



Comisiynydd y
Gymraeg
Welsh Language
Commissioner



Llywodraeth Cymru
Welsh Government

Quality Report

Welsh Language Use Survey

1. Introduction

The Welsh Language Use Survey (WLS) is a self-completion questionnaire which was handed out to all Welsh speakers (aged 3 or over) that were identified in households where a respondent had participated in the National Survey for Wales. The survey sought to obtain information on those who speak Welsh, how fluent they are, who they speak Welsh with, in what situations and how often they use it.

The National Survey for Wales is a large-scale survey of people across Wales, carried out face-to-face in respondents' homes. Each year, the survey is completed by 14,500 people aged 16 and over (around 660 in each local authority), selected at random to ensure the results are representative.

This quality report sets out how the Welsh Language Use Survey adheres to the European Statistical System definition of quality (Section 2), and provides a summary of methods used to compile the output (Section 3). A separate [quality report](#) has been prepared for the National Survey.

2. Summary of quality

This section covers how the Welsh Language Use Survey meets the six dimensions of quality set out by the European Statistical System: relevance; accuracy; timeliness and punctuality; accessibility and clarity; comparability; and coherence.

2.1 Relevance

The degree to which the statistical product meets user needs for both coverage and content.

<p>What it measures</p>	<p>The Welsh Language Use Survey contains a series of questions about Welsh speakers' ability and use of the Welsh language.</p> <p>As the survey is left for Welsh speakers identified in households where someone has participated in the National Survey, demographic information is available from the National Survey about the respondents to the Welsh Language Use Survey which allows detailed cross-analysis of the results.</p> <p>The adult and young person's questionnaires are slightly different to each other and are both available on the Welsh Language Use Survey webpage. This website also contains a copy of the reminder letter that</p>
-------------------------	---

	was sent to households who had not replied.
Mode	A self-completion questionnaire(s) left for all identified Welsh speakers in each household participating in the National Survey. Returned by post. The questionnaire was bilingual and all respondents were given the opportunity to complete it in their language of choice.
Frequency	Continuous over 2 years from April 2013 to March 2015.
Sample size	An achieved sample of 5,700 adults and 1,500 children during the survey period. Exact sample sizes are provided in the technical report .
Periods available	The main reporting year for the survey is based on questionnaires completed between April 2013 and March 2014, and between March 2014 and April 2015 the following year. Local authority figures can be produced by combining the results from the two years of survey fieldwork.
Sample frame	<p>Welsh speakers identified in the National Survey for Wales are used as the sample for the Welsh Language Use Survey.</p> <p>For the National Survey for Wales the addresses are sampled randomly from Royal Mail's small user Postcode Address File (PAF), an up-to-date list of all UK addresses.</p>
Sample design	<p>The sample for the National Survey is stratified disproportionately by local authority to obtain an approximately equal effective sample size of respondents in each local authority each year. The sample for each reporting year is drawn in one stage and first allocated evenly across each quarter, before being allocated to each month within each quarter. The sampling approach ensures that any consecutive four quarters of data is formally unclustered.</p> <p>For addresses containing more than one household, one household is selected at random. In each sampled household, the National Survey respondent is randomly selected from all adults (aged 16 or over) in the household who regard the sample address as their main residence, regardless of how long they have lived there. Once the selected respondent had completed the National Survey, Welsh Language Use Survey questionnaires were left for all Welsh-speaking members of that household.</p>
Weighting	<p>The National Survey's results are weighted to take account of unequal selection probabilities and for differential non-response, i.e. to ensure that the age and sex distribution of the responding sample matches that of the population of Wales.</p> <p>The Welsh Language Use Survey is also weighted to ensure that the age, sex, local authority and Welsh-speaking ability of the responding sample matches that of the population of Welsh speakers identified in the National Survey for Wales. More information about how the results are weighted is provided in the technical report.</p>
Imputation	No imputation.
Outliers	No filtering of outliers.

More detail on the survey design and methodology is set out in the Technical Report for the [National Survey](#) and the [Welsh Language Use Survey](#).

Primary purpose

The main purpose of the Welsh Language Use Survey is to collect information on how those who identify themselves as being Welsh speakers use the language. It asks about their ability, where they learnt to speak Welsh, how often they use it and in what situations. It will provide information about use of the Welsh language and enable comparisons to be made with the previous Welsh Language Use Surveys conducted in 2004, 2005 and 2006.

The Welsh Language Use Survey is funded jointly by Welsh Government and the Welsh Language Commissioner. It is a continuation of research undertaken on behalf of the Welsh Language Board (WLB) between 2004 and 2006 over a series of three surveys. These surveys were organised as a supplementary questionnaire to the Living in Wales survey, conducted between 2004 and 2006 by the Local Government Data Unit ~ Wales on behalf of the Welsh Assembly Government.

Users and uses

The main users of the results are the Welsh Language Commissioner and Welsh Government Ministers and policy teams. There is also a wide range of other users, including: other UK government and local government organisations; other public sector organisations; academics; members of the public; and the voluntary sector, particularly in Wales. The datasets will be deposited at the [UK Data Archive](#) to ensure that the results are widely accessible for research purposes.

The Welsh Government Welsh Language Strategy 2012-17, *A living language: a language for living*, sets out the Welsh Government aspiration to see an increase in the number of people who both speak and use the language in all aspects of daily life. The Welsh Language Use Survey provides the Welsh Government with information to help measure the progress made toward achieving this goal, and identifies areas where policy needs to be strengthened and developed further.

The principal aim of the Welsh Language Commissioner, an independent body established by the Welsh Language Measure (Wales) 2011, is to promote and facilitate use of the Welsh language. The Commissioner's Strategic Plan 2013-15 identifies five strategic objectives, one of which is to influence the consideration given to the Welsh language in policy developments. The results from the Welsh Language Use Survey will provide the Commissioner with important evidence on the use of the Welsh language in households in Wales (and in other situations), ensuring that policy makers have up-to-date statistical information, as well as an understanding of the current position of the Welsh language.

Strengths and limitations

The strengths of the Welsh Language Use Survey include:

- A large sample size. This allows analysis by population sub-groups and local authority areas.
- The National Survey provides a randomly-selected sample with a high response rate of around 70%. This helps to ensure that the results are as representative as possible of people in Wales, including harder-to-reach groups such as younger working people. The survey is weighted to adjust for non-response, which also helps make sure it is as representative as possible.

- The Welsh Language Use Survey was handed to respondents in person, which helped to ensure a high acceptance rate of 82% for the questionnaires, and for a self-completion post back survey, a return rate of 53% (41% for children) (yielding an overall response rate of 48% for adults, 33% for children; this is relatively high for a self-completion survey).
- Sampling errors are small compared to some social survey designs, because the National Survey has a single stage sample of addresses.
- The National Survey contains a range of demographic questions which allow cross-analysis by age, gender etc.
- Where possible, most questions were worded in the same way as they were in the 2004-06 surveys. This means that they are tried and tested, and that results can be compared over time.
- Respondents to the National Survey were also asked whether they would be willing to be re-contacted for further research and whether they would be willing to link their data to other administrative records already held about them. This allows us to make further use of the information we collect.

The limitations include:

- Although the response rate is relatively high, there is still a substantial proportion of sampled individuals who do not take part. This is likely to affect the accuracy of the estimates produced.
- The survey does not cover people living in institutional establishments (e.g. care homes, residential youth offender homes, hostels, and student halls).
- Care has been taken to make the questions as accessible as possible, but there will still be instances where respondents do not respond accurately, for example because they have not understood the question correctly, because of the inherent limitations of relying on self-assessment of linguistic skills and language use, or because for some reason they do not wish to provide an accurate answer. Again, this will affect the accuracy of the estimates produced.
- As the survey is a self-completion survey, there will be several instances where a respondent may answer the questions without reading the instructions. They may provide contradictory answers and they may miss out questions.
- Robust analyses for small areas (i.e. sub-LA level) and groups are not possible without combining two years of data.

Several of the strengths and limitations mentioned above relate to the accuracy of the results. Accuracy is discussed in more detail in the following section.

2.2 Accuracy

The closeness between an estimated result and the (unknown) true value.

The main threats to accuracy are sources of error, including sampling error and non-sampling error.

Sampling error

Sampling error arises because the estimates are based on a random sample of the Welsh-speaking population rather than the whole population of Welsh speakers. The results obtained for any single random sample are likely to vary by chance from the results that would be obtained if the whole population of Welsh speakers were surveyed

(i.e. a census), and this variation is referred to as the sampling error. In general, the smaller the sample size the larger the potential error.

For a random sample, sampling error can be estimated statistically based on the data collected, using the standard error for each variable. Standard errors are affected by the survey design; and can be used to calculate confidence intervals and coefficients of variation in order to give a more intuitive idea of the size of sampling error for a particular variable. These issues are discussed in the following subsections.

Effect of survey design on standard errors

The National Survey is stratified at local authority level, with different probabilities of selection for people living in different local authorities (e.g. a lower probability for someone living in the Cardiff local authority area than someone in Merthyr Tydfil, given the relatively larger size of Cardiff). One of the effects of using this complex design is that standard errors for the survey estimates are generally higher than the standard errors that would be derived from a simple random sample of the same size.

The ratio of the *standard error* of a complex sample to the standard error of a simple random sample (SRS) of the same size is known as the design factor, or “deft”. If the standard error of an estimate in a complex survey is calculated as though it has come from a SRS survey, multiplying that standard error by the deft gives the true standard error of the estimate, given the complex design.

The ratio of the *sampling variance* of the complex sample to that of a simple random sample of the same size is the design effect, or “deff” (which is equal to the deft squared). Dividing the actual sample size of a complex survey by the deff gives the “effective sample size”. This is the size of a SRS that would have given the same level of precision as the complex survey.

Standard errors adjusted for the survey design and design factors for a selection of key National Survey variables are set out in the [Technical reports](#) for each year.

Confidence intervals

Standard errors can be used to calculate confidence intervals for each survey estimate. The confidence intervals for each estimate give a range within which the ‘true’ value for the population is likely to fall (that is, the figure we would get if the survey covered the entire population). The most commonly-used confidence interval is a 95 per cent interval. This means that, in 95 per cent of survey samples, the 95 per cent confidence interval for each sample will contain the ‘true’ figure for the whole population. The larger the confidence interval, the less precise the estimate is.

95 per cent confidence intervals have been calculated for a range of Welsh Language Use Survey tables which are published on [StatsWales](#). These confidence intervals have been adjusted to take into account the design of the survey and are larger than they would be if the survey had been based on a simple random sample. They are calculated as the mean of the estimate plus or minus approximately 1.96 * the standard error of the estimate.¹

Confidence intervals can also be used to help tell whether there is a real difference between two groups. As a rough guide to interpretation, when comparing two groups, if the confidence intervals around the estimates overlap, it can be assumed that the

¹ The value of 1.96 varies slightly according to the sample size for each particular estimate.

estimates are not statistically significantly different – this approach is not as rigorous as doing a formal statistical test, but is straightforward, widely used and reasonably robust.

Note that checking to see whether two confidence intervals overlap is less likely than a formal statistical test to lead to conclusions that there are real differences between groups. That is, it is more likely to lead to "false negatives": incorrect conclusions that there is no real difference when in fact there is a difference. It is also less likely to lead to "false positives": incorrect conclusions that there **is** a difference when there is in fact none. Carrying out many comparisons increases the chance of finding false positives. Therefore, when many comparisons are made the conservative nature of the test is an advantage because it reduces (but does not eliminate) this chance.

Coefficients of variation

To ensure that the estimates reported are robust, coefficients of variation (CVs) are calculated for each estimate. These are reported in the tables of survey results published on [StatsWales](#). The CV is calculated as the standard error multiplied by 100 and divided by the mean, and indicates the robustness of an estimate. The table below shows how we colour code our results according to the CV and how we interpret them.

	Estimate is precise	$0 \leq CV < 5$
	Estimate is reasonably precise	$5 \leq CV < 10$
	Estimate is considered acceptable	$10 \leq CV < 20$
	Estimate is not reliable	$CV \geq 20$
-	Value is suppressed due to small cell size (fewer than 30 responses)	

An estimate with a coefficient of variance greater than 20 is not considered reliable for practical purposes.

In tables of results, values with cell sizes of less than 30 (i.e. where fewer than 30 people selected the relevant answer option(s)) are suppressed and shown as “-”.

Non-sampling error

Non-sampling error covers all differences between the survey estimates and true population values except those arising because of sampling error. Unlike sampling error, non-sampling error is present in censuses as well as sample surveys. Types of non-sampling error include: coverage error, non-response error, measurement error and processing error.

It is not possible to eliminate non-sampling error altogether, and it is not possible to produce statistical estimates of the size of non-sampling error. Substantial efforts have been made to reduce non-sampling error in the National Survey and the Welsh Language Use Survey. Some of the key steps taken are discussed in the following subsections.

Measurement error: question development

To reduce measurement error, harmonised or well-established questions that were previously tried and tested are used in the survey where possible.

Non-response

Non-response is a key source of non-sampling error. Response rates are therefore an important dimension of survey quality and are monitored closely. The quarterly and annual target response rate for the National Survey at national level is 70%, with a contractual minimum of 65% rate at LA level in each reporting year. The achieved

response rate for the National Survey in 2013-14 was 70% and 2014-15 was 64%. More details on achieved response rates are given in the [Technical reports](#).

The Welsh Language Use Survey was handed to respondents in person, which helped to ensure a high acceptance rate of 81% for the questionnaires. The respondents who had not returned a completed questionnaire within two weeks were sent a reminder pack through the post, containing a replacement questionnaire and a cover letter explaining about the survey and inviting them again to take part. If they did not return a questionnaire after being sent a postal reminder, they received an additional reminder, conducted by telephone. This resulted in a return rate of 53% (41% for children) (yielding a response rate of 48% for adults, 33% for children).

It must be remembered that as this is a self-completion questionnaire, there may be instances where respondent's literacy levels may affect their ability and propensity to respond. The interviewers however were briefed to help respondents to complete the survey, should they wish or ask to have assistance.

The survey results are weighted to take account of differential non-response across population subgroups, i.e. to ensure that the age and sex distribution of the responding sample matches that of the population of Welsh speakers in Wales. This step is designed to reduce the non-sampling error due to differential non-response.

Missing answers

Missing answers occur for several reasons, including refusal or inability to answer a particular question, and cases where the question is not applicable to the respondent. This is a particular problem for self-completion surveys. Missing answers are usually omitted from tables and analyses.

Measurement error: interview quality checks

Another cause of bias may be that interviewers systemically influence responses in some way. Extensive interviewer training is provided to minimise this effect, although it remains likely in the initial face-to-face survey that responses will still be subject to effects such as social desirability bias (for example, the respondents may exaggerate their Welsh-speaking ability as their answer may be affected by what the respondent perceives to be socially acceptable). Interviewers are also closely supervised with 12% of interviews being verified through 'back-checking'. Once the respondent has received the Welsh Language Use Survey, the interviewer bias is less likely to have a large impact on data as it is collected through a self-completion questionnaire.

Data scanning and editing

The self-completion questionnaires are sent to the contractor's operations centre, and following scanning, the self-completion questionnaires, data and electronic images are stored on a computer drive and linked to the National Survey data through serial numbers (at both household and individual levels).

Due to the unmonitored nature of self-completion questionnaires, it was necessary to subject all returned questionnaires to a data editing process to ensure that the information recorded was consistent and accurate. Editing was carried out to address instances in which the questionnaire appeared to have been filled in incorrectly; for example where routing has not been followed, where respondents have coded more than one answer where only one was required, or where incompatible answers have been entered.

Processing error: data validation

The main survey outputs are provided in two SPSS datasets, one containing data from the adult questionnaire and another containing data from the children's questionnaire. Each dataset contained weighting variables, which are explained in more detail in the [technical report](#). Each dataset was checked by a member of the survey contractor's data processing team and then by members of the survey contractor's project team. A formalised set of checks on the content and format of the datasets is then carried out by the Welsh Government before the datasets are signed off.

2.3 Timeliness and punctuality

Timeliness refers to the lapse of time between publication and the period to which the data refers. Punctuality refers to the time lag between the actual and planned dates of publication.

The main reporting year for the survey is from April to March of the following year. The publication of the results of the Welsh Language Use Survey is within 9 months after the end of the fieldwork period.

All Welsh Language Use survey outputs are pre-announced on the Welsh Government's [Upcoming calendar](#) web pages, in line with the Code of Practice for Official Statistics.

Should the need arise to change the pre-announced release schedule, the Welsh Government's [Revisions, Errors and Postponements process](#) is followed, and a full explanation made available of the reasons for the change.

2.4 Accessibility and clarity

Accessibility is the ease with which users are able to access the data, also reflecting the format(s) in which the data are available and the availability of supporting information.

Clarity refers to the quality and sufficiency of the metadata, illustrations and accompanying advice.

Publications

All reports published by the Welsh Government will be available to download from the [Welsh Government](#) and [Welsh Language Commissioner](#) webpages. As this survey is jointly funded by the Welsh Language Commissioner, other reports on the survey may be published separately on the two organisations' websites.

Statistics from the National Survey are pre-announced, and are published on the Welsh Government website at 9.30am on the day of publication. All releases are available to download without charge. An RSS feed alerts registered users to each publication. Simultaneously the releases are also published on the National Statistics Publication Hub, and on [StatsUserNet](#). Outputs will be announced and shared on the Welsh Government and StatisticsWales Twitter feeds.

Key results are available to download as spreadsheets from the [StatsWales website](#).

Further information regarding the survey results, for example if you would like to see a different breakdown of results, can be obtained by contacting the Welsh Government via e-mail (welshlanguage@wales.gsi.gov.uk) or the Welsh Language Commissioner

(post@welshlanguagecommissioner.wales), or by phoning the Welsh Government (029 2037 0373) or the Welsh Language Commissioner (0845 6033221).

Disclosure control

Outputs are designed to ensure that individuals are not identifiable from the published results. We adhere to the requirements for confidentiality and data access as set out in Principle 5: Confidentiality, in the Code of Practice for Official Statistics.

Quality assurance

Some weeks before publication, a draft version of each release is shared under restrictive conditions with relevant analytical colleagues from the Welsh Government and the Welsh Language Commissioner, to ensure that the publication is clear, accurate and comprehensive.

Language requirements

All our publications and outputs relating to the Welsh Language Use Survey are published in Welsh and English and adhere to the Welsh Government accessibility policy. We also aim to use Plain English / Cymraeg Clir for our outputs.

UK Data Archive

An anonymised version of the annual dataset (from which some information is removed to ensure confidentiality is preserved), together with supporting documentation, is deposited with the UK Data Archive after the publication of each year's results. These datasets may be accessed by registered users for specific research projects. The UK Data Archive is accessed via the [Economic and Social Data Service \(ESDS\) website](#). Datasets can be downloaded in a number of formats, including SPSS and CSV format.

From time to time, researchers may wish to analyse more detailed data than is available through the Data Archive. Requests for such data can be made to the National Survey team (see contact details below). Requests are considered on a case by case basis, and procedures are in place to ensure that confidentiality is maintained.

Methods and definitions

The Welsh Language Use Survey [Technical report](#) contains an overall description of the way the survey has been undertaken. It provides information on topics such as sampling, fieldwork procedures and weighting. Each survey publication also contains a glossary with descriptions of more general terms used in the output. Copies of the questionnaires are available [here](#).

2.5 Comparability and coherence

The degree to which data can be compared over both time and domain.

The degree to which data that are derived from different sources or methods, but which refer to the same phenomenon, are similar.

The 2013-15 Welsh Language Use Survey (WLS) was a continuation of research previously undertaken between 2004 and 2006. The survey was designed to be carried out in a way that most closely replicated the Welsh Language Use surveys that were carried out in 2004, 2005 and 2006, so that comparisons could be made between the two surveys.

For the 2004-2006 surveys, the Living in Wales survey provided the sample frame (i.e. those identified during the LIW interview as being able to speak Welsh). The Living in Wales survey was conducted (where possible) with the Household Representative Person. This person was defined as the person most 'responsible for the household', who earned the most or the oldest. It was carried out from March to July and had an achieved sample of 7,500 people. The National Survey for Wales was asked of a random adult in the household, continuously throughout the year and had an achieved sample of 14,500. Therefore there were some differences in the way the survey was administered.

It was also noted that a greater proportion of people indicated that they spoke Welsh in the 2014-15 National Survey than in the Living in Wales survey. However a smaller proportion of these Welsh speakers stated that they spoke Welsh fluently, therefore some caution should be used when comparing results between the 2013-15 and 2004-06 surveys.

The number and percentage of fluent Welsh speakers was calculated by multiplying the percentages who were recorded as fluent in the Welsh Language Use Survey by the percentage of Welsh speakers identified in the National Survey for Wales. This method was used to ensure consistency with the analysis carried out for the 2004-06 surveys.

Some new questions were introduced in 2013-14 and some questions that were previously used in the 2004-06 survey have also been subject to minor changes. Where comparisons are available, this is noted clearly in outputs.

The survey results can be compared for different sub-groups (e.g. age, sex, fluency levels, geography, socio-demographic factors). However, some caution should be used when comparing results for sub-groups such as local authorities or for children, since differences are not always significant due to smaller sample sizes.

Feedback or further information

If you would like further information, please visit the [Welsh Government](#) website, or the [Welsh Language Commissioner](#) website. We welcome comments from users of our publications, for example on content and presentation.

Martin Parry
Knowledge and Analytical Services
Welsh Government
Cathays Park
Cardiff
CF10 3NQ

Produced by Knowledge and Analytical Services, Welsh Government
Last reviewed: November 2015