. **PSA Target – Extending Opportunities:** while the method described above can be applied to the economically active adult population, this can only be done for 2006 and subsequent years. NPD data prior to this time are insufficiently complete, and the imputation models cannot be fully replicated. Therefore, the present means for assessing progress against the 2010 PSA Target under "Extending opportunities" (see paragraph 2.3) is to sum the growth from the revised series between 2006 and 2009 with the growth from the old series between 2001 and 2006. Further research will be conducted during summer 2010 to establish whether a better measure of growth between 2001 and 2010 is available.
- 4.9. **Further breakdowns for sub-groups:** previous Statistical First Releases contained a number of supplementary tables setting out differences in highest levels of qualification held by different groups and geographies within the population. However, the process developed to produce the revised England level estimates has not yet been adapted to produce equivalent supplementary tables. BIS is investigating feasibility of making such adaptations, with a view to publishing further tables later in 2010.

5. Comparison of estimates from the new and old methodology

- 5.1. Table 1 shows estimates calculated using the revised methodology for the percentages of working age adults in England holding different levels of qualification. This compares with equivalent estimates from the old methodology in Table 2. The key differences are:
- Results from the revised methodology show higher percentages qualified at the different levels
 - Growth between years is less volatile in the estimates from the revised methodology

Table 1: Estimates of percentages qualified calculated using revised methodology: England

Year	All people aged 19-59/64	Level 4 and above	Level 3 and above	Level 2 and above	Below Level 2 (incl no qualifications)
2006	29,417	32.2 ^P	52.3	71.7	28.3
2007	29,566	33.3 ^P	53.9 ^P	73.0 ^P	27.0 ^P
2008	29,736	33.6 ^P	54.4	73.9	26.1
2009	29,930	35.3 ^P	55.9 ^P	75.6 ^P	24.4 ^P
2009 95% CI (+/-)		0.4	0.4	0.4	0.4

Table 2: Estimates of percentages qualified calculated using old methodology: England

Year	All people aged 19-59/64	Level 4 and above	Level 3 and above	Level 2 and above	Below Level 2 (incl no qualifications)
2006	29,417	30.0	49.4	69.9	30.1
2007	29,566	30.9	50.6	70.7	29.3
2008	29,736	31.1	50.8	71.1	28.9
2009	29,930	33.1	53.1	73.6	26.4
2009 95% CI (+/-)		0.4	0.4	0.4	0.4

5.2. Similarly, Table 3 presents estimates calculated using the revised methodology for the number of economically active adults in England qualified to at least level 2. This compares with equivalent estimates from the old methodology in Table 3, which also shows the 2001 data point for this series.

Table 3: Estimates of numbers of economically active adults qualified calculated using revised methodology

Year	Economically active adults in England	Level 2 and above	Below Level 2 (incl no qualifications)
2006	24,233	18,423	5,810
2007	24,389	18,814 ^P	5,575 ^P
2008	24,682	19,201	5,481
2009	24,691	19,483 ^P	5,208 ^P

P. Provisional estimate - these estimates will be finalised in March 2011

Table 4: Estimates of numbers of economically active adults qualified calculated using old methodology

Year	Economically active adults in England	Level 2 and above	Below Level 2 (incl no qualifications)
2001	23,259	16,098	7,161
2006	24,233	17,933	5,810
2007	24,389	18,207	5,575
2008	24,682	18,433	5,481
2009	24,691	19,099	5,208

ANNEX A: Relevant historical developments of the LFS questionnaire and survey method

1. The LFS is a multipurpose survey aiming to provide good-quality data for a range of labour market measures and related topics. As the UK's largest regular household survey it covers around 130,000 individuals in 60,000 households each quarter. The survey began in 1973, being held every two years until 1983 when it became annual. Since 1992 it has been conducted quarterly.
2. Amongst other labour market information, the LFS collects data on a wide range of current and past qualifications. For most purposes, the level of highest qualification held is of more interest than details about the individual types of qualifications that a respondent holds. In the Education and Training section respondents are firstly asked to list their educational qualifications, by type (see Annex B for the current question, introduced in the Spring 2005 questionnaire). Respondents are then asked additional questions in order to determine the highest level at which individual qualifications are held.
3. Based on their responses, individuals are then assigned to the corresponding National Qualifications Framework (NQF1) level of their highest qualification held (further details of this assignment are in Annex B). These data are used widely across government and in particular for measuring the Skills PSA targets discussed in the main body of this note.

Timeline for development of the LFS

1979

Initially, working age people had responses coded against a list of 13 qualifications, which was assigned to NQF level equivalences as follows:

	NQF level equivalences				
	Level 7-8	Level 4-6	Level 3	Level 2	Below level 2
HIGHQUA					
First or higher degree	Yes				
Member of professional institution		Yes			
HNC,HND		Yes			
Teaching, secondary education		Yes			
Teaching, primary education		Yes			
Nursing qualification		Yes			
OND,ONC			70%	15%	15%
City & Guilds			20%	40%	40%
A level or equivalent			70%	30%	
Trade apprenticeship			50%	50%	
O level or equivalent				52.8%	47.2%
CSE below grade 1					Yes
Other Qualifications			10%	35%	55%
Don't Know	Pro rata	Pro rata	Pro rata	Pro rata	Pro rata
No Qualifications					

¹ www.qca.org.uk/493_15772.html

1983

New codes were added to identify Higher degree and First degree separately, allowing them to be assigned to Level 7-8, and Level 4-6 respectively.

1989

New codes were added to identify Teaching, further education (Level 4-6) and YTS certificates (Below level 2).

1992

Major methodological changes were introduced:

- The sample changed from a largely independent annual sample, to a quarterly rotating panel design - this allowed missing data to be imputed by bringing forward responses from the previous quarter.
- Computer assisted interviewing was introduced, together with data checks during the interview.
- The sample was expanded to include students in halls of residence and residents in NHS accommodation.

The qualifications question was also changed, so that all responses were to pre-coded categories and highest qualification was identified automatically, rather than by post interview coding. Coverage was extended to include people older than working age who were still in employment.

A Don't Know response was no longer permitted, and a new code was added to identify Teaching, level not stated (Level 4-6).

1993

The qualifications questions were revised to collect more detailed data on some qualifications. Awarding bodies such as BTEC (Edexcel), City & Guilds and RSA(OCR) award a variety of qualifications at different levels. Previously, their qualifications were coded as one category even though they include qualifications equivalent to Levels 4, 3, 2 and Below level 2. The list of qualifications was expanded to identify separately the different levels of these qualifications - resulting in improved data on vocational qualifications.

In addition, new questions on the number of A level (2 or more, 1) and GCSE/ O level passes (4 or more, less than 4) were introduced to enable more accurate conversion of these qualifications to NQF level equivalents.

New codes were also introduced to record Diploma in higher education, Scottish CSYS and SCE Highers, were previously reported under A levels.

1995

New question asking number of AS level passes (4+, 2 or 3, 1) introduced.

1996

Following research in 1994/5² the structure of the LFS questions on qualifications was altered with several changes being introduced from spring 1996 onwards. The four main changes were:

- Previously, the survey asked respondents for the three highest qualifications, but now asked respondents for *all* their qualifications. This change was made

² Employment Gazette, January 1998, pp33-39

as respondents are not always clear about which are their highest qualifications.

- The previous questionnaire had a long list of qualifications, in order from highest to lowest. Some qualifications were mentioned in several places as they are available at different levels, eg City & Guilds were mentioned three times (Advanced Craft, Craft and Other). From Spring 1996 onwards, respondents were asked what types of qualification (for example, City & Guilds, RSA, A levels, etc) they had. They were then asked relevant follow-up questions to determine the level of those qualifications.
- NVQs (and SVQs) were not covered in the original list of qualifications, but were asked about in a separate question. They were now included in the main question, although the additional question is still asked.

Additionally, GNVQs were now dealt with explicitly. Previously, they were put with A levels (for Advanced GNVQs) and GCSEs (Intermediate GNVQs). The Don't Know response code was also added back in, and the variable for GCSE/ O level passes was revised to record 5+, and <5 (rather than 4).

1997

From spring 1997, SCE Highers, and SCE Standard/Ordinary grades were treated as separate categories in the qualifications question. An additional question on the number of SCE Highers held was also introduced to more accurately assign respondents to their highest level of qualification held.

The change to calendar quarter data means data for Q1 1997 is unavailable. However the questions have remained the consistent for the rest of the year which means the syntax did not change during this period.

2004

The variable asking the number of GCSEs (A*-C)/O levels held was revised to gather more detailed information. The previous question asked whether people held 'Fewer than 5', or '5 or more'. The revision split the first of these categories into 1-2 and 3-4. This resulted in a much higher proportion of people answering Don't Know than previously.

This led to two sets of adjustments:

Imputation - We brought forward highest qualification data for those not contacted in Spring, deriving a new set of variables (with the suffix 'I' for imputed). This applies to Spring 2004 interviews only.

Num5up - CHAID analysis was undertaken to segment the population based upon their propensity to answer Don't Know to the number of GCSEs/O levels held. Responses to a derived variable NUM5UP, can be summed to provide an estimate of the total numbers holding 5 or more GCSEs (A*-C)/O levels.

The change to calendar quarters means there is currently no Q1 2004 data. When looking at Q2 found that original calculation of imputed variables hadn't included some cases so these variables were rederived and given "corrected" suffix. Therefore great care needs to be taken when using this dataset not to use the "normal" variables which are incorrect (they have been given labels warning not to use). The BIS syntax template is available with syntax covering

the period 1997-2008 and includes a set of syntax specific to Q2 2004. This particular fix has not been implemented on ONS datasets, annual datasets or longitudinal or household datasets..

2005

We updated our methodology to take account of all of the qualifications held by people when assigning them to their highest level of qualification. We derive our own versions of NUMAL, NUMSCE and NUMAS taking account of any other qualifications held that may be at a higher level.

We reverted back to the original version of the variable asking the number of GCSEs (A*-C)/O levels held. As some respondents had previous responses imputed forward, we continued to use Num5up for Spring 2005.

Due to the changes in spring 2005 there is currently no quals variables for Q1 2005. Num5up has been used up to Q2 2005.

2006

Interviewer feedback identified an issue recording Scottish qualifications. Scottish Highers could be coded under 2 different options (and subsequently routed to different options to record the number held). The code Higher (Scotland) was removed, whilst National Qualifications (Scotland) was revised to cover all Scottish qualifications. The follow up variable NUMSCE (number of Highers) was also removed with respondents instead being routed towards ADVHST or HST (number of Advanced Highers or Highers, respectively).

2007

In 2007 there was a change to record not just the highest degree held, but all types. There were problems with the implementation of this which meant that in Q1 07 there was a much larger proportion of other degrees that before and smaller numbers at higher levels (which was remedied by Q3 or certainly Q4 with improved interviewer instructions). This in itself did not affect the overall numbers at L4+ but there was an error in that degrees reported in Q4 06 but not as the first qualification at quals6 were not carried forward to Q1 07 as they should have been in cases where sample members did not respond at Q1 (see LFS User Guide Volume 1 for an explanation of Data Brought Forwards). DIUS wrote syntax to repair this problem – it has not been implemented on ONS datasets.

In Q1 2007 new questions were also added so that not just the highest level of BTEC, C and G and RSA qualification, but also any lower levels of such qualifications were recorded . In doing so the variable name for highest BTEC changed from BTEC to BTEC7 and that in itself required a change to the highest qualification derivation syntax.

No account of 2007 could be complete without mentioning apprenticeships. In Q1 2007 one part of QUALCH5 (the initial questions to get respondents thinking about qualifications) was changed from ...Do you have any qualifications....from a Modern Apprenticeship” to “Do you have any qualifications from an apprenticeship” (Presumably either to reflect the dropping of the “modern” prefix or to help pick up quals from trade apprenticeships or possibly both). However,

at the same time a soft check was introduced at appr4 so that if someone had said they did not have a qualification from an apprenticeship at QUALCH5 but then said they had an apprenticeship at appr4 this was queried. There was also a soft check for the reverse (ie reported qualification from apprenticeship at QUALCH5 but did not report having completed an apprenticeship at APPR4. The net effect of these two checks was a substantial fall in reporting of apprenticeship attainment in Q1 07 with a follow on smaller fall in Q2 (from DBFS in Q1 2007 coming back into the respondent sample). Unfortunately those of us on the good ship FESSA at the time were not aware of this soft check issue and it was not uncovered until late 2007. The soft checks were apparently eventually removed in Q2 2008 although the data does not show the expected change until Q3 2008.

In our Q4 2007 SFR we estimated that the effect of the soft checks had been roughly 0.25 pts off both the overall L3+ and L2+ rates for the 19-59/64 population and 0.2 ppts off both rates for the economically active 18-59/64 population.

2008

The apprenticeship question was amended from APPR4 “Are you doing or have you completed a recognised apprenticeship?” to APPR8 “Are you doing or have you completed a recognised apprenticeship including trade, advanced and foundation modern apprenticeships”. Not aware that this resulted in any great change in response. As mentioned above the “apprenticeship soft checks” were removed in Q2 2008 but without any apparent change in data until Q3 2008. We have not yet assessed the size of the change (eg whether the 0.25 ppts has been reversed) and to some extent the impact may be hard to gauge given the change in wording in Q1 2008.

In Q1 2008 the BTEC question was extended to say “Is your highest BTEC/TEC/EDEXCEL/LQL qualification.....” instead of “Is your highest BTEC qualification...”. This resulted in a change in name of the variable from BTEC7 to BTEC8 which necessitated a change in the highest qualification syntax.

Foundation Welsh Baccalaureates were also added in Q1 2008. The addition of an extra category in the hiqual variable led to the renaming from hiqual5 to hiqual8 which obviously needed reflecting in the syntax and also required change to the template (given the extra row).

ANNEX B: Details of LFS qualifications question and classification

1. LFS respondents are initially asked to identify all the qualifications they hold, by type, via the following question.

QUALS6

Which qualifications do (you think) you have, starting with the highest qualifications?

- 1 degree level qualification including foundation degrees, graduate membership of a professional institute, PGCE, or higher
- 2 diploma in higher education
- 3 HNC/HND
- 4 ONC/OND
- 5 BTEC/ BEC/TEC/EdExcel/LQL
- 6 SCOTVEC, SCOTEC or SCOTBEC
- 7 teaching qualification (excluding PGCE)
- 8 nursing or other medical qualification not yet mentioned
- 9 other higher education qualification below degree level
- 10 A-level/ Vocational A-level or equivalent
- 11 DO NOT USE
- 12 Welsh Baccalaureate
- 13 International Baccalaureate
- 14 NVQ/SVQ
- 15 GNVQ/GSVQ
- 16 AS-level/ Vocational AS level or equivalent
- 17 certificate of sixth year studies (CSYS) or equivalent
- 18 Access to HE
- 19 O-level or equivalent
- 20 Standard/Ordinary (O) Grade / Lower (Scotland)
- 21 GCSE/ Vocational GCSE
- 22 CSE
- 23 Advanced Higher / Higher / Intermediate / Access qualifs. (Scotland)
- 24 RSA/ OCR
- 25 City & Guilds
- 26 YT Certificate
- 27 Key Skills / Core Skills (Scotland)
- 28 Basic Skills (Skills for life / literacy / numeracy / language)
- 29 Entry Level Qualifications
- 30 any other professional / work related qualification / foreign qualifications

APPLIES IF RESPONDENT HAS QUALIFICATIONS

i.e.

IF QUALCH5=1 or 2 or 3 or 4 or 5 or 7

Qualifications from school, college, university, modern apprenticeship, connected with work, government schemes, being educated at home, don't know or question not answered

2. From this information, the LFS processing derives the variable HIQUAL which assigns people to the highest qualification they hold in the hierarchy. There are currently over 40 response codes for HIQUAL.
3. In 1992 a task force drew up the National Education and Training Targets. In order to monitor these targets, data from the LFS had to be converted into estimates of the number of people holding qualifications at each NQF level. Experts within the Employment Department and outside were consulted to come up with the mappings. In some cases the whole grouping was allocated to a level, e.g. holders of degrees were all classified to level 4. In other cases the qualifications related to more than one level, e.g. City and Guilds. Assumptions based on available evidence were made to estimate the number of people holding qualifications at various levels within that group (for example, of people holding City and Guilds qualifications, 40% were

assumed to be at level 1, 40% at level 2, and 20% at level 3 or above). There were also major methodological changes to the data collection and questionnaire structure in 1992 and 1996, which mean that time series analyses will have inconsistencies at these dates.

4. A full overview of the current correct assignment of individuals to the level of highest qualification held is set out in Table B1 overleaf.

Table B1: NQF Level equivalences (2008)

	Level 7-8	Level 4-6	Level 3	Level 2	Below level 2
HIQUAL8					
Higher degree	Yes				
NVQ level 5	Yes				
First degree		Yes			
Other degree		Yes			
NVQ level 4		Yes			
Diploma in higher education		Yes			
HNC, HND, BTEC higher		Yes			
Teaching, further education		Yes			
Teaching, secondary education		Yes			
Teaching, primary education		Yes			
Teaching foundation stage		Yes			
Teaching, level not stated		Yes			
Nursing qualifications		Yes			
RSA/OCR higher diploma		Yes			
Other HE below degree		Yes			
NVQ level 3			Yes		
Advanced Welsh Baccalaureate			Yes		
International Baccalaureate			Yes		
GNVQ advanced			Yes		
A level or equivalent			those with 2+	those with 1	
RSA advanced diploma			Yes		
OND, ONC, BTEC national			Yes		
City & Guilds advanced craft			Yes		
Scottish CSYS			67%	33%	
SCE higher or equivalent			those with 3+	those with 1 or 2	
Access to HE qualification			Yes		
AS level or equivalent			those with 4+	those with 2 or 3	those with 1
Trade apprenticeship			50%	50%	
NVQ level 2				Yes	
Intermediate Welsh Baccalaureate				Yes	
GNVQ intermediate				Yes	
RSA/OCR diploma				Yes	
City & Guilds craft				Yes	
BTEC, SCOTVEC first general diploma				Yes	
O level, GCSE grade A*-C or equivalent				those with 5+	those with <5
NVQ level 1					Yes
Foundation Welsh Baccalaureate					Yes
GNVQ, GSVQ foundation level					Yes
CSE below grade 1, GCSE below grade C					Yes
BTEC, SCOTVEC first / general certificate					Yes
SCOTVEC modules					Yes
RSA/OCR other					Yes
City & Guilds other					Yes
YT, YTP certificate					Yes
Key Skills					Yes
Basic Skills					Yes
Entry Level					Yes
Other Qualifications			10%	35%	55%
Don't Know	Pro rata	Pro rata	Pro rata	Pro rata	Pro rata
No Qualifications					

ANNEX C: Details of new methodology

Creation of amended micro-data set

1. The first step of applying the new methodology requires production of an amended micro-data set. The Decision Rules for creating this are set out below.
2. The following definitions are used here:
 - Address is used to denote an address contained in the sampling frame of LFS
 - Household is used to denote a household unit resident at the address
 - A case (or individual) is a member of the household
3. In LFS, identifier fields work such that case 101740310101 lives in household 101740310100 in address 740010103. This allows us to identify new cases living in households, and new households living at addresses. Thus if an individual moves from one sampled address to another, they would receive a new set of identifiers and so cannot be identified in standard LFS data.
4. In the vast majority of cases, a single household lives at an address. However, 1.6% of the working age population in Q4 2007 lived at a multi-household address.
5. The example below sets out the sequence of steps necessary to create a “boosted” dataset for Q4 2008.

Decision rules

- 1) Create an aggregated data file of all English LFS responses from Q4 2007 to Q4 2009, identifying the year and quarter of the series by a code e.g. 20084 for Quarter 4 in 2008.
- 2) Choose cases identified as living at an address included in the Q4 2008 sampling frame. This produces a dataset containing **all** responses for **all** individuals who ever lived at an address during its time in the sampling frame and who responded to the LFS at least once. Some individuals will be included 5 times.
- 3) Remove anyone for whom a valid highest qualification level is missing or who was not living at a household which responded in Q4 2008 (but did at some other time during the time the address was in the sample).
- 4) Identify the lowest priority series for each household using the following logic, which gives preference to the household that responded in the reference quarter.

IF (series=20084) Priority =1.
IF (series=20083) Priority =2.
IF (series=20091) Priority =3.
IF (series=20082) Priority =4.

IF (series=20092) Priority =5.
IF (series=20081) Priority =6.
IF (series=20093) Priority =7.
IF (series=20074) Priority =8.
IF (series=20094) Priority =9.

5) From 4, identify the lowest priority series for each address

6) If, from 4 and 5 more than one household is recorded at a single household address, choose the household with the lowest priority value, assign any other addresses as duplicates and removed from the dataset. This is to ensure that only one household is included in the dataset for each single-household address.

7) Recalculate calendar and academic ages. Use age recorded in LFS Q4 2008 if a case responded then. Otherwise use 31st October 2008 as a reference date. Repeat the process for academic age and assume academic age to equal calendar age if the former is missing.

8) Derive each individual's economic status as at the reference quarter. In most cases, a response for Q4 2008 will be available. However, for "boost" cases, a value will have to be imputed. Preference is determined using the logic outlined in step 4.

9) Identify the relevant LFS record to use for each case. This will be the latest-in-time personal response in the dataset, or the latest-in-time proxy response if no personal response is available. By latest we mean prioritising wave 5 responses, no matter which quarter they were given.

10) Reweight to Q408 population control totals. Allocate each individual in the boosted dataset to a cell of a 3 dimensional table consisting of age (5 year bands), gender and government office region. Divide the Q408 total for each cell by the number of cases in the dataset.

11) Make ILR adjustments if a higher level can be found in the previous 4 years worth of ILR data.

12) Remove anyone not in either the working age or economically active populations

Production of SFR estimates

6. Upon creation of revised datasets, the revised method sub-divides adults of working age into three groups, and the data used to produce the final estimates varies by these groups.

1. Young people aged 19 (calendar age) to 21 (academic age) who were born in the UK or who came to the UK at academic age 14 or younger
2. Young people aged 19 (calendar age) to 21 (academic age) who came to the UK after academic age 14
3. Adults over the academic age of 21

Source of data for SFR estimates for Sub-populations

	Group 1	Group 2	Group 3
Level 4+	Boosted LFS (unadjusted)	Boosted LFS (unadjusted)	Boosted LFS (adjusted for proxies and "Other")
Level 3+	Published NPD proportion	Boosted LFS (adjusted for "Other")	Boosted LFS (adjusted for proxies and "Other")
Level 2+	Published NPD proportion	Boosted LFS (adjusted for "Other")	Boosted LFS (adjusted for proxies and "Other")

7. Adjustments to groups 2 and 3 for "Other" and proxies are made on the basis of the imputation modelling described in Annex C of the [final research report](#).

ANNEX D: Details of old methodology

1. The previous method was based on using the LFS variable HIQUAL, rather than LEVQUAL, in order to a) handle the issues now dealt with by the script levqualcorrected.sps, and b) provide a basis to apportion those respondents reporting 'other' qualifications (see paragraph 7 below).
2. The sections below set out the key steps to producing final estimates using the LFS data, with relevant SPSS scripts provided at the end of the section.

Calculating level of highest qualification held by different respondent groups

3. To identify how many people in a given group have 'X' qualification, weighted population estimates (using variable PWT03) of HIQUAL are produced using SPSS. The relevant sub population is selected (i.e. 19 – 59 females) using a series of filters, and a frequency table of HIQUAL on this sub group is produced.
4. Aggregation of the various qualifications from HIQUAL into NQF levels (Levels 1 to 4+) is carried out using a spreadsheet process. An initial step is to handle those respondents with ambiguous responses to HIQUAL. Where individual responses to HIQUAL are coded as -8 (No answer) and -9 (Did not apply), respondents are apportioned on a pro rata basis based on the response codes for those people who did respond, including those reporting 'No qualifications'. 'Don't Know' responses are apportioned on a pro rata basis based on those who did respond, but excluding those with 'No qualifications'. This results in revised totals for HIQUAL.
5. In most instances there is a direct relationship from the highest qualification noted in HIQUAL to NQF level. For example a 'first degree' is a level 4+. In some cases additional information is used to assign a correct level to that qualification. This is based on either a) use of data from other LFS questions that follow the initial qualifications question (QUALCH6 see annex) or b) assumptions based on other research evidence.
6. Other LFS variables are used for details such as the number of GCSEs; a full list of derivations is set out below.

Qualification type	Variable used	Number held	Level
A levels / Scottish Advanced Highers	NUMAL / ADVHST	2 or more 1	Level 3 Level 2
Scottish Highers	NUMSCE / HST	3 or more 1 or 2	Level 3 Level 2
AS levels	NUMAS	4 or more 2 or 3 1	Level 3 Level 2 Below level 2
O levels / GCSEs (A*-C)	NUMOL5	5 or more 1 to 4	Level 2 Below level 2

7. Assumptions applied in other cases in order to assign a correct level are:
- Apprenticeships are assigned to Level 3:Level 2 in the ratio 50:50.
 - Other³ qualifications (those which cannot be coded, as well as all Foreign qualifications) are assigned to Level 3:Level 2:Below level 2 in the ratio 10:35:55 (based on analysis from the General Household Survey).
 - Scottish CSYS are assigned to Level 3:Level 2 in the ratio 67:33.

Calculation of progress against PSA targets and supplementary tables

8. The PSA targets to 'Improve the skills of the population by 2011' refer to people aged 19 – 59 (women) and 19 – 64 (men) and this is termed as 'working age adults'.
9. The PSA target to 'Extend opportunities for all' relates to economically active adults, i.e. those people who are working age adults but who also are working or looking for work.
10. In order to calculate the latest progress against these targets, it is necessary to identify the people who fall in the relevant age brackets before running the analyses to identify the levels of highest qualification held for the group (detailed above). The variable AGE on the LFS is used to define these sub-groups.
11. It is also necessary to create a variable to identify the economically active within this group, i.e. those who are employed, self employed and unemployed. This new variable is created using the existing LFS variable INECAC05.
12. The below SPSS syntax is used to create highest qualification tables for the working age adult population (19-54/64) using quarterly LFS data Q1 2008 onwards.

*******FILTERS AND VARIABLES*****.**

weight by pwt07.

* Workage filter.

Compute workage =0.

if (sex=1&range(age,16,64)) | (sex=2&range(age,16,59)) workage =1.

variable labels workage 'Working age identifier'.

* Workage19 filter.

Compute workage19 =0.

if (sex=1&range(age,19,64)) | (sex=2&range(age,19,59)) workage19 =1.

variable labels workage19 'Working age 19 identifier'.

³ Those people with qualifications that do not fit into the existing pre code list are recorded as having Other qualifications. In addition, all foreign and professional qualifications are recorded as Other. Other responses are only used in the calculation of highest qualification if this is the sole qualification reported.

```

* Adult targets definition.
Compute adult=0.
if workage=1 & age>=18 & range(ilodefr,1,2) adult =1.
variable labels adult 'Adult identifier'.

* Create SFR age bands.
DO IF workage=1.
RECODE
  age
  (16 thru 19=1) (20 thru 24=2) (25 thru 29=3) (30 thru 34=4)
  (35 thru 39=5) (40 thru 44=6) (45 thru 49=7) (50 thru 54=8)
  (55 thru 64=9) (ELSE=999) INTO sfrage .
END IF.
VARIABLE LABELS sfrage 'Age bands for quals SFR'.
ADD VALUE LABELS sfrage 1'16-19' 2'20-24' 3'25-29' 4'30-34' 5'35-39' 6'40-44' 7'45-49' 8'50-54' 9'55-64'.

```

```

* Economic activity.
Compute sfrecon =999.
if inecac05 = 1 sfrecon =1.
if inecac05 = 2 sfrecon=2.
if ilodefr = 2 sfrecon =3.
if ilodefr = 3 sfrecon =4.
VARIABLE LABELS sfrecon 'Economic activity for quals SFR'.
ADD VALUE LABELS sfrecon 1'Employee' 2'Self-employed' 3'ILO Unemployed' 4'Inactive'.

```

```

**Create GO regions for England and Country variables.
Compute ukhome=999.
if range(govtor,1,16) ukhome = 1.
if govtor = 17 ukhome = 2.
if range(govtor,18,19) ukhome = 3.
if govtor = 20 ukhome = 4.
VARIABLE LABELS ukhome 'UK home countries'.
ADD VALUE LABELS ukhome 1'ENG' 2'WAL' 3'SCO' 4'NIR'.
DO IF range(govtor,1,16).
Compute sfrgov=999.
if range(govtor,1,2) sfrgov = 1.
if range(govtor,3,5) sfrgov =2.
if range(govtor,6,8) sfrgov = 3.
if govtor = 9 sfrgov = 4.
if range(govtor,10,11) sfrgov = 5.
if govtor = 12 sfrgov = 6.
if range(govtor,13,14) sfrgov = 7.
if govtor = 15 sfrgov = 8.
if govtor = 16 sfrgov = 9.
END IF.
VARIABLE LABELS sfrgov 'GO regions for SFR'.
VALUE LABELS sfrgov 1'NE' 2'NW' 3'YH' 4'EM' 5'WM' 6'EE' 7'LN' 8'SE' 9'SW'.

```

*******QUALIFICATION CALCULATIONS*******.

```

**Recode NAs and DNAs in hiqua18, numal, numas and numol5 to 999.
variable level hiqua18 (ordinal).
RECODE
  hiqua18

```

(-8 = 999) (ELSE=Copy) INTO hiqual8.

RECODE

numal

(-8=999) (ELSE=Copy) INTO numal.

RECODE

numas

(-8=999) (ELSE=Copy) INTO numas.

RECODE

numol5

(-8=999) (ELSE=Copy) INTO numol5 .

compute numal2=999.

if (numal=3 | advhst=-8 | hst=-8) numal2=3.

if (numal=1 | advhst=1 | hst=2) numal2=1.

if (numal=2 | advhst=2 | hst=1) numal2=2.

variable labels numal2 'Numal incorporating Scottish Advanced Highers and Highers'.

value labels numal2 1 'Level 2' 2 'Level 3' 3 'Dont know'.

*A-level adjustment.

*Spring 2005 onwards.

*Create to variable for adjusted A levels.

COMPUTE numaladj = numal2.

VARIABLE LABELS numaladj 'Adjusted number of A levels - those with 1 A-Level but another L3 are moved'.

VALUE LABELS numaladj

1.0000000000000000 "1 A level or equivalent"

2.0000000000000000 "More than one A-level or equivalent quals"

3.0000000000000000 "Don't Know"

4.0000000000000000 "Other L3"

5.0000000000000000 "CSYS and one A-level"

6.0000000000000000 "Trade Apprenticeship and one A-level"

999.00000000000000 "999".

VARIABLE LEVEL numaladj(ordinal).

*Filter on those holding A-levels as their highest qual.

USE ALL.

COMPUTE filter_\$(hiqual8 = 20).

VARIABLE LABEL filter_\$ 'hiqual8 = 20 (FILTER)'.

VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.

FORMAT filter_\$ (f1.0).

FILTER BY filter_\$.

*Recode numal for those with A-levels who also hold other quals at L3 or higher.

DO IF (rsa=2 | ANY(quals601,4,18) | ANY(quals602,4,18) | ANY(quals603,4,18) |

ANY(quals604,4,18) | ANY(quals605,4,18) |

ANY(quals606,4,18) | ANY(quals607,4,18) | ANY(quals608,4,18) | ANY(quals609,4,18) |

ANY(quals610,4,18) | ANY(quals611,4,18)

| candg=1 | btec8=2 | sctvec=2 | numas=3) .

RECODE

numaladj (1=4) (2=2) (3=4) (4=4) (5=5) (6=6) (SYSMIS=SYSMIS) .

END IF .
FILTER OFF.

*Filter on all those holding one-A-level (in the adjusted variable) as their highest qual.
USE ALL.
COMPUTE filter_\$(hiqual8 = 20 & numaladj = 1 | hiqual8 =20 & numaladj = 3).
VARIABLE LABEL filter_\$ 'hiqual8 =20 & numaladj = 1 | hiqual8=20 & numaladj=3 (FILTER)'.
VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_\$ (f1.0).
FILTER BY filter_\$.

*Recode numal for those with 1 A-levels (adjusted) and a completed CSYS.
DO IF (quals601=17 | quals602=17 | quals603=17 | quals604=17 | quals605=17 | quals606=17
| quals607=17 | quals608=17 | quals609=17 | quals610=17 | quals611=17).

RECODE
numaladj (1=5) (2=2) (3=5) (4=4) (5=5) (6=6) (SYSMIS=SYSMIS) .
END IF .
FILTER OFF.

*Filter on all those holding one-A-level (in the adjusted variable) as their highest qual.
USE ALL.
COMPUTE filter_\$(hiqual8 =20 & numaladj = 1 | hiqual8=20 & numaladj = 3).
VARIABLE LABEL filter_\$ 'hiqual8 =20 & numaladj = 1 | hiqual8 = 20 & numaladj = 3
(FILTER)'.
VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_\$ (f1.0).
FILTER BY filter_\$.

*Recode numal for those with 1 A-levels (adjusted) and a completed Trade Apprenticeship.
DO IF appr8=1 | appr8=3.
RECODE
numaladj (1=6) (2=2) (3=6) (4=4) (5=5) (6=6) (SYSMIS=SYSMIS) .
END IF .
FILTER OFF.

*Filter on those with SCE highers as their highest qual.
USE ALL.
COMPUTE filter_\$(hiqual8 = 25).
VARIABLE LABEL filter_\$ 'hiqual8 = 25 (FILTER)'.
VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_\$ (f1.0).
FILTER BY filter_\$.

*Recode numaladj for those with Highers who also hold other quals at L3 or higher.
DO IF (quals601=18 | quals602=18 | quals603=18 | quals604=18 | quals605=18 |
quals606=18 | quals607=18 | quals608=18 | quals609=18 | quals610=18 | quals611=18 |
numas=3) .
RECODE
numaladj (1=4) (2=2) (3=4) (4=4) (5=5) (6=6)(SYSMIS=SYSMIS) .
END IF .
FILTER OFF.

*Filter on those holding 1 or 2 Scottish Highers (adjusted) and a Trade Apprenticeship.

USE ALL.

COMPUTE filter_\$(=(hiqua18 =25 & numal2 = 1) | (hiqua18 =25 & numal2 = 3)).

VARIABLE LABEL filter_\$_ '\$=(hiqua18 =25 & numal2 = 1) | (hiqua18 =25 & numal2 = 3)
(FILTER)'.

VALUE LABELS filter_\$_ 0 'Not Selected' 1 'Selected'.

FORMAT filter_\$_ (f1.0).

FILTER BY filter_\$_.

*Recode numaladj for those with 1 or 2 Highers who also hold a Trade Apprenticeship.

DO IF appr8=1 | appr8=3.

RECODE

numaladj (1=6) (2=2) (3=6) (4=4) (5=5) (6=6) (SYSMIS=SYSMIS) .

END IF .

FILTER OFF.

**AS adjustment.

*Spring 2005 onwards.

*Create new variable for adjusted AS levels.

COMPUTE numasadj = numas .

VARIABLE LABELS numasadj 'Adjusted number of AS levels - those with 1 AS & Other L2 are
moved'.

VALUE LABELS numasadj

1.0000000000000000 "1 AS level"

2.0000000000000000 "2 or 3 AS levels or equivalent quals"

3.0000000000000000 "4 or more AS levels"

4.0000000000000000 "Don't Know"

5.0000000000000000 "Other L2"

6.0000000000000000 "Trade apprenticeship"

999.00000000000000 "999".

VARIABLE LEVEL numasadj(ordinal).

*Filter on those AS level as their highest qual.

USE ALL.

COMPUTE filter_\$(=(hiqua18 = 27 & any(numas,1,2,4)).

VARIABLE LABEL filter_\$_ 'hiqua18 = 27 & any(numas,1,2,4) (FILTER)'.

VALUE LABELS filter_\$_ 0 'Not Selected' 1 'Selected'.

FORMAT filter_\$_ (f1.0).

FILTER BY filter_\$_.

*Recode numas for those with 1 AS level who also hold other quals at L2 or higher.

DO IF (ANY(appr8,1,3) | nvql5=2 | RANGE(gnvq4,2,3) | rsa=3 | candg=2 | btec8=3 | sctvec=3
| numol5 = 2).

RECODE

numasadj (1=5) (2=5) (3=3) (4=5) (5=5) (6=6) (SYSMIS=SYSMIS) .

END IF .

FILTER OFF.

*Filter on those holding 1 or 2 to 3 AS levels as their highest qual.

USE ALL.

COMPUTE filter_\$(=(hiqua18 = 27 & any(numas,1,2,4)).

VARIABLE LABEL filter_\$_ 'hiqua18 = 27 & any(numas,1,2,4) (FILTER)'.

VALUE LABELS filter_\$_ 0 'Not Selected' 1 'Selected'.

FORMAT filter_\$_ (f1.0).

FILTER BY filter_\$_.

```
*Recode numas for those with AS levels who also hold a Trade Apprenticeship.
DO IF appr8=1 | appr8=3.
RECODE
  numasadj (1=6) (2=6) (3=3) (4=5) (5=5) (6=6) (SYSMIS=SYSMIS) .
END IF .
FILTER OFF.
```

*Numol 5 - Including the 999s in the value labels.

```
VALUE LABELS numol5
1.00000000000000 "Fewer than five"
2.00000000000000 "Five or more"
3.00000000000000 "Don't Know"
999.000000000000 "999"
```

```
VARIABLE LEVEL numol5(ordinal).
```

*****FINAL TABLES*****.

**Highest qualification tables.

temporary.

select if workage19=1.

* Custom Tables.

CTABLES

```
/VLABELS VARIABLES=sex ukhome sfrgov hiqua18 DISPLAY=DEFAULT
/TABLE hiqua18 [COUNT F40.0] BY sex > (ukhome [C] + sfrgov [C])
/CATEGORIES VARIABLES=sex hiqua18 ORDER=A KEY=VALUE EMPTY=INCLUDE
TOTAL=YES POSITION=BEFORE MISSING=EXCLUDE
/CATEGORIES VARIABLES=ukhome [2, 3, 4, 1, OTHERNM] EMPTY=INCLUDE
TOTAL=YES LABEL='UK' POSITION=BEFORE
/CATEGORIES VARIABLES=sfrgov ORDER=A KEY=VALUE EMPTY=INCLUDE.
```

**Highest qual = A levels.

temporary.

select if (hiqua18 = 20 | hiqua18=25) & workage19=1.

* Custom Tables.

CTABLES

```
/VLABELS VARIABLES=sex ukhome sfrgov numaladj DISPLAY=DEFAULT
/TABLE numaladj BY sex > (ukhome [C][COUNT F40.0] + sfrgov [C][COUNT F40.0])
/CATEGORIES VARIABLES=sex ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=YES
POSITION=BEFORE MISSING=EXCLUDE
/CATEGORIES VARIABLES=ukhome [2, 3, 4, 1, OTHERNM] EMPTY=INCLUDE
TOTAL=YES LABEL='UK' POSITION=BEFORE
/CATEGORIES VARIABLES=sfrgov ORDER=A KEY=VALUE EMPTY=INCLUDE
/CATEGORIES VARIABLES=numaladj ORDER=A KEY=VALUE EMPTY=INCLUDE
TOTAL=YES POSITION=BEFORE MISSING=INCLUDE.
```

*Highest qual = AS levels.

temporary.

select if hiqua18 = 27 & workage19=1.

* Custom Tables.

CTABLES

```
/VLABELS VARIABLES=sex ukhome sfrgov numasadj DISPLAY=DEFAULT
```

```

/TABLE numasadj BY sex > (ukhome [C][COUNT F40.0] + sfrgov [C][COUNT F40.0])
/CATEGORIES VARIABLES=sex ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=YES
POSITION=BEFORE MISSING=EXCLUDE
/CATEGORIES VARIABLES=ukhome [2, 3, 4, 1, OTHERNM] EMPTY=INCLUDE
TOTAL=YES LABEL='UK' POSITION=BEFORE
/CATEGORIES VARIABLES=sfrgov ORDER=A KEY=VALUE EMPTY=INCLUDE
/CATEGORIES VARIABLES=numasadj ORDER=A KEY=VALUE EMPTY=INCLUDE
TOTAL=YES POSITION=BEFORE MISSING=INCLUDE.

```

**Highest qual = GCSEs.

temporary.

select if hiqua18 = 35 & workage19=1.

* Custom Tables.

CTABLES

```

/VLABELS VARIABLES=sex ukhome sfrgov numol5 DISPLAY=DEFAULT
/TABLE numol5 BY sex > (ukhome [C][COUNT F40.0] + sfrgov [C][COUNT F40.0])
/CATEGORIES VARIABLES=sex ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=YES
POSITION=BEFORE MISSING=EXCLUDE
/CATEGORIES VARIABLES=ukhome [2, 3, 4, 1, OTHERNM] EMPTY=INCLUDE
TOTAL=YES LABEL='UK' POSITION=BEFORE
/CATEGORIES VARIABLES=sfrgov ORDER=A KEY=VALUE EMPTY=INCLUDE
/CATEGORIES VARIABLES=numol5 ORDER=A KEY=VALUE EMPTY=INCLUDE
TOTAL=YES POSITION=BEFORE MISSING=INCLUDE.

```

weight off.

compute freq=0.

if range(hiqua18,1,49) freq=1.

value labels freq 1 'Unweighted base for percentages' 0 'Ignore, no info on hiqua1'.

variable level freq (ordinal).

temporary.

select if workage19=1.

* Custom Tables.

CTABLES

```

/VLABELS VARIABLES=sex ukhome sfrgov freq DISPLAY=DEFAULT
/TABLE freq BY sex > (ukhome [C][COUNT F40.0] + sfrgov [C][COUNT F40.0])
/CATEGORIES VARIABLES=sex ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=YES
POSITION=BEFORE MISSING=EXCLUDE
/CATEGORIES VARIABLES=ukhome [2, 3, 4, 1, OTHERNM] EMPTY=INCLUDE
TOTAL=YES LABEL='UK' POSITION=BEFORE
/CATEGORIES VARIABLES=sfrgov ORDER=A KEY=VALUE EMPTY=INCLUDE
/CATEGORIES VARIABLES=freq ORDER=A KEY=VALUE EMPTY=INCLUDE TOTAL=no
MISSING=INCLUDE.

```