



Department
for Education

Continuing vocational training survey: CVTS 5

**Technical report
October 2017**

June Wiseman and Emma Parry, BMG Research

Contents

Introduction	3
Background to the research	3
Objectives	3
Methodology	3
Standardisation	3
The universe under investigation	3
Sample completed	5
Questionnaire development	5
Sampling process	6
Response rates	6
Editing and control	8
Treatment of non-response	8
Impact of imputation and estimation	11
Standard error and confidence intervals	11
Note on rounding	11
Weighting process	12
Survey scope/definitions	12
Defining Continuing Vocational Training	12
Quality control	14
Database	14
Measurement errors: Questionnaire design	15
Interviewer briefings	16
Data processing	17
Appendix I: Population	19
Sample:	19
Appendix II: Sample	22
Appendix IV: Sampling Methodology	27
Appendix V: Weighting Factors	30
Appendix VI: Stratum Switchers	31
Appendix VII Call Outcome	32

Introduction

Background to the research

The fifth Continuing Vocational Training Survey (CVTS5) has been undertaken in the UK by BMG Research on behalf of the Department for Education as part of a pan-European research exercise to determine the level of CVT within private and voluntary sector organisations.

The survey has been undertaken every five years since 1995, with The Statistical Office of the European Community (Eurostat) co-ordinating it across 28 European countries (the (now) EU 27 plus Norway).

Objectives

The overarching aim of CVT5 is to conduct a survey of employers to explore the nature and extent of the vocational training that they provide.

Important research issues for which CVTS data are needed include the organisation and management of CVT in enterprises, the role of social partners, assessment of skill/training needs, volume of CVT and possible interaction with IVT, incentives for enterprises to provide CVT, costs and financing of CVT in enterprises, obstacles for enterprises in providing CVT, and the costs and financing of CVT in enterprises.

The survey has been carried out so that it conforms to specifications laid down by Eurostat, ensuring data from the UK is comparable with that of other participating countries.

Methodology

Standardisation

In order to preserve comparability across participating countries, the survey employed the European standard questionnaire, with minor additions and wording changes to maximise its efficacy in the UK. In addition the survey used a standard list of classifications across countries for defining types of training and the scope of the survey was standardised using the NACE classification, common sampling units i.e. enterprises and specified size bands, as well as time reference periods.

The universe under investigation

For the purposes of the research, the business population was defined as follows:

- Enterprise-based;
- Excluding businesses with fewer than 10 employees.

The population of interest for use by Eurostat in reporting comparisons between all participating countries belong to 20 NACE Rev. 2 categories including B, C10-C12, C13-C15, C17-C18, C19-C23, C24-C25, C26-C28 and C33, C29-C30, C16+C31-32, D-E, F, G(45), G(46), G(47), I, H, J, K(64, 65), K(66), L+M+N+R+S.

Other sectors were excluded from the survey. These other sectors included NACE Rev. 2 categories A, O, P and Q¹. These were included in previous surveys as additional sample but are optional. Data relating to these sectors is not of interest to Eurostat.

NACE	SIC 2007	NACE/SIC description²
B05-B09	5-9	Mining and quarrying and support activities
C10-C12	10-12	Manufacture of food products, beverages and tobacco
C13-C15	13-15	Manufacture of textiles and textile products Manufacture of leather and leather products
C17-C18	17-18	Manufacture of pulp, paper and paper products, Printing of newspapers
C19-C23	19-23	Manufacture of coke oven products Manufacture of flat glass
C24-C25	24-25	Manufacture of basic iron and steel and of ferro-alloys Manufacture of metal structures and parts of structures
C26-C28 and C33	26-28, 33	Manufacture of electronic components, Repair of fabricated metal products Manufacture of engines and turbines, except aircraft, vehicle and cycle engines
C29-C30	29-30	Manufacture of motor vehicles, Building of ships and floating structures
C16+C31-C32	16, 31, 32	Sawmilling and planing of wood, Manufacture of office and shop furniture, Striking of coins
D-E	35-39	Electricity, gas, steam and air conditioning supply, Water supply; sewerage, waste management and remediation activities
F	41-43	Construction

1

A 01-03 Agriculture, forestry, fishing
O 84 Public administration
P 85 Education
Q 86-88 Human health and social work

² "The UK SIC is based exactly on NACE but, where it was thought necessary or helpful, a fifth digit has been added to form subclasses of the NACE four digit classes." UK Standard Industrial Classification of Economic Activities 2007 (SIC 2007): Structure and Explanatory Notes. Office for National Statistics, December 2009

NACE	SIC 2007	NACE/SIC description²
G45	45	Sale of cars and light motor vehicles
G46	46	Agents involved in the sale of agricultural raw materials, live animals, textile raw materials and semi-finished goods
G47	47	Retail trade in non-specialised stores with food, beverages or tobacco predominating
H	49-53	Transportation and storage
I	55-56	Accommodation and food service activities
J	58-63	Information and communication
K64-K65	64-65	Financial and insurance activities, Life insurance
K66	66	Administration of financial markets
L + M + N + R + S	68-82, 90-96	Real estate, renting and business activities, Professional, scientific and technical activities, Administrative and support service activities, Arts, entertainment and recreation, Other service activities

Sample completed

In total, 3,315 interviews were conducted.

Questionnaire development

The European standard questionnaire was used as the basis for the survey. The questionnaire employed, with annotations for additional questions, codes and wording changes compared with previous surveys, is included in Appendix VII.

In order to maximise the opportunity to obtain key data regarding the total number of hours worked by employees and the total labour costs within an organisation, questions were added in order to calculate these statistics. This was in line with Eurostat's guidelines.

Also, in order to obtain detailed numeric information from respondents who were not able to provide this information without consulting other sources, a data sheet was assembled which was emailed or faxed to respondents for completion prior to an arranged interview appointment. This enabled respondents to collate information in readiness to relay that information to a BMG interviewer. The data sheet employed is included in Appendix VIII.

Prior to the process of fully piloting the questionnaire, 5 cognitive interviews were undertaken to explore respondents' understanding of the terminology and phrases used in questioning and issues around the ease of supplying the data requested. In addition, respondents were asked to give feedback on the data sheet. Following the cognitive pilot, minor wording and coding changes were made to the questionnaire and the data

sheet was redesigned in a simpler format with minimal text. The report on the cognitive interviews is available separately.

Average interview lengths varied depending on whether training was provided for employees or not. Those that did not provide training had a far shorter interview of around 7 minutes. Those that did provide training recorded an average interview length of about 28 minutes.

Sampling process

To inform the distribution of the sample by size and sector, the latest (November 2015) data from the Inter-Departmental Business Register (IDBR) covering the United Kingdom and the distribution of businesses was used and is included in Appendix I.

The sample structure was calculated to specifications designed by the Statistical Office of the European Community (Eurostat) to ensure consistency of approach across all participating countries.

This structure ensured that the sample achieved is reflective of the known population of enterprises at a level of detail defined by the 20 NACE groups and across 6 business size bands (in previous CVT surveys only 3 size bands were used).

The sample definition was based on the population data, to which a formula was applied which takes into account estimated response rates and propensity to fund or arrange training within a size and sector cell. This resulting 'target' within each size and sector cell was the number of contacts that were to be issued in order to achieve an expected number of interviews, based on a net 40% response rate (gross 50%, including 'deadweight'). The number of achieved interviews depended on the actual response rate, which did not always conform to the anticipated response rate.

The number of contacts issued, the number of interviews achieved and the response rate by sector and size with regard to the core sample are included in Appendix II. This information pertaining to the additional sample is included in Appendix III.

The CVTS5 Manual provided detailed instructions in determining the sample structure and the numbers to work to. This is included in Appendix IV.

Response rates

Whilst in calculating the likely sample that will be achieved from a set of contacts, an overall estimated response rate is assumed, in reality the response rate varies between NACE/size cells. The number of achieved interviews against the targets set was monitored by NACE/size and steps taken to maximise response rates. These steps involved calling at varying times and on varying days of the week; evening and weekend interviewing was carried out where necessary; appointments were made with potential

respondents; and contacts were called at least 10 times before being discarded as a non-response.

A response rate of 25% was achieved overall. This response rate is not adjusted to take into account cases where repeated calls were made but interviewers were unable to speak to someone to ask them to take part in the survey. When these calls are taken into account and only the 'contacted' sample is used to calculate the figure, the response rate was 37% overall.

In maximising response rates a number of strategies were employed. Calls were made at varying times and on varying days of the week; evening and weekend interviewing was carried out where necessary; appointments were made with potential respondents; and contacts were called at least 10 times before being discarded as a non-response.

The option of completing the survey online was also offered. This was set up to allow respondents to complete the survey in their own time and at their own pace, taking breaks and going back into the survey when they needed to search for data or wished to pause for another reason.

BMG Research employs a fully trained team of interviewers, most of whom are permanently employed on a full time basis and have extensive experience of labour market surveys amongst employers. They are experienced in persuading people to participate and were provided with a survey introduction which stressed the long term significance of the project and the importance of employers' contributions to informing developments in training policy.

BMG Research also provided a helpline service for the purpose of the survey, using a dedicated and experienced team who were fully briefed to handle respondents' questions.

All outcomes were logged to industry standard codes i.e. completed; refused; no reply; answer machine; ring backs but unsuccessful; ceased trading; wrong number. The breakdown of final call outcomes for each contact is provided in Appendix VI.

The achieved response rates are also to be found in Appendix VI.

Editing and control

The use of a Computer Assisted Telephone Interviewing (CATI) system to collect data allows for the incorporation of logic checks into the script. For each numeric variable collected, an acceptable value range was applied. This range was based on the number of employees across an organisation (all of the variables being directly affected by this factor) and the range of responses given in CVTS4 (modified to reflect expected increases over time). When responses fell outside of this range, the interviewer was prompted to double-check the response before moving forward to the next question. In some cases, this resulted in a revised response, in others the response would not be changed as the respondent had confirmed it to be correct.

Following data collection, data was checked to ensure that values given were valid and credible and that there was consistency between different variables. These checks were undertaken on a case by case basis. Where values were considered to be incorrect or inconsistent with other variables, these cases were flagged for call-backs to make another attempt to obtain the correct figure.

Treatment of non-response

There are two types of non-response:

- Unit non-response, where no survey data are collected for a unit.
- Item non-response, where some data are collected for a unit but some values are missing.

The sampling process operates on the basis of a predetermined number of contacts being issued and those contacts being called repeatedly until they either complete an interview or opt out of the survey (i.e. refuse to take part). There will, of course be other call outcomes, most notably, ring-backs because there is no one with the necessary knowledge to respond to the survey at that time.

The impact of unit non-response is to reduce the sample size overall and within NACE and size categories. The extent to which responses are obtained varies by NACE and size and this leads to certain NACE and size categories being over- or under-represented in the data. It is therefore necessary to apply weighting factors to the data to ensure that NACE and size categories have a weight in the data that is equal to their weight had a representative proportion responded to the survey.

Item non-response is dealt with through the process of imputation. Where there is still a missing value after direct methods (i.e. call-backs) have been attempted, a value is imputed according to rules laid down in the CVTS5 Manual as follows:

Core variables, for which no missing value was accepted or imputation permitted include:

A1	Actual NACE-code of the enterprise
A2tot	Total number of persons employed at end of 2010
B1a	Provision of internal CVT courses
B1b	Provision of external CVT courses
B2aflag, B2bflag, B2cflag, B2dflag, B2eflag	Provision of 'other' forms of CVT
F1tot	IVT participants usually employed in the enterprise

Key variables, for which every effort to avoid missing values should be made, but imputation was permitted include:

A4	Total number of hours worked in 2015 by persons employed
A5	Total labour costs (direct and indirect) of all persons employed in 2015
C1tot	Total CVT course participants
C3tot	Paid working time (in hours) spent on all CVT courses
C7sub	CVT costs sub-total
C7tot	Total costs CVT
PAC	Personal absence costs

In addition, the following variables were identified as being appropriate for imputation, some of which feed into derived key variables, such as C7tot and PAC.

A2m	Total number of males employed at end of 2010
A2f	Total number of females employed at end of 2010
B2a	Total number of participants in planned training through guided on-the-job training
B2b	Total number of participants in planned training through job-rotation, exchanges, secondments and study visits
B2c	Total number of participants in planned training through attendance at conferences, workshops, trade fairs and lectures
B2d	Total number of participants in planned training through participation in learning or quality circles
B2e	Total number of participants in planned training by self-directed learning
B5a	Amount in £ sterling contributed to collective or other funds for Vocational Training activities
B5b	Amount in £ sterling received in financial grants or subsidies for Vocational Training activities
C2m	Total number of male participants in CVT courses

C2f	Total number of female participants in CVT courses
C3i	Total number of paid working hours spent on internally managed Vocational Training courses
C3e	Total number of paid working hours spent on externally managed Vocational Training courses
C7a	Cost of fees and payments for Vocational Training courses
C7b	Cost of travel and subsistence payments for Vocational Training courses
C7c	Cost of labour for internal trainers for Vocational Training courses
C7d	Cost of training centre, training premises or specific training rooms

The following rules for imputation to address item non-response were set and adhered to:

- When a record contained less than 50% of variables presented then this record was considered to be a unit non-response.
- For a single NACE/size cell, imputations were not allowed if more than 50% of the responding enterprises had missing data for more than 25% of quantitative variables.
- For a single NACE/size cell imputations were not allowed on a quantitative variable if the proportion of responding enterprises for that particular variable were less than 50%.
- For a single NACE/size cell imputations were not allowed on a qualitative variable if the proportion of responding enterprises for that particular variable is less than 80%.

Imputations were carried out on the following variables (which are referenced in the questionnaire included in Appendix VII):

A2m, A2f, A3tot, A4, A5, B2a, B2b, B2c, B2d, B2e, B5a, b5b, C1tot, C2m, C2f, C3tot, C3i, C3e, C4, C7a, C7b, C7c, C7d, C7sub, C7tot, PAC

Firstly, analysis by NACE and size was undertaken to determine where imputation was possible for each variable within each NACE and size cell. The larger the organisations the less likely imputation was allowed due to the small size of the samples within the cells.

The methods used for imputation were as follows:

- Where an absolute was not provided but a banded response was, the mid point of the banded answer was used.
- Where there was no valid banding given, the mean of the valid responses given for the relevant NACE/size cell was used for the missing value.

Impact of imputation and estimation

It should be noted that the process of imputation and the fact that some respondents could only estimate some characteristics of their training (such as its costs or the amount of time which employees spent in training) introduces a margin of error into the data (over and above normal sampling error). Some estimates in the survey should, therefore, not be read as having pinpoint accuracy but as general indications of employer behaviour.

Standard error and confidence intervals

In an ideal world when views are sought, everyone would be asked. This would involve a census. It is an expensive approach and time-consuming and impractical, as it is very difficult to get hold of everyone in a target population. Consulting a sample of a target population is more cost-effective and achievable but does introduce a level of standard error, where the statistics gathered from a sample of the target population deviate from those that would be gathered from a census.

Standard error is calculated on the basis of two different elements; the sample size and the statistic itself. The larger the sample, the smaller the size of the standard error. The maximum standard error for a given sample is based on a statistic of 50%. The standard error is usually calculated to a confidence level of 95% (i.e. we can be confident that 95% of responses would fall within a given range of responses). Based on a reported statistic of 50%, the overall sample of 3,315 for this survey is subject to a standard sampling error of +/-1.7%. Thus, if all businesses within the population were asked, we would be 95% confident that the reported statistic would fall within a range of 48.3% to 51.7%.

A table which presents the level of standard error for a range of statistics, based on the total sample and key sub-samples, is included in Appendix X.

Note on rounding

Throughout this report percentages have been rounded to the nearest whole integer. Furthermore average cost and cost per hour figures have been rounded to the nearest 10. The exceptions to rounding concern hourly costs and values for hours spent in training. These have been reported to two decimal points (in the case of figures that are based on £ sterling) and to one decimal point (in the case of figures that relate to parts of hours). This enables the reporting of small differences between sample groups.

Weighting process

As already mentioned, the data was weighted to ensure that unit non-response was compensated for in NACE/size cells. Population estimates based on business enterprises from IDBR were used to provide the weighting factors. The weighting factors applied are presented in Appendix V.

Survey scope/definitions

Defining Continuing Vocational Training

A key issue for the survey is that it distinguishes CVT from other forms of training with which the survey is not, or is less, concerned. These other forms primarily comprise Initial Vocational Training (IVT) and what is usually known as 'on-the-job' or informal training.

Continuing vocational training (CVT)

Continuing vocational training is a training measure or activity which has as its primary objective, the acquisition of new competencies or the development and improvement of existing ones and which is financed at least partly by the enterprises for their persons employed who either have a working contract or who benefit directly from their work for the enterprise such as unpaid family workers and casual workers. Persons employed holding an apprenticeship or training contract must not be taken into consideration for CVT. (These could be relevant candidates for IVT). The training measures or activities must be planned in advance and must be organised or supported with the specific goal of learning. Random learning and initial vocational training (IVT) are explicitly excluded.

Initial vocational training (IVT)

Initial vocational training (IVT) is a work-based training measure or activity for apprentices/trainees. It leads to a formal qualification. The measures are often financed (partly or wholly) by the enterprise although this is not a mandatory condition. Apprentices/trainees often have a special training contract.

Continuing vocational training measures and activities

Continuing vocational training measures and activities includes CVT courses and "Other" forms of CVT.

CVT courses

CVT courses are typically clearly separated from the active work place (learning takes place in locations specially assigned for learning, a class room or training centre). They exhibit a high degree of organisation (time, space and content) by a trainer or a training institution. The content is designed for a group of learners (e.g. a curriculum exists). Two distinct types of CVT courses are identified:

- internal CVT courses
- external CVT courses

‘Other’ forms of CVT

“Other” forms of continuing vocational training are typically connected to the active work and the active work place, but they can also include attendances at conferences, trade fairs etc. for the purpose of learning. They are often characterised by a degree of self-organisation (time, space and content) by the individual learner or by a group of learners. The content is often tailored according to the learners’ individual needs in the workplace. In the context of CVTS3 the following types of other forms of CVT are identified:

- Planned training through on-the job-training
- Planned training through job-rotation, exchanges, secondments or study visits
- Planned training through participation in learning or quality circles
- Planned training through self-directed learning
- Planned training through attendance at conferences, workshops, trade fairs and lectures

Excluded learning

Random learning

Random learning can occur in everyday life. It is not an activity which is intentionally planned in advance and is not bound to special or specific places (e.g. classes) or to mediators (e.g. teachers). Random learning can be considered as a natural learning mechanism. Learners may often not be aware that they have learnt something.

Other excluded training

Training in the enterprise for the benefit of a participant who is not a member of the staff of the company nor registered in the formal education and training system that has a work based component, for instance:

- Training of unemployed people at the request of (and nearly always funded by) employment authorities and agencies
- Voluntary stage of a person who wants to be trained
- Short (a few weeks) exposure to work environment, part of a curriculum, but without structured programmes and knowledge and skills acquisition, intention and assessment. As a practical approach, a minimum period of time (e.g. 6 months) might be a useful discerning criterion
- Internships (e.g. for physicians, ...) and periods of practise after courses and before recognition (e.g. for lawyers, ...) should be understood as **work** and not as CVT or IVT

Quality control

There is a focus on maintaining quality and accuracy at every stage of the research process. Firstly at the database stage:

Database

The sample used was provided to the DfE by the Office for National Statistics (ONS), who sourced the data from their Inter-Departmental Business Register (IDBR), which is defined by ONS as:

“A list of UK businesses maintained by National Statistics (NS) and combines the former Central Statistical Office (CSO) VAT based business register and the former Employment Department (ED) employment statistics system. It complies with European Union regulation 2186/93 on harmonisation of business registers for statistical purposes.”

The main strengths of the IDBR are highlighted in its 2001 Review, with the three most relevant to this study being:

- that it is updated frequently [weekly] from administrative sources
- the existence of systems that regularly update the structures of large businesses
- its excellent coverage, which is due to the use of two comprehensive administrative sources [HM Customs and Excise and Inland Revenue]
- that extensive work is undertaken to reduce the level of duplication arising from the use of multiple administrative sources.

In order to minimise errors arising from discrepancies between the sampling frame and the target population and sub-populations (such as under-coverage, over-coverage and misclassification), the sample frame received from IDBR was subject to stringent checks from BMG Research, which brought to light a number of duplications.

The sample frame was subsequently re-counted and these counts were then used to form the basis of the sample structure, rather than the original counts which included duplications. The table in Appendix I illustrates the profile of the sample frame from the IDBR.

Stratum switchers

The survey included a further element of quality control in that it asked respondents to confirm the industry sector in which they operate and the current number of employees. In the majority of cases, respondents confirmed that the industry sector recorded against their organisation on the original sample frame was correct. It was also the majority that confirmed that even if the number of employees reported did not match the number on the original sample frame exactly, the actual number remained within the boundaries of the same size stratum. In a minority of cases, the information given by the respondent

led to the organisation being classified within a different sector and/or size stratum than indicated on the original sample frame. The cases are termed 'stratum switchers'.

The table in Appendix VI presents the proportions of all organisations surveyed that switched stratum within each NACE sector and each size band.

Measurement errors: Questionnaire design

Errors in measurement that may have occurred at the stage of data collection are covered in this section.

The quality control process in BMG's fieldwork began at the questionnaire and sample design stages. Field teams were fully consulted over these issues to ensure that the questionnaire and sample design were unambiguous and as far as possible interviewer friendly.

The questionnaire used in CVTS5 solicited detailed figures on training participants and their demographic make up, training costs and the breakdown of costs and hours spent in training, broken down into sub-groups and as a proportion of all paid working hours.

It was therefore agreed at an early stage that a method of reducing the level of item non-response should be designed, as the complex information solicited may be off-putting for respondents who could not easily, if at all, produce accurate statistics on their company training provision. A number of calculation questions were devised to aid respondents who would otherwise provide a "don't know" response to two key questions that may prove difficult to answer - total working hours and total labour costs. The questionnaire is attached in Annex A.

This proved to be a useful addition to the questionnaire, with many respondents able to respond using the calculations rather than recording a non-response for these questions. This boosted the response to these questions, minimising the level of imputation required and therefore helping to develop a more accurate data set.

CATI method

The data collection stage also benefitted by largely delivering the questionnaire using a CATI method (computer aided telephone interviewing), allowing the addition of logic checks on the data at the point of the interview. This was achieved using Conformat software, building the survey directly into the CATI operating system. This system has the advantage of ensuring that the code frames, open-ended structures and logic checks were built in from the outset. CATI also ensures that data are only input once, at source of collection, increasing the integrity of the processed data. It also ensures that mandatory data is provided and recorded.

The telephone script was then validated by the Project Manager, Associate Director and a Team Leader from the telephone contact centre. The CATI system had pre-set inbuilt logic checks that ensure respondents' answers are justified and are checked as the survey progresses. For example, figures for total employee count were logic checked against later questions looking at how this number is split into age bands and gender so that the sums correlate. The logic checks were again repeated when the data were processed in Merlin.

Subsequent NACE coding makes use of the accurately captured information for detailed coding, using the information recorded verbatim.

Interviewer briefings

The second key stage in the field quality assurance process is the interviewer briefing. In terms of briefing, all interviewers worked from BMG's contact centre in Birmingham, where all briefings were undertaken. BMG employs a stable and substantial field resource, with many interviewers having been employed by the agency for a number of years. The interviewing team is fully conversant with many of the concepts covered by the survey, which meant that a detailed briefing covering issues such as skills and training needs etc was not necessary, and the briefing could therefore concentrate on the specific issues relevant to CVTS5 and to the terminology and definitions it uses.

BMG conducted a briefing for all interviewers prior to starting fieldwork. Each briefing session lasted for approximately two hours, and as part of our quality control procedures, all interviewers who are briefed according to procedure signed a briefing record form.

As a minimum, the verbal briefing covered:

- Project objectives, client profile, and uses for the data collected.
- A review of the questionnaire, interviewing and routing instructions, and issues to look for when recording verbatim or 'other' responses.
- Specific reference to the terminology and definitions used in the survey.
- A profile and introduction to the field in which they are working, and any issues of note.
- A full explanation of the sampling methodology being adopted, including sample design, and quotas.
- Timetable for the project and fieldwork deadlines.
- The field monitoring and QC procedures adopted by BMG.

During the fieldwork stage, all interviewers were supervised by BMG's team leaders, with a ratio of 10 interviewers to one team leader. The outcomes of all field supervision were recorded on the Supervisor Report form which is audited under the industry accredited (IQCS) scheme.

The CATI system software recorded the outcome for every call made, and where an interview was not secured, records the reason for this (eg fax, modem, line engaged). The responses recorded on the system were analysed by BMG's field team leaders to ensure that interviews were completed appropriately, and that the sampling methodology is being properly adhered to.

Remote interviewer monitoring

BMG team leaders listen in remotely, and watch inputting on master terminals, to at least 5% of all telephone interviews, and the outcomes of these checks are again recorded on the Batch Control Sheet. 'Listening-in' and live on-screen monitoring is a systematic process whereby 5-10% of all interviews are listened in to on a daily basis, to confirm that:

- The data code input, grammar, accuracy of spelling, etc.
- The interview was conducted with the appropriate level of respondent.
- That the interviewer conducted themselves in a polite and professional manner, and that the questionnaire was fully adhered to.
- The information is being recorded accurately on the questionnaire.

Data processing

Processing system

After the data collection stage, cleaned and merged data files were validated by a data processing executive. This process of logic checking the data is highly effective using Merlin software. Merlin is a specialist software product for use in the processing of surveys from data entry stage to production of outputs. One of the strengths of the Merlin system is its powerful validation facility, which is used to check and report any data inconsistencies, prior to the main report being run.

Data processing quality checks

At the end of the validation stage any errors encountered in CATI entry were recorded on a Data Entry Evaluation form. The telephone supervisor provided necessary feedback to interviewers, or took other appropriate action. This could initially include instigating a higher level of monitoring on individual telephone interviewers.

The data processing executive, followed by researchers in the account team then carried out a thorough check of the data, checking specifically that;

- Survey weights were applied properly and produced accurate data.
- All required sector/size breakdowns were evident within the data.
- That all labels and table specifications were accurate.

- That the data map specification was properly adhered to and that all sample bases in the data were accurate.
- That there were no logic issues evident in the data - subsequent logic checks involved checking that the responses provided were broadly comparable across respondents in the same size and sector cell. Within these, any responses lying outside of the normal distribution were checked at a micro level, with the result that some call backs were made where verification of outlying responses was required

Appendix I: Population

Sample:

Number of enterprises in the sample frame, by NACE and size (source: IDBR November 2015)							
NACE	Size						All
	10-19	20-49	50-249	250-499	500-999	1000+	
All	120555	61632	30040	3507	1752	1664	219150
B05-B09 - Mining and quarrying and support activities	115	102	70	15	14	17	333
C10-C12 - Manufacture of food products, beverages and tobacco	1064	863	704	111	62	66	2870
C13-C15 - Manufacture of textiles and textile products Manufacture of leather and leather products	810	562	279	31	4	2	1688
C17-C18 - Manufacture of pulp, paper and paper products, Printing of newspapers	1253	810	470	48	14	10	2605
C19-C23 - Manufacture of coke oven products Manufacture of flat glass	1718	1355	1160	147	79	40	4499
C24-C25 - Manufacture of basic iron and steel and of ferro-alloys Manufacture of metal structures and parts of structures	3175	2233	1104	84	27	12	6635
C26-C28 and C33 - Manufacture of electronic components, Repair of fabricated metal products Manufacture of engines and turbines,	2467	1996	1362	167	58	45	6095

Number of enterprises in the sample frame, by NACE and size (source: IDBR November 2015)							
NACE	Size						All
	10-19	20-49	50-249	250-499	500-999	1000+	
except aircraft, vehicle and cycle engines							
C29-C30 - Manufacture of motor vehicles, Building of ships and floating structures	339	275	344	74	39	39	1110
C16+C31-C32 - Sawmilling and planing of wood, Manufacture of office and shop furniture, Striking of coins	2005	1162	612	57	26	7	3869
D-E - Electricity, gas, steam and air conditioning supply, Water supply; sewerage, waste management and remediation activities	1068	552	309	39	12	38	2018
F - Construction	11098	4793	1844	169	67	66	18037
G45 - Sale of cars and light motor vehicles	3807	1666	764	101	51	39	6428
G46 - Agents involved in the sale of agricultural raw materials, live animals, textile raw materials and semi-finished goods	9913	5397	2518	233	95	78	18234
G47 - Retail trade in non-specialised stores with food, beverages or tobacco predominating	11468	4575	1638	166	110	192	18149
H - Transportation and storage	4498	2831	1489	162	103	115	9198
I - Accommodation and food service activities	19682	9906	2985	305	121	135	33134
J - Information and communication	5645	3298	1689	184	93	88	10997
K64-K65 - Financial	490	273	284	51	45	73	1216

Number of enterprises in the sample frame, by NACE and size (source: IDBR November 2015)							
NACE	Size						All
	10-19	20-49	50-249	250-499	500-999	1000+	
and insurance activities, Life insurance							
K66 - Administration of financial markets	1606	874	601	99	50	45	3275
L + M + N + R + S - Real estate, renting and business activities, Professional, scientific and technical activities, Administrative and support service activities, Arts, entertainment and recreation, Other service activities	38334	18109	9814	1264	682	557	68760

Appendix II: Sample

In order to achieve the required number of interviews, contacts were issued as outlined in the table below:

Sample to load	Size						All
	10-19	20-49	50-249	250-499	500-999	1000+	
NACE	10-19	20-49	50-249	250-499	500-999	1000+	All
All	3466	3388	2592	1541	1076	1022	13085
B05-B09 - Mining and quarrying and support activities	115	102	70	15	14	17	333
C10-C12 - Manufacture of food products, beverages and tobacco	177	175	135	88	62	66	703
C13-C15 - Manufacture of textiles and textile products Manufacture of leather and leather products	298	284	132	31	4	2	751
C17-C18 - Manufacture of pulp, paper and paper products, Printing of newspapers	224	218	133	48	14	10	647
C19-C23 - Manufacture of coke oven products Manufacture of flat glass	202	200	133	88	76	40	739
C24-C25 - Manufacture of basic iron and steel and of ferro-alloys Manufacture of metal structures and parts of structures	175	174	145	84	27	12	617
C26-C28 and C33 - Manufacture of electronic components, Repair of fabricated metal products Manufacture of engines and	152	151	136	101	58	45	643

Sample to load	Size						
NACE	10-19	20-49	50-249	250-499	500-999	1000+	All
turbines, except aircraft, vehicle and cycle engines							
C29-C30 - Manufacture of motor vehicles, Building of ships and floating structures	143	138	123	74	39	39	556
C16+C31-C32 - Sawmilling and planing of wood, Manufacture of office and shop furniture, Striking of coins	196	192	143	57	26	7	621
D-E - Electricity, gas, steam and air conditioning supply, Water supply; sewerage, waste management and remediation activities	207	197	150	39	12	38	643
F - Construction	135	134	114	102	67	66	618
G45 - Sale of cars and light motor vehicles	141	139	133	83	51	39	586
G46 - Agents involved in the sale of agricultural raw materials, live animals, textile raw materials and semi-finished goods	209	208	148	116	95	78	854
G47 - Retail trade in non-specialised stores with food, beverages or tobacco predominating	228	226	148	92	85	94	873
H - Transportation and storage	155	154	138	82	76	77	682
I - Accommodation and food service	165	164	150	91	80	82	732

Sample to load	Size						
NACE	10-19	20-49	50-249	250-499	500-999	1000+	All
activities							
J - Information and communication	142	141	119	90	79	78	649
K64-K65 - Financial and insurance activities, Life insurance	137	128	117	51	45	73	551
K66 - Administration of financial markets	130	128	103	91	50	45	547
L + M + N + R + S - Real estate, renting and business activities, Professional, scientific and technical activities, Administrative and support service activities, Arts, entertainment and recreation, Other service activities	135	135	122	118	116	114	740

The achieved number of interviews was as follows:

Achieved sample	Size						All
	10-19	20-49	50-249	250-499	500-999	1000+	
All	961	1009	726	308	174	137	3315
B05-B09 - Mining and quarrying and support activities	8	18	2	0	2	0	30
C10-C12 - Manufacture of food products, beverages and tobacco	61	55	23	13	6	10	168
C13-C15 - Manufacture of textiles and textile products Manufacture of leather and leather products	40	43	12	4	2	1	102
C17-C18 - Manufacture of pulp, paper and paper products, Printing of newspapers	67	66	16	7	2	1	159
C19-C23 - Manufacture of coke oven products Manufacture of flat glass	54	69	50	15	11	5	204
C24-C25 - Manufacture of basic iron and steel and of ferro-alloys Manufacture of metal structures and parts of structures	58	62	43	13	3	1	180
C26-C28 and C33 - Manufacture of electronic components, Repair of fabricated metal products Manufacture of engines and turbines, except aircraft, vehicle and cycle engines	44	61	56	22	10	2	195
C29-C30 - Manufacture of motor vehicles, Building of ships and floating structures	32	29	16	8	4	6	95
C16+C31-C32 - Sawmilling and planing of wood, Manufacture of office and shop furniture, Striking of coins	65	66	28	6	4	1	170

Achieved sample	Size						
NACE	10-19	20-49	50-249	250-499	500-999	1000+	All
D-E - Electricity, gas, steam and air conditioning supply, Water supply; sewerage, waste management and remediation activities	30	59	28	8	2	3	130
F - Construction	43	44	45	15	13	6	166
G45 - Sale of cars and light motor vehicles	40	35	30	14	5	4	128
G46 - Agents involved in the sale of agricultural raw materials, live animals, textile raw materials and semi-finished goods	67	62	55	25	6	15	230
G47 - Retail trade in non-specialised stores with food, beverages or tobacco predominating	76	71	44	19	14	15	239
H - Transportation and storage	35	45	51	15	19	19	184
I - Accommodation and food service activities	48	46	58	33	11	12	208
J - Information and communication	51	40	38	12	1	3	145
K64-K65 - Financial and insurance activities, Life insurance	24	36	12	7	6	1	86
K66 - Administration of financial markets	39	36	32	7	1	3	118
L + M + N + R + S - Real estate, renting and business activities, Professional, scientific and technical activities, Administrative and support service activities, Arts, entertainment and recreation, Other service activities	79	66	87	65	52	29	378

Appendix IV: Sampling Methodology

The calculation used to determine the number of contacts to be issued in order to achieve the required number of interviews by size and sector is as described in the CVTS 5 Manual Version 1-2. It is presented below.

A sample size shall be calculated to ensure a maximum half-length of 0.2 of the 95 % confidence interval for the estimated parameters, which are a proportion of "training enterprises" (after allowance for the non-response rate in the sample) for each of the 60 stratified elements identified above (120 stratified elements for Member States with 50 million inhabitants or more).

The following formula may be used in determining the sample size:

$$n_h = 1 / \left(\left(C^2 * te_h + \frac{1}{N_h} \right) * r_h \right)$$

Where:

n_h = the number of sampling units in the stratum cell, h

r_h = the anticipated response rate in the stratum cell, h

c = maximum length of half the confidence interval

te_h = the anticipated proportion of training enterprises in the stratum cell, h

N_h = the total number of enterprises (training and non-training) in the stratum cell, h

The derivation of the formula, which is used for calculating sample sizes in each stratum is presented below. In CVTS, we are mainly interested in estimating proportions, for example the proportion of training enterprises offering external courses, p_h .

$$C = 2 \sqrt{p_h(1-p_h) \left(\frac{1}{n_h r_h} - \frac{1}{N_h} \right)}$$

$$\frac{C^2}{2^2} = p_h(1-p_h) \left(\frac{1}{n_h r_h} - \frac{1}{N_h} \right)$$

Assume that $p_h = 0.5$

$$\begin{aligned} \Rightarrow C^2 &= \frac{1}{n_h r_h} - \frac{1}{N_h} \\ \frac{1}{n_h r_h} &= C^2 + \frac{1}{N_h} \\ n_h r_h &= 1 / \left(C^2 + \frac{1}{N_h} \right) \\ n_h &= 1 / \left(\left(C^2 + \frac{1}{N_h} \right) * r_h \right) \end{aligned}$$

Allocate the sample sizes in such a way that the expected number of training enterprises will be roughly the same in each stratum.

$$n_h = 1 / \left(\left(C^2 * te_h + \frac{1}{N_h} \right) * r_h \right)$$

Sample sizes should be calculated for a maximum **C-values of 0.2** on the basis of a **20 x 3 (NACE x size) matrix** (20 x 6 matrix in Member States with 50 million inhabitants or more).

The sample size was determined by a number of factors including the population, the anticipated response rate and the propensity to provide training. These were all factors that might introduce bias unless accounted for in the sampling methodology.

In determining the structure of the sample, a C-value of 0.188 and a response rate of 30% was used. This was revised down from the original response rate of 40% half way through fieldwork when it became clear that the original response rate would not be achieved. These variables, when input into the formula, gave an expected return (i.e. number of achieved interviews) of close to 4,000.

The resulting sample structure is illustrated in the grids below.

Anticipated sample	Size						
NACE	10-19	20-49	50-249	250-499	500-999	1000+	All
All	1040	1016	778	462	323	307	3926
B05-B09	35	31	21	5	4	5	100
C10-C12	53	53	41	26	19	20	211
C13-C15	89	85	40	9	1	1	225
C17-C18	67	65	40	14	4	3	194
C19-C23	61	60	40	26	23	12	222
C24-C25	53	52	44	25	8	4	185
C26-C28 and C33	46	45	41	30	17	14	193
C29-C30	43	41	37	22	12	12	167
C16+C31-C32	59	58	43	17	8	2	186
D-E	62	59	45	12	4	11	193
F	41	40	34	31	20	20	185
G45	42	42	40	25	15	12	176
G46	63	62	44	35	29	23	256
G47	68	68	44	28	26	28	262
H	47	46	41	25	23	23	205
I	50	49	45	27	24	25	220
J	43	42	36	27	24	23	195
K64-K65	41	38	35	15	14	22	165
K66	39	38	31	27	15	14	164
L + M + N + R + S	41	41	37	35	35	34	222

Appendix V: Weighting Factors

The weighting factors detailed in the tables below are based on IDBR population statistics and their application to the data addresses issues of non-response.

NACE	10-19	20-49	50-249	250-499	500-999	1000+	All
All	0.5501	0.2812	0.1371	0.0160	0.0080	0.0076	1.0000
B05-B09	0.0005	0.0005	0.0003	0.0001	0.0001	0.0001	0.0015
C10-C12	0.0049	0.0039	0.0032	0.0005	0.0003	0.0003	0.0131
C13-C15	0.0037	0.0026	0.0013	0.0001	0.0000	0.0000	0.0077
C17-C18	0.0057	0.0037	0.0021	0.0002	0.0001	0.0000	0.0119
C19-C23	0.0078	0.0062	0.0053	0.0007	0.0004	0.0002	0.0205
C24-C25	0.0145	0.0102	0.0050	0.0004	0.0001	0.0001	0.0303
C26-C28 and C33	0.0113	0.0091	0.0062	0.0008	0.0003	0.0002	0.0278
C29-C30	0.0015	0.0013	0.0016	0.0003	0.0002	0.0002	0.0051
C16+C31-C32	0.0091	0.0053	0.0028	0.0003	0.0001	0.0000	0.0177
D-E	0.0049	0.0025	0.0014	0.0002	0.0001	0.0002	0.0092
F	0.0506	0.0219	0.0084	0.0008	0.0003	0.0003	0.0823
G45	0.0174	0.0076	0.0035	0.0005	0.0002	0.0002	0.0293
G46	0.0452	0.0246	0.0115	0.0011	0.0004	0.0004	0.0832
G47	0.0523	0.0209	0.0075	0.0008	0.0005	0.0009	0.0828
H	0.0205	0.0129	0.0068	0.0007	0.0005	0.0005	0.0420
I	0.0898	0.0452	0.0136	0.0014	0.0006	0.0006	0.1512
J	0.0258	0.0150	0.0077	0.0008	0.0004	0.0004	0.0502
K64-K65	0.0022	0.0012	0.0013	0.0002	0.0002	0.0003	0.0055
K66	0.0073	0.0040	0.0027	0.0005	0.0002	0.0002	0.0149
L + M + N + R + S	0.1749	0.0826	0.0448	0.0058	0.0031	0.0025	0.3138

Appendix VI: Stratum Switchers

The following table summarises the proportion of all respondents who provided information in the survey which suggested that their actual NACE designation and/or size band differed from that provided on the database.

NACE CODE	NACE AND SIZE CHANGE	NACE ONLY	SIZE ONLY	SIZE CODE	SIZE CHANGE
1	0.1%	0.1%	0.1%	1	8%
2	0.1%	0.1%	1.0%	2	5%
3	0.2%	0.2%	0.6%	3	5%
4	0.2%	0.3%	0.6%	4	4%
5	0.2%	0.4%	0.9%	5	1%
6	0.1%	0.3%	0.8%	6	0%
7	0.2%	0.3%	1.2%	Total	23%
8	0.2%	0.3%	0.6%		
9	0.1%	0.2%	0.8%		
10	0.0%	0.2%	0.8%		
11	0.2%	0.3%	1.0%		
12	0.2%	0.1%	0.5%		
13	0.5%	0.6%	1.2%		
14	0.2%	0.5%	1.4%		
15	0.1%	0.2%	1.2%		
16	0.3%	0.2%	1.6%		
17	0.3%	0.0%	0.6%		
18	0.1%	0.1%	0.6%		
19	0.2%	0.2%	0.8%		
20	0.4%	0.3%	3.1%		
Total	3.7%	5.0%	19.5%		

Appendix VII Call Outcome

Call outcome	CVTS5
Completed interviews	3,315
Appointments	110
Quits	394
Refusal	4,947
Not available during fieldwork	89
Total contacted	8,855
No answer	286
Ring backs	1,924
No one available to interview	391
Total live contacts	2,601
Unobtainable	1,479
Less than 10 employees	154
Total loaded	13,089
Response rate on contacted sample	37%
Response rate on total sample	25%



Department
for Education

© Department for Education

Reference: DFE-RR754A

ISBN: 978-1-78105-823-7

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

Any enquiries regarding this publication should be sent to us at:

alicia.heptinstall@education.gov.uk or www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications