

Differences in estimates of Welsh Language Skills

Lucy Haselden

Ethnicity and Identity Branch, ONS

1 Aims of the paper

The proportion of people able to communicate in Welsh is of considerable interest to policy makers, particularly in Wales. They want to monitor schemes that promote the Welsh language and assess the advantages that can be gained by speaking Welsh.

Since there is a large amount of regional variation in Welsh language usage, monitoring has to be done at local authority level; so data sources which can be analysed at this level are needed.

The two key current sources of this information are the Census and the Welsh Local Labour Force Survey (WLLFS). They both ask the same simple questions on whether the respondent can understand, speak, read and/or write Welsh and both produce data that can be analysed at local authority level. However, they produce very different estimates of how many people have these skills.

The aim of this paper, therefore, is to:

- establish the level of these differences;
- give possible explanations for these differences; and
- indicate whether either source gives a better measure of Welsh language skills.

2 Background to the questions on Welsh language

2.1 The Welsh language question on the 2001 Census

A Welsh language question has been asked in Wales on every Census held since 1891.

The 2001 Census Welsh language question appeared immediately after the question on ethnicity and cultural background on the Census form in Wales. It differed slightly from the 1991 question as it asked “Can you speak Welsh?” as opposed to “Do you speak Welsh?”, which was asked in 1991. It also asked respondents whether they could understand spoken Welsh, which had not been included previously.

The question was:

<p>Can you understand, speak, read or write Welsh?</p> <p>◆✓ all the boxes that apply</p> <p><input type="checkbox"/> Understand spoken Welsh</p> <p><input type="checkbox"/> Speak Welsh</p> <p><input type="checkbox"/> Read Welsh</p> <p><input type="checkbox"/> Write Welsh</p> <p><input type="checkbox"/> None of the above</p>

The ethnicity question, which immediately preceded the Welsh language question, had a tick box for ‘White –British’ but none of the tick boxes contained the word ‘Welsh’. Following a media campaign against the lack of a Welsh tick box, ONS responded by inviting respondents to express their Welsh identity by writing in “Welsh” against the appropriate ‘Any other ...’ category.

2.2 The Welsh language question on the 2001 Labour Force Survey

The Annual Local Area Labour Force Survey (ALALFS) is the largest UK household survey and the sample size has recently been increased in order provide reliable estimates at national, regional and local authority level. In particular, the sample in Wales has been increased to facilitate analysis at local authority level there. This part of the survey is known as the Welsh Local Labour Force Survey (WLLFS).

A set of questions on Welsh language skills was introduced in March 2001. These were designed to be as comparable as possible to the 2001 Census question but some adaptations were made to make them suitable for face-to-face and telephone interviewing (as opposed to the self-completion form on the 2001 Census).

The questions on Welsh language skills in the WLLFS are asked immediately after the question on national identity¹ and before the question on ethnicity.

The WLLFS Welsh Language questions are:

Can you understand spoken Welsh? Yes No
Can you speak Welsh? Yes No
Can you read Welsh? Yes No
Can you write Welsh? Yes No

2.3 Welsh language questions on other surveys

A number of other surveys have recently included questions on Welsh language skills. Although these may not have been widely used to inform policy, it is useful to compare their methodologies and results with those of the 2001 Census and the Welsh Local Labour Force Survey.

By comparing all these surveys, we can see the full range of estimates and determine to what extent survey methodology might account for different estimates.

Surveys discussed in this paper are:

- The Welsh Social Survey (Welsh Office, 1992)
- The Welsh Omnibus Survey (Beaufort Research on behalf of ONS, September 2000 and November 2000)

¹ The national identity question in Wales asks “What do you consider your national identity to be? Please choose as many or as few as apply”:

1. Welsh
2. English
3. Scottish
4. Irish
5. British
6. Other answer

- The Welsh Household Panel Survey (which is part of the British Household Panel Survey) (ESRC, 1999)
- The Welsh Household Interview Survey (Welsh Office, 1997)

Table 1 shows details the exact wording of Welsh language questions, the preceding questions (to give context), who was asked the question, the mode of interviewing and details of the sample for all these key sources.

Table 1 Details of Welsh language skills questions, by source; 1991- 2002

	Phrasing of Welsh language question	Answer categories	Preceding question	Population asked the question	Mode of interview	Sample size and details
Census (1991)	“Does the person speak, read or write Welsh?” Please tick the appropriate box(es)	1. Speaks Welsh 2. Reads Welsh 3. Writes Welsh 4. Does not speak, read or write Welsh	Name, sex and date of birth	All residents of Wales aged 3 and over	Self-completion form	N = 2,723,623 Full count
Welsh Social Survey (1992)	“Do you speak Welsh?” “Can you read Welsh?”	<ul style="list-style-type: none"> • Yes • No 	Do you read Welsh language.. Books Magazines Local papers (papurau bro)	Given by the householder on behalf of all aged 3-15, and asked individually of all over 15	Self-completion form, but 80.5% completed during face-to-face-interview	27,720 completed questionnaires from a stratified random sample.
Welsh Social Survey (1992)	“Which of the following best describes you:”	<ol style="list-style-type: none"> 1. I can’t speak any Welsh and never have been able to 2. I can’t speak Welsh now although I could once 3. I can only speak a little Welsh 4. I can speak a fair amount I’m fluent in Welsh... 5.but never or hardly ever speak it 6. but speak it only occasionally 7. and speak it about half the time 8. and speak it all or most of the time 				
Welsh Household Interview Survey (1997)	Can (name of respondent speak Welsh?)	<ol style="list-style-type: none"> 1. Yes 2. No 	Name, age, sex and marital status	Residents of households in Wales aged 3 and over	Face-to-face interview	N=28360 interviews from a stratified random sample of 40000 people (70.9% response)
Welsh Household Panel Survey (1999)	“Can you understand, speak, read or write Welsh?”	<ol style="list-style-type: none"> 1. Understand spoken Welsh 2. Speak Welsh 3. Read Welsh 4. Write Welsh 5. None of the above 	Country of parents’ birth	Residents of households in Wales aged 16 and over	Face-to-face interview	N = 4,368 Two-stage clustered probability design and systematic sampling
Welsh Omnibus Survey (Sep 2000)	“Do you speak Welsh? If Yes: “Is that fluently?”	<ol style="list-style-type: none"> 1. Yes, fluently 2. Yes, not fluently 3. No 	Cultural background and ethnic group (no specific Welsh category)	Residents of households in Wales aged 16 and over.	Face-to-face interview	N = 1002 Quota sample, with interlocking quota controls of age and social class within sex

Continued over page

Table 1 (continued) Details of Welsh language skills questions, by source; 1991- 2002

	Phrasing of Welsh language question	Answer categories	Preceding question	Population asked the question	Mode of interview	Sample size and details
	“Do you understand Welsh?”	1. Yes 2. No				
	“Can you read or write Welsh?” (code all that apply)	1. Read Welsh 2. Write Welsh 3. No, neither				
Welsh Omnibus Survey (Nov 2000)	“Can you understand spoken Welsh?” “Can you speak Welsh?” “Can you read Welsh?” “Can you write Welsh?”	For each question: 1. Yes 2. No	Questions regarding a utility and a charity. Questions on national identity and ethnic group appeared earlier in the questionnaire	Residents of households in Wales aged 16 and over	Face-to-face interview	N = 1005 Quota sample, with interlocking quota controls of age and social class within sex
Welsh Local Labour Force Survey (2001/02)	“Can you understand spoken Welsh?” “Can you speak Welsh?” “Can you read Welsh?” “Can you write Welsh?”	1. Yes 2. No	National identity	Residents of households in Wales (tables produced for those aged 3 and over)	Face-to-face or telephone interview	N = 48,809 Random probability sample
Census (2001)	“Can you understand, speak, read or write Welsh?” ✓ all the boxes that apply	1. Understand spoken Welsh 2. Speak Welsh 3. Read Welsh 4. Write Welsh 5. None of the above	Ethnic group and cultural background	All residents of Wales (tables produced for those aged 3 and over)	Self completion form	N = 2,8005,701 Full counts

3 Comparing results of Welsh language questions

Tables 2 and 3 show the different estimates for understanding, speaking, reading and writing Welsh that the various sources have produced. These tables show the extent of the differences in the estimates between the sources.

Table 2 shows estimates for those aged 16 and over and Table 3 shows estimates for those aged 3 and over. Not all sources produced results for both these age groups, therefore each table only shows estimates where they are available. They show:

- The 2001 Censuses produced the lowest estimates for the proportions understanding, speaking, reading and writing Welsh than for other sources collected since 1997 (**Tables 2 and 3**).
- WLLFS estimates were appreciably higher than Census figures. Although sampling errors are not currently available for these estimates, the sheer size of the sample means that sampling errors alone are unlikely to explain all this difference (**Tables 2 and 3**).
- The variation in estimates is smallest for writing Welsh and largest for understanding Welsh (**Tables 2 and 3**).
- Both Welsh Omnibus surveys produced similar estimates to the WLLFS for reading and writing Welsh. The higher estimates of understanding and speaking Welsh on the September 2000 omnibus can be explained by differences in the question wording (**Tables 1 and 3**).

Table 2 Comparison of estimates of the proportion understanding, speaking, reading and writing Welsh produced by different sources; people aged 3 years and over

Wales									Percentages
	Census (1991)	Welsh Social Survey (1992)	Welsh Household Interview Survey (1997)	Welsh Household Panel Survey (1999) [†]	Welsh Omnibus Survey (September 2000) [†]	Welsh Omnibus Survey (November 2000) [†]	Welsh Local Labour Force Survey (2001/02)	Census (2001)	Range (for surveys where all 4 skills were asked about)
Understands spoken Welsh	Not asked	Not asked	Not asked	-	-	-	35	24	24-35 (11)
Speaks Welsh	18	22	20	-	-	-	30	21	21-30 (9)
Reads Welsh	16	23	Not asked	-	-	-	29	20	20-29(9)
Writes Welsh	14	Not asked	Not asked	-	-	-	25	18	18-25 (7)

[†] Data not available for this age group

Table 3 Comparison of estimates of the proportion understanding, speaking, reading and writing Welsh produced by different sources; people aged 16 years and over

Wales									Percentages
	Census (1991)	Welsh Social Survey (1992)	Welsh Household Interview Survey (1997)	Welsh Household Panel Survey (1999)	Welsh Omnibus Survey (September 2000)	Welsh Omnibus Survey (November 2000) [†]	Welsh Local Labour Force Survey (2001/02)	Census (2001)	Range (for surveys where all 4 skills were asked about)
Understands spoken Welsh	Not asked	Not asked	Not asked	27	35 ^{††}	30	30	21	21-35 (14)
Speaks Welsh	17	19	18	22	30	25	24	17	17-27(10)
Reads Welsh	15	21	Not asked	21	26	24	25	17	17-26 (11)
Writes Welsh	13	Not asked	Not asked	17	20	20	21	15	15-21 (8)

^{††} Unlike the other surveys, only non-fluent Welsh speakers and those who said they did not speak the language at all were asked if they could understand Welsh. In order to make the results comparable with the other surveys, the figures shown here have been derived so that the proportion of people who could understand Welsh = proportion of people who can speak Welsh fluently + non-fluent Welsh speakers who said they could understand Welsh.

Tables 4 and 5 concentrate on the differences in Welsh language skills estimates between the WLLFS and the 2001 Census. They compare the differences by age and by local authority area, listed geographically:

- WLLFS is considerably higher than the 2001 Census for all estimates (**Tables 4 and 5**).
- Differences between the Census and the WLLFS are not consistent across age groups. There is a greater difference between the survey estimates for those aged under 25 than for older age groups (**Table 4**).
- The difference between the WLLFS results and the Census results also varied considerably geographically. The greatest discrepancies could be found in Merthyr Tydfil; Ceredigion; Rhondda, Cyon, Taff; Conwy and Gwynedd. There was no obvious common characteristic between these areas as they were geographically spread, had differing proportions of people with Welsh language skills, different Census response rates and different age profiles (**Table 5**).
- Despite the fact that the difference between WLLFS and Census estimates for each of the Welsh language skills varied by local authority area, both sources showed the Isle of Anglesey and Gwynedd as having the highest proportions of people understanding Welsh and both showed Monmouth, Torfaen, Blaenau Gwent and Newport as having the lowest proportion of people understanding Welsh (albeit ranked slightly different) (**Table 5**).

Table 4 Comparison of estimates of the proportion understanding, speaking, reading and writing Welsh on the 2001/2002WLLFS and the 2001 Census across different age groups

percentages

Age groups	Understanding Welsh			Speaking Welsh			Reading Welsh			Writing Welsh		
	WLLFS	Census	Difference	WLLFS	Census	Difference	WLLFS	Census	Difference	WLLFS	Census	Difference
3 to 15	51	35	16	48	38	10	41	35	6	41	34	7
16 to 24	39	25	14	35	22	13	35	23	12	32	21	11
25 to 34	28	20	8	23	16	7	22	16	6	19	13	6
35 to 44	27	19	8	21	14	7	20	14	6	17	12	5
45 to 64	29	20	9	22	16	6	23	16	7	18	13	5
65 and over	31	22	9	24	20	4	26	19	7	20	15	5
Total	35	24	11	30	21	9	29	20	9	25	18	7

Table 5 Comparison of estimates of the proportion understanding, speaking, reading and writing Welsh on the 2001/2002WLLFS and the 2001 Census across local authorities; All people aged 3 and over

percentages

	Understanding Welsh			Speaking Welsh			Reading Welsh			Writing Welsh		
	Census	WLLFS	Difference	Census	WLLFS	Difference	Census	WLLFS	Difference	Census	WLLFS	Difference
Isle of Anglesey	64	74	10	60	66	6	54	59	5	51	55	4
Gwynedd	69	83	14	69	78	9	64	72	8	61	69	8
Conwy	35	50	15	29	41	12	27	36	9	25	31	6
Denbighshire	31	44	13	26	38	12	25	34	9	22	29	7
Flintshire	17	30	13	14	26	12	14	24	10	12	20	8
Wrexham	19	28	9	15	23	8	15	21	6	12	18	6
Powys	25	34	9	21	28	7	20	28	8	17	25	8
Ceredigion	55	70	15	52	60	8	49	58	9	45	54	9
Pembrokeshire	24	30	6	22	25	3	20	23	3	18	21	3
Carmarthenshire	56	64	8	50	53	3	46	51	5	40	46	6
Swansea	18	27	9	13	20	7	14	21	7	11	17	6
Neath Port Talbot	23	33	10	18	24	6	19	26	7	15	21	6
Bridgend	14	25	9	11	18	7	14	21	7	10	17	7
Vale of Glamorgan	13	27	14	11	20	9	12	21	9	10	18	8
Rhondda, Cyon, Taff	16	31	15	12	24	12	15	27	13	12	22	10
Merthyr Tydfil	13	31	18	10	25	15	12	26	14	9	23	14
Caerphilly	12	23	11	11	20	9	12	20	8	10	17	8
Blaenau Gwent	9	20	11	9	18	9	9	16	7	8	15	7
Torfaen	10	19	9	11	17	6	10	16	6	9	14	5
Monmouthshire	9	18	9	9	16	7	9	15	6	8	13	5
Newport	9	20	11	10	18	8	9	16	7	8	15	5
Cardiff	13	25	12	11	22	11	12	21	9	10	18	8
Total	24	35	11	21	30	9	20	29	9	18	25	7

4 Explanations for the differences

4.1 A loosely defined concept

The concepts of ‘understanding’, ‘speaking’, ‘reading’ and ‘writing’ Welsh are poorly defined. For example, a person who says they understand Welsh may only understand a few words or may be completely fluent in the language. Respondents will have applied their own definitions and will have decided for themselves whether their skills meant they could answer “yes” or tick the appropriate box.

The definitions that people have used will not only vary from person to person, but they will be heavily influenced by the questionnaire design. Subtle changes in the question wording and context may affect the answers people give. The effects of question wording, routing and context are discussed in some detail below.

Of all these concepts, ‘understanding’ Welsh is perhaps the most poorly defined because it could encompass a very wide range of skills. It is not surprising, therefore that it produced the widest range of estimates from 21% of adults on the 2001 Census to 30% of adults on the WLLFS. ‘Writing’ Welsh is a slightly better defined skill as it requires some demonstrable ability and therefore it has a narrower range of results: 21% of adults [aged 16 and over] said they could write Welsh on the WLLFS compared to 15% on the 2001 Census.

The 1992 Welsh Social Survey provides some evidence as to how ‘Do you speak Welsh?’ may be interpreted. Whereas 21.5% said they spoke Welsh, 13.4% described themselves as fluent, 3.5% described themselves as able to speak a fair amount and 17% said they could only speak a little. Thus, 33.9% said they could speak at least a little Welsh while 16.9% said they could speak a fair amount or were fluent.

4.2 Question wording and answer categories

For most of these sources, the question wording was very similar but where there were differences, the sources produced very different results.

The September 2000 Welsh Omnibus asked:

Do you speak Welsh? If “Yes” Is that fluently?. The answer categories were:

1. Yes, fluently
2. Yes, not fluently
3. No

The question allowed respondents who were uncertain about their skills, a third option, that is, ‘not fluently’. This meant that more people could be included in the overall “Yes” group and so the survey gave a much higher estimate of the proportion of people speaking Welsh: some 35% said they did speak Welsh on this survey compared with 30% on the WLLFS and 21% on the 2001 Census who said they could.

There was also a change in the wording of the question from ‘Can you...’ to ‘Do you..’. One would normally expect that more people ‘can’ speak Welsh (which is about ability) than ‘do’ (which is about usage). However, in this case the question wording effect is counteracted by the larger effect of the increased number of answer categories.

The 1992 Welsh Social Survey asks if respondents ‘do’ speak Welsh and if they ‘can’ read Welsh. Twenty-two per cent of respondents said they **did** speak Welsh but a higher proportion, 23% said they **could** read Welsh. On every other source, the proportion able to speak Welsh was higher than the proportion able to read it.

Therefore it is likely that the change from 'do' to 'can' resulted in an increase in the proportion of people reporting that they had that skill.

4.3 Question routing

The September 2000 Welsh Omnibus also had slightly different question routing to all the other sources. On all the other sources, a question on understanding Welsh was asked first and was asked of all respondents. On the September 2001 Omnibus survey, this question was only asked of those who said they did not speak Welsh fluently. It was assumed that all fluent Welsh speakers could also understand the language. Again, this approach produced much higher estimates of the proportion of people understanding Welsh than on other surveys (35% compares with 30% on the WLLFS and 21% on the Census).

4.4 Question context

The answers a respondent gives to one question may be affected by the context it is asked in. The context will vary according to the main subject matter of the survey, the social and political climate of the day, the way the survey was introduced as well as the answers respondents have given to previous questions.

On these surveys, the context of the Welsh language questions varied. Not only did they cover very different subject matter but some surveys included questions about Welsh identity while others did not.

On the WLLFS and the November 2000 Welsh Omnibus survey, respondents had been asked about their national identity earlier on in the questionnaire. This national identity question gave respondents an opportunity to say they were Welsh. It is possible that having just said they were Welsh, the respondents felt more inclined to say that they had Welsh language skills. These two surveys produced very similar estimates for understanding, speaking, reading and writing Welsh.

The Welsh Household Panel Survey did not ask about national identity but the questions on Welsh language were immediately preceded by questions about the respondents' parents' country of birth. Again, if respondents had both parents born in Wales and were born in Wales themselves, it may have meant they were more likely to say they understood, spoke, read or wrote Welsh. The WHPS produced slightly lower estimates for each of the Welsh language skills than the WLLFS and November 2000 Welsh Omnibus Survey but higher estimates than the 2001 Census (27% of WHPS respondents aged 16 and over said they could understand Welsh compared to 30% of the same age group on the WLLFS and 21% of Census respondents).

The 2001 Census question was immediately preceded by a question on ethnic group. Although this question did allow respondents to write in that they were Welsh, there was no Welsh tick box. Unlike on the WLLFS and the November 2000 Welsh Omnibus, respondents were not encouraged to think about their Welsh identity before answering the question on Welsh language. It is therefore not surprising that a smaller proportion of people said they had Welsh language skills on the 2001 Census than on the WLLFS. It is also worth noting that respondents to the Welsh Household Interview survey were also not asked about their Welsh identity before answering a question on whether they spoke Welsh (this survey did not ask about understanding, reading or writing Welsh). The Census produced only a slightly higher proportion of people speaking Welsh than the WHIS: 21% against 20% of those aged 3 and over.

The context of a question can also be affected by current affairs for example if there is a lot of media or political interest in Welsh language issues.

4.5 Survey mode

The 2001 Census form was completed by respondents, usually without reference to an enumerator or any other person employed by ONS. Every effort was made to ensure that the question was easy to understand and was ‘user friendly’ but, as with all self-completion forms, data quality is unlikely to be as high as in a face-to-face interview. In particular, since most questions on the Census asked for a single answer, respondents may have overlooked the instruction that they should tick all the answers that applied for this particular question. Indeed, preliminary analysis of the Census data shows that 3% of those aged 3 and over in Wales indicated that they were able to speak Welsh but could not understand it. It is probable that many of those people could understand spoken Welsh and either forgot to tick the box or thought that it was implicit. Census processing did not attempt to ‘second guess’ respondents.

On the WLLFS, interviews are either held face-to-face with respondents in their own home or by telephone. Telephone interviewers are based in Titchfield (England) and are closely monitored. They follow the same questionnaire as those who conduct face-to-face interviews. Field interviewers tend to be locally recruited, but not necessarily able to speak the language. It is clearly not possible to monitor face-to-face interviews in the same way as telephone interviews but field interviewers are occasionally accompanied on household visits by a trainer.

In every area and age band, a greater proportion of WLLFS respondents interviewed by telephone said they were able to understand, speak, read or write Welsh than those interviewed by face to face interviewer: some 35% of telephone respondents said they could speak Welsh compared with 30% of face-to-face respondents. There were similar differences for understanding and writing Welsh. Furthermore, 36% of those interviewed by telephone said they could read Welsh compared with 27% of those interviewed face-to-face. This difference was most marked in younger age groups.

Since the questionnaires are identical for both telephone and face-to-face interviews, this is unlikely to be as a result of a questionnaire effect. Respondents who are not very confident of their Welsh language abilities might give a different answer to an interviewer on the relatively anonymous medium of the telephone than they would give to an interviewer who came to their house.

4.6 Proxy data

Although more than half (54%) of the responses given on the WLLFS are given in person, all children’s responses and a large proportion of adult responses are given by proxy. The WLLFS records whether a proxy interview was given on behalf of a child under 16, on behalf of a spouse or partner or on behalf of someone else.

The proportion of adults who give proxy responses varies considerably by age. More than half (53%) of all responses given by 16-24 year olds are by (non-spousal) proxies and another 4% were proxy responses given by spouses or partners. This compared with about a third of responses were given by proxy for other age groups. In most of these cases, it was a spouse or partner who answered on their behalf.

Proxy responses given by people in this 16-24 year age group vary significantly from those given in person: only 32% of personal respondents said they could understand Welsh compared to 46% of (non-spousal) proxy respondents. This may be because parents answering on behalf of their children often assume that their children have learnt Welsh at school and so must be able to understand it, even though some of these children may not have been so confident about their skills.

The Census form must be signed by a householder, who is responsible for ensuring that all the details on it are correct. There have been no published studies on who completed in the household completed the 2001 Census form or how well they checked the data. However, It is likely that the pattern of proxies was different for the Census than it was for the WLLFS and this might explain some of the differences between the sources, especially for those aged under 25 years.

4.7 Survey coverage

Most surveys only cover the population living in households but the Census also covers the population living in communal establishments (such as hospitals, schools, prisons and army barracks). People living in communal establishments are slightly less likely to understand Welsh than those in households. Although this difference in coverage is thought to only have a very small effect, it does mean that surveys which only cover households will always appear to have a slightly higher proportion of people understanding Welsh than are recorded the Census.

5 Other factors which may affect estimates on different surveys

There are a number of other factors which should be borne in mind when comparing estimates for Welsh language ability. These include:

- **Sample design.** Surveys that use quota samples are likely to produce slightly different results than those using probability sampling. Ability to speak/read/write/understand Welsh varies according to geographical location, social class and age. If the quota sample does not adequately take account of these factors, it is likely to bias the results. Furthermore, there may be other factors related to Welsh language ability that are not adequately accounted for in the quota sample.
- **Sampling errors.** If the sample size for a particular survey is small, there may be a large margin of error around the estimate. Where sampling errors are large, surveys may appear to produce different estimates, but the differences may not be significant. No sampling errors are currently available for WLLFS estimates for Welsh language skills. However, the sheer size of the survey means these are unlikely to be more than one or two percentage points and will not explain the difference between Census and WLLFS figures on their own.
- **Non-response bias.** If the characteristics of those who respond to a survey are different from those who do not, non-response bias may result. This is more likely in surveys with low response rates and can effect estimates. ONS surveys are routinely weighted to take this non-response into account. Surveys that are not weighted in this way may produce different results.
- **Differences in routing.** Not all surveys ask everyone if they speak/read/write/understand Welsh and instead only route certain people to these questions and then make assumptions about other people. This will affect estimates.
- **Differences in mode.** The mode (face to face interview, telephone interview self-completion questionnaire) in which data are collected can affect the response people give.
- **Differences in edit routines.** In computer-assisted personal interviewing, the interview program often contains a number of checks that will alert interviewers to inconsistencies in their respondent's replies. For example, it may be possible to alert interviewers to check a respondent's answer if they say they speak Welsh but do not understand it. Similarly, self-completion

questionnaires often require editing to deal with inconsistent or implausible responses. In both of these cases, different editing routines may elicit different results.

- **Differences over time.** It is notable that the more recent data produces higher estimates than data that was collected some time ago. The recent and on-going policies promoting the Welsh language are likely to have resulted in estimates for 2000/2001 being higher than those for earlier years

6 Which source of data to use

There is no one preferred source for Welsh language data. There are differences between the Census and the WLLFS estimates for understanding, speaking, reading and writing Welsh but these arise because of differences in the way that the questions are interpreted and it is important to understand these differences when interpreting the data. Respondents with limited ability in Welsh are more likely to say they can understand, speak, read or write Welsh if they have already been asked about their Welsh identity, as on the WLLFS. Where respondents have not been prompted in this way, as on the 2001 Census, respondents with limited ability in these skills will be less likely to say that they have them.

The differences between the Census and survey results are not consistent by age or region. For this reason the results from the two sources should always be treated separately.

7 Conclusions

The WLLFS estimates of the proportion of people with Welsh language skills are considerably higher than the respective Census estimates. For some age groups and regions there is more than 10 percentage points difference in these estimates. The differences between the sources are most marked for young people aged 24 or under.

There is no single explanation for this difference, but it is likely to have been caused by a combination of factors including: survey mode, context of Welsh language questions, the effect of proxies, editing procedures and survey coverage.

Neither source can be considered more accurate, but the WLLFS estimates are likely to be more inclusive of people with limited skills, or who are not confident about those skills, than the 2001 Census.

The WLLFS and the Census estimates should never be mixed because they appear to be measuring slightly different things and the difference between the estimates produced from the two sources vary by age and region.

Sample size means that both the Census and the WLLFS are good indicators of Welsh language skills, especially when tracking changes over time. However, detailed appraisal of Welsh language policy will require more detailed questions.