European Foundation for the Improvement of Living and Working Conditions

First European Quality of Life Survey

Methodological review

Wyattville Road, Loughlinstown, Dublin 18, Ireland. - Tel: (+353 1) 204 31 00 - Fax: 282 42 09 / 282 64 56
email: postmaster@eurofound.europa.eu - website: www.eurofound.europa.eu

Introduction

This report is one of two reports written as technical assistance in preparing the second European quality of life survey (EQLS-2). While this report is a review of the first round of the EQLS in comparison to the European Social Survey (ESS), the second report provides recommendations to improve the data quality. Thereby the 'methodological review' forms the basis for the 'recommendations'.

Both reports are divided into four parts: sampling, fieldwork, questionnaire and data set. This report compares the EQLS-1 with the ESS in these four dimensions, and the 'recommendations' make specific proposals for improving quality in each of these dimensions.

The methodological report is based on previous research (Kohler, 2006a, 2007). This research revealed that the key methodological weakness of EQLS-1 lay in its higher risk of sampling bias, compared to other similar surveys. In a comparison with five other social surveys, including the ESS and the Eurobarometer 62.1, the EQLS 2003 showed the highest rate of failure against so called 'internal criteria of representativeness' and the highest over-representation of women.

Sampling

Sample sizes

Sample sizes must meet two basic needs: estimates of reasonable precision, and ability to calculate estimators separately for groups of substantive interest. The first condition can be reviewed by an evaluation of standard errors. Generally, the higher the sample size the smaller are the standard errors, and the more precise are the estimators. For the second condition one needs to look at the frequency of the subpopulations of substantial interest. The less frequent these subpopulations are, the higher the sample sizes have to be to observe a reasonable number of observations from these subpopulations.

The sample sizes of the EQLS approximate 1000 observations and around 600 in the smaller countries (CY, EE, LU, MT, SI). The actual number of completed interviews ranges from between 591 in Estonia to 1071 in Slovakia. The ESS strives for "effective" sample sizes of around 1500 and 800 for smaller countries, which demands higher sample sizes for the multi-stage samples than for the simple random samples. In the ESS 2002, the actual sample sizes vary between 1207 in Italy and 2919 in Germany[1], and in 2004 between 578 in Iceland and 3036 in the Czech Republic.

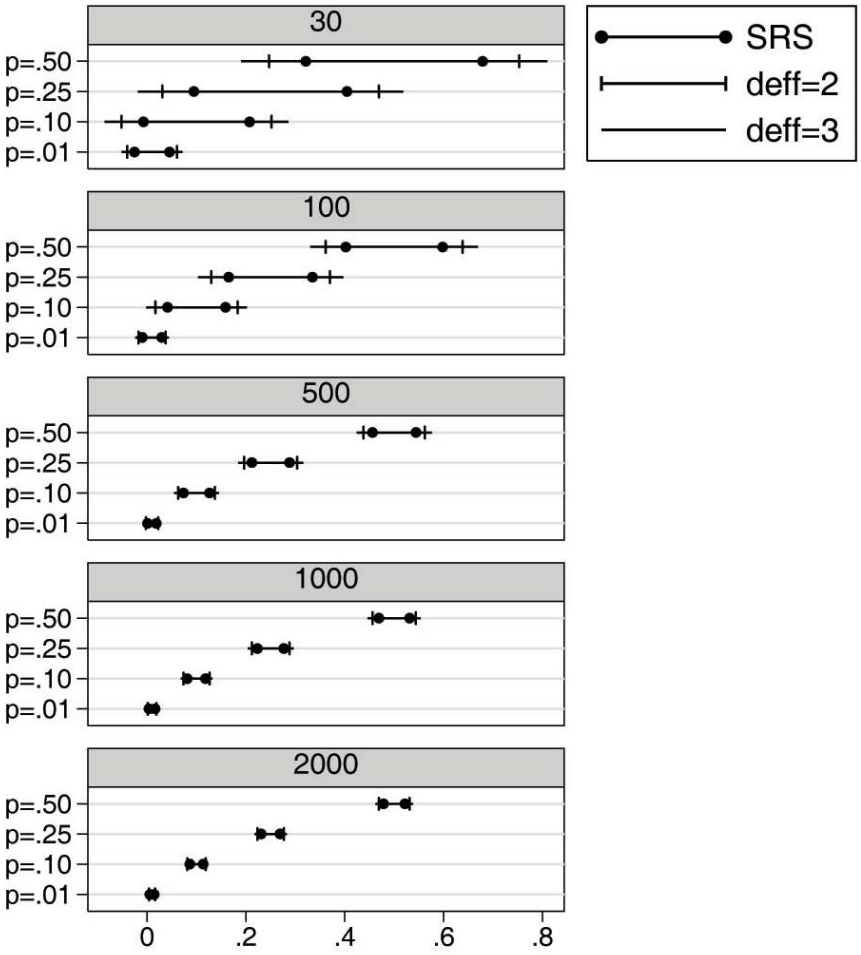
Are the samples sizes of the EQLS large enough to obtain reasonable small standard errors? To answer this question one has to understand that standard errors vary with the sample size, the distribution of the observed characteristic in the target population, and the sampling design. The standard errors are larger for small sample sizes and for characteristics with a high variation; and they are larger for multi-stage samples than for simple random samples. The question must therefore be answered with reference to these features.

The Foundation's publications show a preference for reporting fractions, like the fractions of persons that are satisfied with their lives (Böhnke, 2005), the fractions of young persons living independently (Saraceno et al., 2005), or the fractions of deprived persons (Fahey et al., 2005). Figure 1.1 illustrate how the confidence intervals around such fractions vary with the variance of the observed characteristic in the target population[2], the sample size and the design effect. The confidence intervals are large for small sample sizes, for fractions that are closer to 50 per cent, and for samples with design-effects (i.e. multi-stage samples).

[1] Actually, the German sample is two samples, one for West Germany and one for the East Germany

[2] If the fraction of the observed characteristic is p, the variance is p(1 - p).

Figure 1: *Confidence intervals by observed fraction, sample size and design-effect*



The typical confidence intervals for the EQLS can be found in the fourth graph of figure 1. The width of the confidence interval for fractions of around 50 per cent is around 6 percentage points for SRS, and more than 10 percentage points when there is a design effect of 3. For the smaller countries we are in the third graph, where the corresponding confidence bounds are around 9 and 15 percentage points respectively. For the ESS the typical confidence intervals can be taken from the fifth graph, where these numbers are around four and nine percentage points.

Overall, the difference in precision between sample sizes of 1000 and 2000 is relatively small, but the difference between sample sizes of 500 and 1000 is quite large. One might conclude that the sample sizes of large EQLS-countries are high enough for estimators of the entire sample. There is, however, no justification for less precise estimators in the smaller countries.

For an evaluation of the ability to break down estimators into subpopulations one needs to know the frequencies of these subpopulations. In the past, the Foundation had been particularly interested in the quality of life of poor or marginalised persons. These sub-populations are already rare in the target population, and they are even less frequent in the samples because persons from lower strata tend to be under-represented in populations surveys. Table 1 shows the numbers of observations in the EQLS of various groups.

In the majority of countries the EQLS observed fewer than 100 unemployed persons. The typical standard errors for this subgroup must therefore be taken from the first or second graph of figure 1. For an observed fraction of around 50 per cent within the group of unemployed persons, the confidence interval will have a width of 20 up to 60 percentage points, depending on the exact numbers of observations and on our assumptions about the design effect. The precision of estimators calculated for unemployed persons must be considered very low. The situation is a bit better for unskilled workers and the low educated. It is practically impossible to gain reliable estimates for persons living in single parent households.

Table 1: *Number of observations of selected subgroups in the EQLS*

	Unemployed	Unskilled	Low education	Single parent		Unemployed	Unskilled	Low education	Single parent
AT	35	146	160	44	**IE**	67	184	209	58
BE	96	108	172	57	**IT**	28	70	311	15
BG	131	180	208	10	**LT**	113	159	161	50
CY	7	30	155	9	**LU**	8	42	144	19
CZ	44	87	105	27	**LV**	86	134	120	37
DE	71	116	85	26	**MT**	10	42	146	1
DK	52	180	61	46	**NL**	42	76	140	44
EE	50	90	37	16	**PL**	145	87	195	21
ES	52	136	419	20	**PT**	69	207	571	25
FI	73	87	208	51	**RO**	48	97	278	19
FR	56	43	174	38	**SE**	25	91	106	49
GB	73	206	290	45	**SI**	38	57	111	10
GR	44	80	433	11	**SK**	96	127	128	18
HU	58	235	278	22	**TR**	79	71	491	25

Increasing the sample size will improve the precision of all estimators, including those calculated for rare subpopulations. However, the sample size must be markedly increased to gain substantial increases in the precision of estimators for these rare subpopulations. To get one additional observation from a rare subpopulation, one has to raise the sample size by 1/p, with p being the fraction of persons from a subpopulation in the target population. For example, it will take a sample size of 1481 observations to get 100 unemployed persons in Germany. In Italy this would require a sample size of 3586, and in Luxembourg of over 7000.

Stratification

Stratified samples will be cheaper than Simple Random Sampling (SRS), if the survey costs vary between strata. They will have smaller standard errors than SRS, if the variation of the observed characteristics differ between strata. Stratified samples allow calculation of estimators for each strata separately, and they can be used to oversample rare strata. However, stratification will only be sensible if suitable sampling frames exist for each stratum (Schnell et al., 279f).

The EQLS used stratified sampling throughout (Ahrendt, 2003, 3). Strata were defined by “region and urbanisation”. It is not documented how regions and degrees of urbanisation were categorised. There is also no information on the sampling frames used for the stratification.

Stratification was also used in most countries in the first two rounds of the ESS. Exceptions are the countries that applied SRS (DK, EE, FI, SE, SK), as well as Ireland in ESS-2, and the Netherlands. All other countries have used stratified samples, whereby the definition of strata vary between countries. Stratification by region and/or urbanisation was used in most countries. More specific stratification rules were used in Luxembourg, Slovenia, and in the Czech Republic (ESS-2). Luxembourg stratified by socio-economic characteristics based on its social security register. Slovenia used regions and types of settlements, based on information provided by Slovenia's Central population register. The Czech Republic stratified communities according to their numbers of streets (European Social Survey, 2004a, 2006)

The ESS samples differ considerably in the number of strata being used. Germany applied a micro-stratification with more than 1500 strata defined by region and population size. Sampling points have been selected from within these so called "layers" with a probability proportional to size. This had the effect that some strata end up with no sampling points, while the large towns contain more than one. Poland and Ireland (ESS-1), on the other hand, only used two strata. Poland stratified communities into towns above and below 50.000 inhabitants. Ireland drew samples for the north-west and the south-east of the country separately.

The advantages of stratified samples will be of importance only if target persons within strata are similar, while persons in different strata are dissimilar. Evidently the positive effect of stratification therefore depends on the specific characteristic under observation. For the stratification used in the EQLS, characteristics that vary with region and urbanisation are likely to profit from the stratification. So far, there is little evidence that many of the characteristics observed by the EQLS have this property. The advantage of stratified sampling must be considered as small. The stratification the ESS used for Luxembourg is more promising, although it is unlikely to find suitable sampling frames to implement it in other countries.

There is also no evidence that the stratification by region and urbanisation causes many disadvantages. Disadvantages might arrive, if the stratification were done based on invalid or outdated information. Information for regional stratification is relatively easy to access and gets regularly updated by statistical offices. Disadvantages from stratification by region and urbanization are not expected. To verify this claim a variable containing stratum identifiers is needed. Unfortunately, such variables are neither available for the EQLS, nor for the ESS.

Clustering

Multi-stage samples tend to have a so called design-effect, which means inflated standard errors of estimators as a result of the homogeneity within the primary sampling units ("clustering"). Generally, the more homogenous the primary sampling units are, and the more elements belong to one primary sampling unit, the larger the standard error gets. It is not uncommon that the standard error of multi-stage samples are more than 1.4 times higher than those of simple random samples. It would require sample sizes of two times the size of a similar SRS sample to achieve the same precision of estimators as with an SRS sample (Kish, 1965).

Apart from these disadvantages, multi-stage samples are often indispensable because of the lack of sampling frames for the entire target population. Furthermore, multi-stage samples reduce interviewers' travelling costs.

The EQLS used multi-stage sampling in all countries. For 18 of the 28 countries this is a consequence of the random-route technique, which implies at least a three stage design: first sampling of geographical areas (sampling points), secondly an application of random route, and thirdly a selection of persons at the identified addresses. It is not

documented whether the sample points within which random route was applied were selected by more than one step. At least two stages were used in Finland, Ireland, Italy, Sweden, Czech Republic, Estonia, Hungary, Latvia, Poland, and Rumania. These countries seem to have selected the target persons directly from a frame of individuals.[3]

The EQLS does not document the definition of the sampling units at each selection stage. The number of sampling points and the selection mode within a household is the only specific information about the application of multi-stage sampling. The number of sample points varies between 56 in Sweden and 379 in Finland. Selection of persons within household is either done by the "last/next birthday rule", or by a variant of the Kish selection grid (Ahrendt, 2003, 4).

Table 2: M*ulti-stage sampling in the ESS*

Country	Rounds	Primary sampling unit	Secondary sampling unit	Tertiary sampling unit
Simple Random Sampling				
DK	1,2			
FI	1,2			
EE	2			
HU	2	ESSes over 80.000)		
PL	1,2	(in cities over 80.000)		
SK	2			
SE	1,2			
Two stages				
BE	1,2	Clusters		
CZ	1	Households		
DE	1,2	Communities		
HU	1,2	Settlements		
LU	1,2	Tax units		
NL	1,2	Postal delivery points		
PL	1,2	Towns/Villages		
SI	1,2	Clusters of Enumeration Areas		
ES	2	Electoral sections		
Three stages				
AT	1,2	Clusters	Households	
FR	1,2	Communities	Households	
GR	1,2	Area Units	Households	
IE	1,2	Agg. district electoral divisions	Addresses	
PT	2	Administrative areas	Addresses	
ES	1	Electoral sections	Households	
GB	1,2	Postcode sectors	Electoral sections	
Four stages				
CZ	2	Towns	Streets	Dwellings
IT	1	Municipality	Electoral precincts	Addresses
PT	1	Localities	? ("PSU")	Addresses

[3] The field work technical report is not very clear here.

Table 3: *Kish's design effects for ESS 2004 using interviewers as cluster information*

Country	Gender	Life satisfaction	Income	Country	Gender	Life satisfaction	Income
AT	1.1	1.6	1.8	**HU**	1.2	1.9	2.2
BE	0.9	1.2	1.5	**IE**	0.9	3.4	6.1
CZ	1.4	2.4	3.9	**LU**	1.3	3.6	3.0
DE	1.0	1.8	3.7	**NL**	0.8	1.0	1.3
DK	0.8	1.4	1.5	**PL**	0.8	1.4	1.8
EE	1.2	3.3	4.2	**PT**	1.4	10.2	9.3
ES	0.8	2.1	6.3	**SE**	1.1	1.2	1.6
FI	0.7	1.0	1.8	**SI**	1.2	1.0	1.4
GB	0.8	1.3	1.7	**SK**	0.8	1.6	2.1
GR	1.8	2.7	7.6				

The ESS did not used multi-stage sampling in all countries. Instead, SRS was used in Denmark, Finland and Sweden, and in 2004 also in Estonia and Slovakia. SRS was also used for large cities in Hungary and Poland. All other countries applied variants of multi-stage sampling. Two stages were used in nine cases, three stages in seven, and finally four stages in three cases (table 2). Thereby, the specific regulations vary considerably between countries. Geographical units, i.e. communities or electoral districts were used as primary sampling units (PSUs) in most cases, but there are also examples of households or the like (CZ in 2002, LU, NL).

The consequences of the multi-stage sampling design for the precision of estimators depend on the homogeneity of the elements within the sampling units. To be more concrete it is necessary to calculate the design-effects, which requires that identifiers for the sampling points are included in the data set. Unfortunately this requirement is neither met by the EQLS nor by the ESS. For the ESS 2004 it is possible to be a little more precise by using the interviewer numbers as a proxy variable for the sampling points. This is possible because fieldwork institutes often assign all target persons of a sampling point to just one interviewer. The downside of such a procedure is that interviewers are a source of design-effects themselves (Schnell and Kreuter, 2005). It is therefore unclear how far the calculated design-effects are due to geographical clustering, or due to interviewer effects.

Table 3 shows Kish's design effects (Kish, 1965). The numbers quantify how much larger the sample size of a multi-stage sample should be to get estimates of the same precision as a comparable SRS. The numbers in the table indicate that design effects are larger for life satisfaction than for gender, and even larger for income. In addition, design effects tend to be larger for the three and more stage samples than for the SRS samples[4] and the two stage samples. Especially the design effects for Greece and Portugal are so large that one is inclined to attribute them to interviewer misbehaviour, either in the selection of target persons, or in conducting the survey.

Unfortunately it is not possible to calculate design effects for the EQLS. Given that most countries used at least three stage sampling, one might expect high design effects for the EQLS. Moreover, as one can assume that the primary sampling units are geographical areas, one can expect high design effects for income (and correlated characteristics) in all countries with strong regional disparities.

[4] The design effects of the SRS samples must be fully attributed to the interviewers.

Sampling frames and random route

The population of available sampling frames is usually not equal to their target population, i.e. the target-population will be under- and/or over-represented by the frame population. Consequently, selecting elements from such frames will lead to problems of under- or over-coverage of the sample. Information about the sampling frame is absolutely necessary to evaluate the sampling quality.

Table 4 compiles information about available sampling frames for the selection of households, addresses, dwellings or individuals.[5] Most of this information stems from the methodological documentation of the ESS, which has regularly used population or address registers from administrative bodies as sampling frames. However, some noteworthy exceptions exist. In two countries, United Kingdom and the Netherlands, addresses were drawn from a list of postal delivery points, provided by the postal services of the respective country. In Austria the telephone book was used for the selection of addresses, which was then combined with a subsequent random address. A merged list of subscribers for electricity, gas, radio, television or telephone was used for the Czech Rep. in ESS-1.

The ESS also provides some remarks about the coverage of the sampling frames being used. Given these remarks, the documentation provides several problematic aspects. The target population of the ESS is all persons aged 15 and above resident within private households, regardless of their nationality, citizenship, language, or legal status in the participating countries. In this respect the population registers like those from Sweden or Slovenia show over-coverage of the persons living in institutions. However, this over-coverage can be easily corrected during field work. More problematic is the under-coverage of persons that are not entitled to vote in the electoral registers used for Ireland and Italy. Albeit the sampling frame is only used to draw addresses, persons not entitled to vote will only be included in the sample, if they live at an address of someone, who is entitled to vote.

Systematic losses because of under-coverage cannot be ruled out even in countries with more reliable sampling frames. This might be a reason to abstain from the selection of households from a frame by applying random route. However, the field work institute of the EQLS also gives little attention to the documentation of sampling frames for countries, where a frame has been used. For several countries the fieldwork technical report mentions that a “database” has been used as sampling frame, which tells little. Some countries used the same frame as the ESS. Sweden used the so-called “GfK master sample”.

[5] Frames for the selection of geographical units are considered as unproblematic.

Table 4: *Sampling frames of the EQLS and ESS*

	Frame in ESS	Remarks	EQLS
AT	Telephone book	Yearly updated. Covers 90% of households.	None
BE	Population register	"Excellent coverage"	Database
CZ	SIPO Address register	Used in ESS 1. Merged subscribers to electricity, gas, radio, television or telephone. Coverage about 98%. Used in ESS 2. All buildings that have a house number. Free access via Internet.	
DE	Population registers	Only available for communities.	None
DK	Population register	99.9% coverage	None
EE	Population register		Ditto
ES	Continuous census	Compiled from compulsory notifications of relocations at the municipalities. Notification allows people to access public services.	None
FI	Population register		Ditto
GB	Address register	Address database for Royal Mail, containing all known UK postal addresses and their associated Postcodes and Delivery Point Suffix information. Daily updated by Royal Mail. Coverage around 97% of individuals.	None
GR	Census	Only used for PSU selection. Address random within PSU.	None
HU	Population register	Provided by the Central register and election office. Less accurate in large cities	Ditto
IE	Electoral register	Yearly updated register of electors. Only used for selection of addresses.	Ditto
IT	Electoral register		Ditto
LU	Household register	List of "tax units", i.e. one or more persons treated collectively for tax purposes. Covers 91% of resident adults. Exceptions being EU civil servants and employees of international organisations and foreign banks.	None
LV	-		Database
NL	Address register	Address list provided by the Dutch postal service	None
PL	Population register	Logs of the Ministry of Internal Affairs and Administration. Does not cover institutional population, foreigners working on black market, homeless, and people temporarily not available.	Ditto
RO	-		Electoral register
SE	Population register	Register of all persons living in Sweden. Includes persons living in institutions	GfK Mastersample
SI	Population register	Covers 99% of residents with permanent address. Institutionalised persons are included with their permanent addresses, but are unlikely to be reached.	None
SK	Population register	Covers 99.99% of residents. Does not cover homeless persons and those living in institutions.	None

As indicated in table 4, most EQLS-countries have not used a frame for the selection of households. These countries used random route, and this is also true for the countries not listed in the table. Random route can be used to approximate a random sample if no sampling frame is available. The extent to which random route approximates a true random sample depends on the definition of the rules for the random walk, and even more crucially on the control of the interviewers by the field work institute.

The EQLS used standard random for the selection of addresses. Within each sampling point a maximum of 20 addresses had to be selected. The interviewer had to record each contact and the outcome of each visit on a contact sheet. Each address had to be contacted up to three times on different times of a day. "Substitutions were not allowed" (Ahrendt, 2003, 3).

The disadvantages of random route are well known and often described: the directions of the random walk can be ambiguous making it difficult to replicate the ascertainment of addresses. This makes it difficult to control the selection process, and opens the possibility for interviewers to violate the rules of the random route to ease their work (Schnell et al., 286). Results shown elsewhere (Kohler, 2006a, 2007) indicate that the sample quality of the EQLS suffers from problems caused by interviewer failure in the selection of target persons.

Address random as it was done for the ESS samples in France and Austria could decrease the problems of random route. The ESS also demonstrated yet another way to overcome the problems of sampling without sampling frames. In Greece, city blocks have been selected from an electronic map at the first stage. A complete list of households was then compiled for each of these relatively small sampling points. The field work institute randomly selected the target households from these self made sampling frames. Such a procedure however requires good electronic maps as frame for the selection of sample points.

In the field of survey sampling it seems better to use sampling frames rather than random route because with the former one at least has information on the amount of under-coverage. With random route one hardly knows what's going on, and, in fact, what is going on could undermine the whole sample.

Unequal sampling probabilities

In simple random sampling all elements are drawn with the same known sampling probability. Multistage samples often deviate from this. If the sampling points differ in size, they must be either drawn with a probability proportional to their size ("PPS"), or the elements within the sampling points must be drawn with PPS. Otherwise elements from large sample points will be under-represented. The same happens, if one person is selected from households of different sizes; persons from large households are under-represented in the sample.

The under- or over-representation from the sampling design does not pose a principal problem. To deal with them, each observation has to be weighted by the reciprocal value of its sampling probability. Hence, all sampled persons that live in a household with two persons (from the target population) must be weighted by $\left(\frac{1}{2}\right)^{-1} = 2$, and those from three person households by $\left(\frac{1}{3}\right)^{-1} = 3$

Likewise, for multi-stage sampling without PPS the weighting for the observations is $\left(\frac{N_s}{N}\right)^{-1}$

with N being the population size and NS the population size in the sampling point.

Unit non-response is another source of unequal sampling probabilities. It arises from target persons, who are not at home at any visit, who are unable to answer, or who refuse to answer. One can say that not-at-homes, unable to answers and refusals have a lower sampling probability than other persons. However, unlike disproportional sampling probabilities stemming from the sampling design, the unequal sampling probabilities that arise from unit non-response are not known. It is common practice to perform a so-called redressment, meaning to weight a sample such that certain socio-demographic variables are as frequent as in the target population. Substantially, the method assumes that the population is divided into homogenous subgroups that can be measured by socio-demographic variables, and that the non-response

within these subgroups is completely at random. Empirical studies of these assumptions have shown that they are not realistic (Schnell, 1993). Technically, redressment requires that the data-set contains some variables for which the true distribution is known from external sources (also see p 17).

The EQLS used multi-stage sampling throughout. Albeit it is not documented, it is very likely that the sampling point selected on the first stage do not have equal size. We assume here that the fieldwork institute selected the sampling points with PPS; no further weighting is then necessary.

In the majority of countries, the EQLS selected one target person from households of different sizes. Sampling probabilities therefore vary with household size. Apparently the weighting variable included in the EQLS data set does not account for these unequal sampling probabilities. Albeit the fieldwork technical report does not document the construction of the weighting variable, the variables vary with age and gender, but not with the number of persons above 18 living together in a household. Hence, the weighting variable of the EQLS only provides redressment. It should be combined with a weighting variable that accounts for different household sizes, which can be easily constructed from the household grid. The problematic side of the combination, however, is that the redressment weights should be calculated after having applied the design weights.

The ESS also applied PPS for the multi-stage samples. To account for unequal sampling probabilities stemming from the selection of target persons within households a design-weight is included in the data set. This design weight also accounts for unequal sampling probabilities of West and East Germany. Redressment was not done for the ESS.

Summary

The sampling of the EQLS and the ESS followed different strategies. The Foundation proposed random-route technique in the tender, and asked for better proposals by the fieldwork agencies. Hence, they asked for what can be considered as the minimum acceptable sampling design, but allowed better techniques. This had the effect that the strict probability sampling was applied in only 10 of the 28 countries, primarily in Eastern Europe. Given the experience of the ESS it can be shown that strict probability sampling was not used in several countries, where suitable sampling frames exist. Hence, the request for “better alternatives” in the tender was not entirely successful.

The ESS starts with the preposition that sample designs may be chosen flexibly and there is no need for similarity of sample designs (Kish 1994, 173). They seek to find the best available sampling design in each country, whereby the main criteria are known probabilities of selection for all population elements. Deviations from strict probability sampling must be justified. The strategy of the ESS had the effect that random route was used only in very few exceptional cases.

The major problems of the Foundation’s sampling strategy are

- higher standard errors than necessary,
- little knowledge about over- or under-coverage problems,
- and little ability to control the selection of respondents by the interviewer.

Together these problems might have led to the large sampling biases described elsewhere (Kohler, 2007). Some of the problems might be dealt with by changes in the field work, or by a better documentation of the sampling process. However, many of the problems are built in, and can only be diminished by use of better sampling frames. This however will require a flexible approach to sampling, and the ESS might act as a model here.

The major problem of the ESS sampling strategy is that it is difficult to implement from a perspective of a central survey administration; this is especially true for the evaluation of sampling frames. The ESS strategy must assume a certain level of trust in the local study administrators. In the case of the ESS this trust might be legitimate as the local organisers share the interest of the central organisers. For the EQLS the local partners are commercial field work organisations, whose first priorities may not be low standard errors. If the Foundation decides to go in the direction of the ESS sampling strategy they need to implement control mechanisms. With respect to the sampling design the Foundation could set up an advisory committee that evaluates the sampling designs of each local field work agency before it is actually used. A similar procedure was also used by the ESS.

Fieldwork

This section examines instructions for the field work that may have affected data quality. We distinguish between instructions that influence the response rates and instructions that influence the data quality.

Instructions that affect the response rates

High response rates are often seen as a sign of good data quality, because non-systematic sampling biases should be more likely when there is high unit non-response. It is therefore sensible to implement specific instruments to maximise the response rates. Some of these instruments are discussed here and how far they have been implemented in the EQLS and the ESS.

Inflating response rates is a means to hide low response rates. The general rule to calculate a response rate is

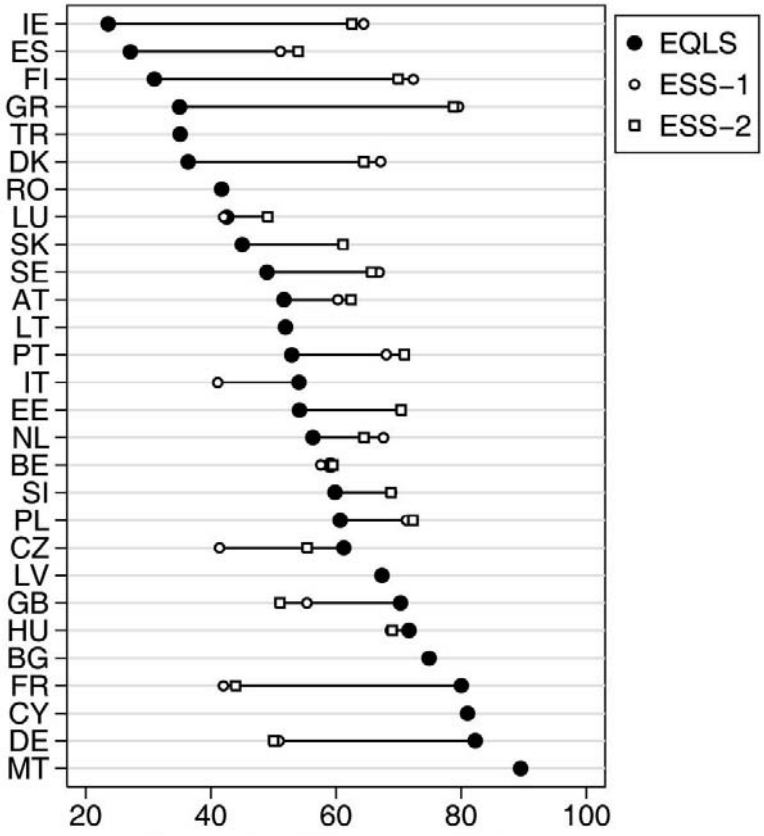
$$\frac{\#\text{ of achieved interviews}}{\#\text{ of issued sample units} - \#\text{ of non eligible sampling units}}, \qquad (2.1)$$

but the applications of this formula differ in their definition of the non-eligible sampling units. Obviously, the more sampling units are categorised as "non-eligible" the higher the response rates. If sampling units that belong to the target population are termed "non-eligible", the response rates are "inflated", and they lose their value as a means to assess the sample quality.

EQLS counted all sampling units living in household that could not been contacted after three visits or due to language problems as "non-eligible". As these sampling units definitely belong to the target population, the EQLS response rates must be considered as inflated.[6] In the ESS the definition of non-eligible sampling units is much narrower. They are defined by (1) addresses that are not residential (institutions, business buildings), (2) not occupied, (3) other non-eligible addresses, and by (4) target respondents that moved abroad or (5) have deceased. Here, only the "other non-eligible addresses" are probably occupied by persons that do in fact belong to the target population. The actual numbers of sampling units belonging to that category are below 10 in most countries. One might be worry about response-rate inflation only in France (108), Germany (100), Ireland (173), Slovenia (74) and Switzerland (165).

[6] The numbers calculated by the EQLS are sometimes called 'cooperation rates'.

Figure 2: *Harmonised Response Rates of EQLS and ESS by country*



A recalculation of the response rates using a stricter definition decreases the average response-rate from 59 per cent (Ahrendt, 2003, 2) to 55 percent. Figure 2 shows that these harmonised response rates are considerably lower than those of the ESS in most countries. The EQLS response rates are higher than in the ESS in only five countries: Italy, Czech Rep., United Kingdom, France and Germany. In three of these countries, Germany, France, and United Kingdom, the number of households that could not be contacted is said to be zero. As these numbers are fairly unrealistic, they might indicate a further inflation of the response rate stemming from substitutions during the sampling (see p.8).

A higher number of call-backs can reduce the number of no contacts considerably. It is, however, necessary that call-backs are made at different times and/or week days. In comparative survey projects the number of call-backs might vary from country to country, as ease of contact differs. For example, according to (Michel and Jaak, no year), the number of contacts in the United Kingdom, Switzerland, Spain and Portugal should be higher than in other countries. The EQLS prescribed three visits 'at different times of the day' (Ahrendt, 2003, 3), including weekends. Due to the application of standard random route there is a risk that addresses, where nobody could be contacted on the first visit, are not contacted a second time but substituted immediately by a nearby contact-able address. While the central organisation of the ESS prescribed four visits at the target address, several countries actually prescribed five or even seven. Moreover ESS-countries vary in whether they have prescribed visits at the evening and/or on weekends.

Advance letters are used to inform target persons that they have been selected for a survey, and that they will be contacted by an interviewer in due time. Advance letters are seen as a means to reduce the number of refusals. The EQLS used letters that were put in the postbox in cases where the household could not be contacted. The ESS sent advance letters and a brochure about the contents of the survey to all sampled persons or household in most countries.

Incentives are used to reduce respondents' expected costs of participation, which might in turn reduce the refusal rate. Incentives might be little presents, cash, lottery tickets etc. It has been shown that even modest incentives help to improve the response rate. Experiments with incentives show that cash or flexible incentives have the highest effect on survey participation. The EQLS did not use any incentives, while the ESS used little gifts, lottery tickets or cash in most of the Western European countries.

Converting refusals

Non-response studies have shown that refusals are often caused by the circumstances and the mood of the potential respondent at the time of the initial survey request. A repeated attempt to persuade initially reluctant persons to reconsider the survey request can be quite successful. The second attempt must be done by an experienced interviewer. The documentation of the EQLS does not indicate any operations to convert refusals, although that possibilty is mentioned in the project instructions. The ESS generally recommends to apply refusal conversions, and the majority of countries followed that recommendation. The precise strategy is documented in the ESS documentation reports.

Refusal avoidance

It has been shown that more experienced interviewers tend to achieve higher response rates than those with less experience, and that interviewers who are confident about their ability to elicit cooperation tend to achieve higher response rates. One way to avoid refusals therefore is to use experienced interviewers, and/or to spend some effort in briefing the interviewers for the study. Only interviewers with at least one year of experience were used for the EQLS; the average experience of the field force was five years. Interviewer training is said to be 'thorough'. The ESS recommends to select experienced interviewers as much as possible. It prescribes a day or half day of personal briefing sessions of all interviewers, covering all aspects of the field procedures, including a session on doorstep introduction and discussions on encouraging participation.[7] The documentation of ESS includes the numbers of interviewers that have participated in the briefing section. Finally it recommends low assignment rates for interviewers, and hourly pay rates for interviewers; the latter was so far only applied in five countries, however (FI, NO, NL 2004, SE, CH).

Instructions that affect the data quality

The quality of the data achieved with a population survey crucially depend on the reliability and validity of the instruments (questions) being used. However, it is not the questions alone that affect data quality; especially the interviewers are important in this respect. The way interviewers read out questions, the way they react to answers, and even the way they look can have an effect on the measured values; interviewers might even cheat when completing questionnaires. Fieldwork instructions can help to reduce such interviewer effects substantially.

Personal briefing of interviewers is much more effective than written specifications in harmonising interviewers' behaviour. The EQLS has not documented whether there were any personal briefing sessions. However, each EQLS interviewer received 'project instructions' with extensive advice on each step of their work. The project instructions also include further comments on many survey questions. The ESS prescribes personal briefing sessions of at least half a day's length. It documents how long the actual briefing sections were, how many interviewers participated and what topics have been discussed.

[7] In Germany, 26 Interviewers participated in an "Refusal Avoidance Training" of 2 hours length. The training reduced the number of reported refusals by nearly 7%, but increased the number of non-contacts (29).

Back-checks involve re-contacting a target person to find out whether he was really interviewed, whether the interviewers have behaved well, and probably whether the interviewer recorded answers correctly. Fieldwork institutes regularly perform back-checks to evaluate the quality of their interviewers. Intensive and regular back checking is likely to prevent interviewers from cheating, or working inaccurately. The EQLS does not document precise back-checking regularities. According to the tender document there is a control of at least 10% of each interviewer's work by telephone call-backs. It should be noted that back-checks cannot control the selection of respondents for the random route samples. The ESS reports that it has performed back-checks in most participating countries. It documents how the back-checks have been performed, and what results they produced.

Payments could be made on an hourly pay rate or per completed interview. The latter gives an incentive to work rather quickly, potentially encouraging interviewers to fill out parts of the questionnaire on their own. Hourly pay rates can increase the overall cost of a survey project, especially in countries with high proportions of potential unit non-response. Payments per interview combined with bonus payments depending on the outcomes of the back-checking procedures might be a compromise between the two. It is, however, unlikely that fieldwork institutes are willing to deviate from their common payment structure for one single academic survey. The payment structure of the EQLS is not documented. The ESS used payment on a per interview basis in all but five (2002: four) countries. Bonus payments for achieving high response rates have been applied in a number of countries.

Interviewer identification numbers that are included in the data set might reduce the number of faked interviews. Letting interviewers know that their identification numbers are stored in the data set might increase their commitment not to cheat. Neither the EQLS nor the ESS-1 includes the interviewer numbers in the data set. Only the ESS-2 does it.

Summary

Overall, the comparison of the fieldwork instructions between the EQLS and the ESS show the following differences:

1. The ESS put more effort than the EQLS into fieldwork instructions likely to reduce unit non-response than the EQLS.
2. The ESS invested more into mechanisms to control the work of the interviewers than the EQLS.
3. The ESS documents its field work instructions more completely than the EQLS.

The efforts spent for reducing non-response and controlling the interviewers cost money as well as time. With a fixed budget there should be some flexibility for the negotiation of the specific fieldwork instructions. It might therefore be sensible to ask the tender applicants for a thorough documentation of the intended field work instructions. With this information the Foundation might get an idea which fieldwork agency offers the "best" instructions. It also increases the means to control the work of fieldwork agencies, which is absolutely necessary.

Questionnaire

Background variables

Background variables serve multiple purposes in a social survey. They are used to describe the social structure of societies, to match treatment and control groups in causal analyses, to assess the quality of a realised sample, and they form the basis for the construction of weights. In order to fulfil all these tasks, background variables of comparative surveys must have two properties:

- They must measure the same theoretical concept in each country.
- They must measure a concept that is comparable with other sources available for the countries.

Two strategies could lead to background variables that have these properties. One can either ask the same questions (ASQ) or equivalent, but not identical questions (Braun and Mohler, 2003, 104). Which of the two is better is different for each background variable.

Household and household composition

Households are defined as the smallest "units of persons living together and reacting as cells of society. Households ... act in society like individuals" (Bien and Quellenberg, 2003, 279). In this definition, living in the same dwelling unit, probably with a shared kitchen, and some common living arrangements are the central elements. Unfortunately, the use of the term "household" in common speech is not nearly as clearly defined. 'Household' is sometimes used as a synonym for families, for persons that have meals together, for persons living in one dwelling, etc. The uses of the term household differ between countries, and also between social groups within countries. Questions that ask for the size and composition of household therefore need to make clear what sort of "units" they have in mind.

The EQLS starts the questionnaire with a question on household size: "Including yourself, can you please tell me how many people live in this household?". A clear and explicit definition of the term household is given in the project instructions for the Interviewer. In these instructions the household is defined as:

> *One person living alone or a group of people living at the same address in a non-institutional dwelling, and having that address as their only or main residence, who either share at least one main meal a day or share the living accommodation (or both).*
>
> *Included are: people away on holiday, away working or in a hospital for less than 6 months; school-age children at boarding school; students in household sharing private accommodation.*
>
> *Excluded are: people who have been absent for 6 months or more, students away at university or college; temporary visitors.*

The ESS asks questions on the household composition in the last section of the questionnaire. After a short transition, the first question on household was: "Including yourself, how many people-including children-live here regularly as members of this household?". Here, some idea of the definition of a household is already mentioned in the question. Empirically, the ESS records more persons above age 16 in the household grid, while there is little difference in the number of persons below age 16.

Income

Income is defined as "receipts of an economically active unit (person, community, co-operative) during a specific period (day, week, month, year)" (Warner and Hoffmeyer-Zlotnik, 2003, 308). This includes wages from employment and

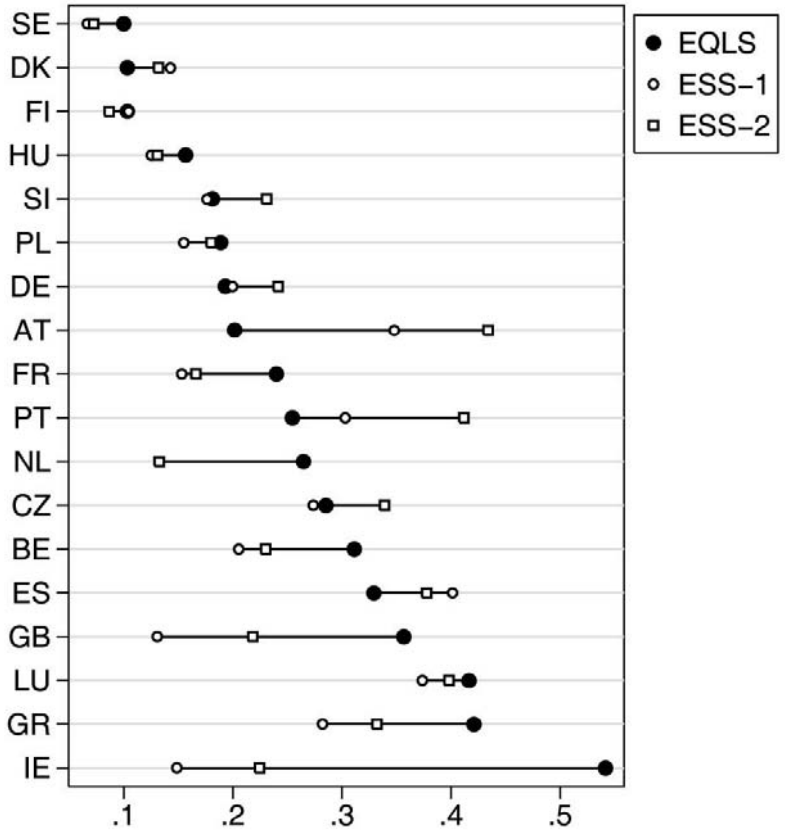
business, returns from property and capital, as well as public or private transfers. Household income is the sum of all income of persons who live in the same household (see p.17), including the transfers provided for the household. Net income is income after taxes and obligatory contributions to the social security system have been deducted. Disposable income is net income reduced by certain fixed costs to measure the opportunities of households to consume the receipts. The operationalisation of income requires a decision about what sort of income one is interested in. As the EQLS asked for household's net income, we will only deal with the operationalisation of this income type here.

The measurement of (household) income with a survey has to overcome three major difficulties:

- Questioning about income is a sensitive issue. Therefore the item non-response is generally high.
- Income can have different sources, and people tend to forget one income source or another.
- Deducting taxes is a very difficult task, especially because the tax due is usually calculated from yearly income, while wage earnings tend to be paid on a weekly or monthly base.

The severity of these problems varies with social strata and country. It is therefore difficult to say what the "best" survey instrument for measuring income might be. Surveys that are primarily concerned with income generally ask for each possible source of income in turn. Surveys like the EQLS and the ESS use income as background, so that there is little room to ask many questions on income. Both the EQLS and the ESS therefore used the same strategy to measure income. They first ask for the sources of income, which somehow alerts respondents to the possible income sources. Afterwards they ask the respondents to add up all the income sources, to deduct the taxes, and to give an approximation of the weekly, monthly or yearly net income. The difference of the EQLS and ESS is that the latter uses a card for the different income sources, and that the EQLS uses 19, the ESS only 12 income intervals. Both operationalisations try to give the respondent the feeling that the interviewer cannot encode the substantial meaning of the given answer.

Figure 3: *Item non-response on household income in EQLS, ESS-1 and ESS-2*



Give the similarity of instruments, we do not expect large differences in the amount of missing values between the ESS and the EQLS. An empirical check largely verifies this claim (figure 3.1). The EQLS has a higher proportion of item non-response in some countries, and lower proportions in other countries. On average, the EQLS lost 26 per cent of observations on the income question, the ESS-2 24 per cent, and the ESS-1 21 per cent.

Education

In empirical analyses education is either used as a measure of achieved skills, or as a measure for formal education (certificates). As it stands, two major international categorical systems for the classification of educational attainment have been developed. One is the so called CASMIN scheme, developed in the early 1970s by the research project "Comparative Analysis of Social Mobility in Industrial Nations" (Brauns et al., 2003). The other is the International Standard Classification of Education (ISCED), also developed in the early 1970s by UNESCO, and updated in 1997 (UNESCO, 2003). Both classification schemes are not meant as an instrument to be included in a survey questionnaire, but as a "methodology that translates national educational programmes into an internationally comparable set of categories" (UNESCO, 2003, 198). The most important consequence of this is that education must not be ascertained by asking the same question (ASQ) in each country. Instead, each country must ascertain the educational programs that exist in their country with a suitable instrument. The resulting data will be then recoded to the corresponding classification.

The disadvantage of both classification schemes is that they require an extra step of harmonisation. The development of classification rules for each country is a serious scientific enterprise that requires in-depth knowledge of each country's educational system. Fortunately standard rules have already been developed, and especially the ISCED rules should be available for each country of the EU. It should be noted, however, that these rules are not without problems.

The advantages of using country-specific instruments for education are:

- It is easier for respondents to name the educational certificate they have received, than to evaluate whether this certificate is of a lower or higher degree.
- Using country-specific instruments allows to construct both the CASMIN and the ISCED classification.
- Using country-specific instruments is open to new developments of international classification schemes.
- Using country-specific instruments allows comparison with national marginals, and the construction of weights for redressment.

The EQLS used an ASQ-approach to ascertain data on education. Two questions were asked, one for the terminal education age, the other for a self-assessment of the level of education. During the data cleaning process it was found that the latter had serious problems of validity; it was therefore recommended not to use it (Nauenburg and Mertel, 2004, 8). Therefore terminal education age is the only usable information on education in the EQLS. Unfortunately, the usefulness of these variable for within country analysis and cross country comparison is rather weak (Hoffmeyer-Zlotnik and Wolf, 2003, 248).

In contrast to the EQLS, the ESS employed a country-specific approach to ascertain education (European Social Survey, 2004b, 44). It thereby applies country specific names for the certificates[8]. In addition it asked a second question on the years spent in education. The answers on the country-specific questions are recoded to a modified version of ISCED-97. As the ESS data contain the country-specific variables and the ISCED-97 it is possible to use it as a model for the EQLS.

[8] Country specific show cards with the respective certification names are available on the Internet.

Table 5: *Categories of employment status in EQLS and ESS*

EQLS	ESS
At work as employee or employer/self-employed	In paid work (or away temporarily) (employee, self-employed, working for your family business)
Employed, on child-care leave or other leave	
At work as relative assisting on family farm or business	
Unemployed less than 12 month Unemployed 12 month or more	Unemployed and actively looking for a job Unemployed, wanting a job but not actively looking for a job
Unable to work due to long-term illness or disability	Permanently sick or disabled
Retired	Retired
Full time home-maker/responsible for ordinary shopping and looking after the home	Doing housework, looking after children or other persons
In education (at school, university, etc.)/student	In education, even if on vacation (not paid for by employer)
	In community or military service
Other	Other

Employment status

Practically every social survey contains one or more questions on employment status. Although all these questions somehow ask for the respondent's connectedness to working life, there is no generally accepted operationalisation of employment status.

Table 5 shows the categories used to classify respondents' employment states in the EQLS and the ESS. In both surveys these categories were presented to the respondent on a show card. The ESS asked the respondent to choose the category that best describes his or her situation of the last seven days best. The EQLS asked the respondent to choose his or her "principal economic status"[9] from the card.

Some of the answer categories of the EQLS and ESS denote equivalent employment states. The major difference concerns unemployment. The EQLS distinguishes between long- and short term unemployment, while the ESS distinguishes between unemployed people that are actively looking for a job and those that are not; the duration of unemployment is ascertained in a separate question. Despite that, the EQLS used finer grades for employment. Obviously it largely depends on the substantial questions, which of the two classifications is better. From a methodological point of view it seems sensible that actively looking for a job must be taken into account before asking for the duration of unemployment. However, there is no major problem with the EQLS question other than this.

Occupation

"Ever since it was recognised that the division of labour is the kernel of social inequality, stratification researchers have developed ways to derive social status measures from information on occupations" (Ganzeboom and Treiman, 2003, 159). Some of the more frequently applied strategies are the EGP class scheme (Erikson et al., 1979; Erikson and Goldthorpe, 1992)[10], the Standard International Occupational Prestige Scale (SIOPS), and the International Socio-Economic Index of Occupational Status (ISEI). All these strategies have in common that they start from a detailed

[9] The wording "economic status" is a technical term that misfit the answer categories. It should be substituted by a better term from common speech.

[10] The EGP class scheme is currently revised by members of the project "European Socio-economic Classification (ESeC)" (**http://www.iser.essex.ac.uk/esec/**)

classification of occupations, namely the International Standard Classification of Occupations (ISCO). This means that one first has to collect enough information about occupation to be able to construct the ISCO codes.

The standard format for collecting information on occupation is spread over several questions. The questions dealing explicitly with occupations are open questions. Thereby the interviewer has to ensure that the respondent provides enough ISCO-compatible information. The answers to the open questions are then either recoded by hand or by assistance of special computer programs (Braum and Mohler, 2003, 109).

Collecting information on international occupational coding systems is relatively complex. Therefore there exist several attempts to construct a closed question. Such attempts, for example, present the categories of the EGP scheme directly and add some examples to help respondents to identify their own status position on the scale. "This procedure, however, will lead to a loss in precision. Whether this is acceptable depends on how important the variable occupation is in a given project" (Braum and Mohler, 2003, 110).

The EQLS did not aim to classify using any of the above mentioned standard scales for occupation. Instead it used a single closed question, which was taken from the Eurobarometer series (D15). The ESS used three open questions on occupation, namely: "What is/was the name or title of your main job?", "In your main job, what kind of work do/did you do most of the time?", and "What training or qualifications are/were needed for the job?". The answers to these questions were then recoded to the ISCO-88. EGP, ISEI or SIOPS scales are not delivered with the data, but can be easily constructed with the help of computer routines that are available on the Internet.

Translation

The goal of translation is to apply the same questionnaire in each country, only somehow in different languages. Thereby a so-called "covered translation" is requested, i.e. a translation which reads like a text that has been originally written in the target language (Harkness and Schoua-Glusberg, 1998, 104). Translating a questionnaire might be done by one translator for one language ("one-for-one translation"), or by a several translators who make independent translations of the same questionnaire ("committee translation"). Both these forms might start from a finalised source language questionnaire (SLQ), or from a draft version allowing for changes in the SLQ in response to translation problems ("advance translating"). Techniques for quality assessments of the translated questionnaires include back translations, committee assessments, and statistical analyses (Harkness and Schoua-Glusberg, 1998, for an overview).

Little is known about how much these techniques affect the quality of the translations. It is likely that advance translating helps to avoid translation problems linked to the source text formulations. Advance translating can also help to minimise the cultural anchoring of the SLQ. There is little doubt that committee translating is less vulnerable to subjective judgements than one-for-one translation. However, both these techniques are costly and difficult to implement in the context of multi-lingual survey projects. Regardless of the translation technique, it is recommended to give the translators additional information about the intended content of each questions. Exchange between researchers, questionnaire designers, target language implementers and translators is always sensible (Harkness and Schoua-Glusberg, 1998, 96).

For quality assessment of the translated questionnaire committee assessment is generally recommended, whereby the committee should be formed by monolinguals, bilinguals and survey experts (Harkness and Schoua-Glusberg, 1998, 113f). Back translating, i.e. translating the translated questionnaire back into the language of the SLQ, is considered as less good for various reasons. It is said to often fail in picking up important problems of the translated text, and back translation does not help to improve a translation if problems are found. In any case back-translation can be only one step in a quality assessment that needs to be accomplished by others. From our look on the literature, the best practice

for quality assurance would be to have a committee assessment, that uses back-translations and statistical tests (split ballot assessments, double administration tests) as working material.

The EQLS started from a finalised SLQ in English, which was accompanied by "detailed explanations of many concepts included in the questionnaire" (Ahrendt, 2003, 5). The SLQ was translated into the main language of each participating country. More than one language was used in Belgium (French and Dutch), Spain (Spanish, Galician, Catalan and Basque), Luxembourg (Luxembourgish, French and German), Estonia (Estonian, Russian), Latvia (Latvian, Russian), Lithuania (Lithuanian, Russian) and Slovakia (Slovakian, Hugarian). The translation itself was passed to the fielding institute, which made two parallel translations as an intermediate step. The harmonised translated version was back-translated to English by professional translators. Discrepancies between the SLQ and the back-translated version "were discussed and solved with the local agencies". Further assessment of the target language questionnaire were done by a pilot study, during which 25 interviews had been carried out in each participating country. Overall, the translation procedure of the EQLS used elements of committee translations, and committee assessments combined with back-translation. This overall technique seems to guarantee a fairly good translation quality; however much depends on how the overall technique was carried out. The final translated questionnaires are publicly available from the Website of the Economic and Social Data Service.[11]

The ESS also started from a finalised SLQ in English. The SLQ was translated into each language spoken by more than 5 per cent of the persons of the participating countries. Unlike the EQLS the translation was not passed to the field work institutes. Instead, the national coordinators were responsible for the translation, guided by a team of translation specialists. The translation guidelines prescribes committee translation and committee assessment (Harkness, 2006). They also prescribe some characteristics the translators should have, how they should be selected and trained, and how their work has to be documented.

Whatever about the detail of the specific regulations, the major difference between the EQLS and the ESS is that the EQLS passed the translation to the field work institute. The general consequence of this decision is that the Foundation lost some control over the translation process and the final product. For an ongoing project like the EQLS this has the disadvantage that it is difficult to learn something from the translation of previous rounds for the translation of future rounds. The knowledge will be accumulated at the field work institute instead, and will get lost if the Foundation changes the field work institute. On the other hand, a decentralised translation process like that of the ESS seem to be not feasible for the central organised structure of the EQLS. In this case the field work institute should document the translation process with a list of deliverables. Harkness (2006, 12) provides a model for such a list.

Summary

The EQLS used an ASQ-approach for all background variables. From a methodological point of view this approach has its shortcomings, especially for the ascertainment of education, and to a lesser degree with occupation. Basically it requires that the respondents themselves to do some of the work that should be done by the researcher. It is however a more substantive question, whether the additional burden of harmonising country-specific answers to a common scale is really needed.

Regarding the translation we conclude that the translation method used for the EQLS should have resulted in reasonably good translations.

[11] http://www.data-archive.ac.uk/doc/5260/mrdoc/5260userguide2.pdf

Dataset

Data entry

- To analyse the answers of the respondents to the survey questions, they must be provided in a machine readable form. Three major ways exist to read in the question answers. Each has its advantages and disadvantages:
  - Transferring data from the questionnaire with a data entry system. The main disadvantage is the risk of transmission errors. Professional data entry institutes often use double entry to control the inputted data. Depending on the software used to input the data, error and consistency checks can be applied. A recommended system for data entry is the freeware program EpiData (**http://www.epidata.dk**). An alternative are self programmed data entry masks in a language like PHP, which allows to input the data in a SQL database (i.e. MySQL) over the Internet. It will however require a substantial amount of programming to implement consistency checks by this way of data entry.
  - Machine readable questionnaires can be used to automate the data entry process completely. The advantage is that the data entry will be finished rather quickly and error-free. The main disadvantage is the high price of hard and software. Moreover, machine readable questionnaires require accurate work of the interviewers when marking the answers. Finally, this method requires inclusion of some specific markers in the questionnaire, which can make respondents suspicious.
  - CAPI. In computer aided personal interviews the data entry step is bypassed by entering the answers of the respondent directly into the computer. The main advantage of this technique is that the data sets can be finalised rather quickly. The downside is that data entry errors can result in a serious disturbance of the interview.

The EQLS has used a centrally designed web site to input the data from the questionnaires. The research partners of most countries have used this web site, while Denmark, Ireland, Luxembourg, Sweden, Cyprus, Czech Republic and Poland have not. According to Ahrendt (2003, 6) the data of the seven countries that have not used the Internet had to be revised several times, so that overall the Internet mask is seen as more advantageous.

The ESS have used CAPI in some countries. It is not known which technique was used to input the data in the countries that have used PAPI.[12]

Variables to be included

Besides the answers to the questionnaire, a data set from a population survey should include some additional information.

Variables for the correct calculation of standard errors

Commonly, standard errors are calculated on the basis of formulae developed for SRS. For complex samples these formulae regularly underestimate the standard error, with the consequence that confidence intervals appear smaller than they are, and differences are termed as "significant" too early. Three variables are needed to be able to calculate correct standard errors: an identifier for the primary sampling unit, an identifier for the strata, and a variable holding the design-weights. The EQLS currently contains none of these variables, although the design-weights can be constructed from the household grid, and the information in the fieldwork technical report. The ESS contain design weights and, in 2004, interviewer numbers that can be used as a proxy variable for the primary sampling units.

[12] The inclusion of sampling point identifiers might be forbidden by data protection laws. An alternative technique to obtain correct standard errors from complex survey requires a set of balanced replicative weights.

Variables to control the data quality

It is common practice that interviewers have to complete a so-called interviewer questionnaire at the end of each interview. The questions in these interviewer questionnaires can provide invaluable information on the quality of the data. To be able to make use of this information, the data in the interviewer questionnaire should be added to the data sets. The EQLS does not provide data from an interviewer questionnaire, while the ESS provides this information in a separate file that can be merged to the main data set.

Variables to control the sample selection bias

Unit non-response is among the biggest threats to the quality of population surveys. Non-response studies are necessary to assess the extent of the sample selection bias, to find out its sources, and how to battle against the problem for further rounds. To perform such non-response studies it is necessary to provide the data of the contact forms for those who have responded, and for those who have not. Unfortunately this is very seldom done. Neither the EQLS nor the ESS provides any contact form data to the "normal user". The ESS researchers have however collected and analysed these data, and the report for the first round of the ESS is publicly available (Michel and Jaak, no year).

Summary

From a methodological point of view the EQLS data set provides too little information. It is neither possible to calculate correct standard errors, nor to demonstrate control of the work of the fieldwork institutes and interviewers. We strongly recommend to ask for such information for future rounds of the EQLS.

Regarding the data entry we do not put any recommendations here. From our point of view it is better to let each fieldwork agency use their preferred way to do it. They should build on their in- house experience here.

Ulrich Kohler, Wissenschaftszentrum Berlin für Sozialforschung (WZB); **Rory Fitzgerald**, Centre for Comparative Social Surveys, City University, London; **Arnošt Veselý**, Centre for Social and Economic Strategies, Charles University, Prague; **Dorothy Watson**, Economic and Social Research Institute, Dublin; **Femke DeKeulenaer**, Gallup Organisation Europe, Brussels

Appendix

The WZB has invited Rainer Schnell, a leading German survey specialist to comment on the EQLS-1 questionnaire. This appendix documents some of the critical remarks that came out of this consultancy.

Ill-defined terms

Several questions of the EQLS use terms that are ambiguous.

- Q5 asks for the size of the place where people usually works. It is left unclear, whether this should be the "company", the "commercial unit", or even something like a "building site".
- Q17 asks for rooms of the accommodation including the "rooms used solely for business". This might include store-rooms, shops, barns, etc.
- Q23 and Q24 ask for participation in the past month and the past year respectively. This can be interpreted quite different depending on the interview is at the end or the beginning of a month/year.
- Q27 asks for the trust in social security system. The term "trust" is a sociological concept that is not as clearly defined outside the sociological world. Less ambiguous might be to ask for a subjective evaluation of the likelihood of benefits.
- Q29 asks for tensions between social groups. It is unclear whether the term "tensions", or "cleavage" should be used. "Tensions" between men and women can be interpreted quite different as "tensions" between poor and rich.
- Q34, Q35 and Q37 ask behavioural frequency questions without defining the time frame of the question.
- Q45 ask for difficulties of getting to a doctor or "medical specialist". A medical specialist might be anything from a physician to healer or even pharmacist.
- Q48 ask for the training course within the "last year". This can be interpreted quite different depending on the interview is at the end or the beginning of a year.
- Q51 ask for the English reading ability. The reading ability can be limited by visual defects.
- Q53 ask for travelling school/work travelling time. It is left unclear whether this is asked for "per day" or "per trip".

Completeness of answer categories

Several scales used in the EQLS are incomplete in that they do not contain their natural extreme points:

- Q11 uses a likelihood rating scale to observe the likelihood of a job loss in the next 6 month, without mentioning "no chance" or "certain to happen".[13]
- Q13 behavioural frequency question on working conditions misses the extreme value "daily".
- Q26 uses terms like "more than once a week", or "once a year" to ask for the frequency of church attendance, but did not set the anchor "daily".

[13] An alternative five point scale might use the terms "no chance", "unlikely", "moderate chance", "likely" and "certain". A similar seven point scale would use "no chance". "very unlikely", "unlikely", "moderate chance", "likely", "very likely" and "certain to happen". It has been shown that the latter scale performed as well or better than other scale for perceptions of health hazard susceptibility.

- Q34, Q35 and Q37 ask behavioural frequency questions without defining the anchor for the lower extreme level. Instead the questions use “less often” (than several times a year), and “less often” (than once and twice a month) respectively. The scale misses the option “never”.
- Q36 asks for persons that support the respondent in specific life situations. The “partner” is not mentioned in the list of persons.

Response-Sets

- Q31 ask for general life satisfaction after the item battery on anomie. This might pose a position effect.
- Q39 ask for the equity of housework. The answers might be influenced by the presence of the partner during the interview. This needs to be controlled with by interviewer questionnaire.
- Q57 ask whether people feel safe when walking around in their area at night. For Germany, it has been shown that such questions are very vulnerable to response errors. A somewhat better alternative are questions asking for the personal risk of being victimised.[14]

[14] Examples are in Keane (1992) or in the “International Crime Survey” 1996.

Bibliography

Ahrendt, D., *The Quality of Life Survey*. On behalf of the European Foundation for the Improvement of Living and Working Conditions., fieldwork Technical Report of Intomart GfK, Hilversum, 2003

Bien, W. and Quellenberg, H.. *How to Measure Household and Family*, in Hoffmeyer-Zlotnik, J. H. P. (Ed.) Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables., New York: Kluwer Academic, 2003, pp. 279-293.

Bohnke, P., *First European Quality of Life Survey: Life Satisfaction, Happiness and Sense of Belonging*, Luxembourg: Office for Official Publications of the European Communities, 2005

Braun, M. and Mohler, P. P., *Background Variables*, in Harkness, J. A., Van de Vijver, F. J. R. and Mohler, P. P. (Eds.) Cross-Cultural Survey Methods, Wiley Series in Survey Methodology, Hoboken: Wiley, 2003, pp. 101-115.

Brauns, H., Scherer, S. and Steinmann, S., *The CASMIN Educational Classification in International Comparative Research*, in Hoffmeyer-Zlodnik, J. H. and Wolf, C. (Eds.) Advances in cross national comparison, New York: Kluwer Academic, 2003, pp. 221-244.

Diefenbach, M., Weinstein, N. D. and O'Reilly, J., *Scales for assessing perceptions of health hazard susceptibility. Health Education Research*, 8, 1993, 181-192.

Erikson, R. and Goldthorpe, J. H., *The Constant Flux: a Study of Class Mobility in Industrial Societies*, Oxford: Clarendon Press, 1992

Erikson, R., Goldthorpe, J. H. and Portocarero, L., *Intergenerational Class Mobility in Three Western European Societies*. British Journal of Sociology, 30, 1979, 415-441.

European Social Survey (2004a). *ESS Documentation Report 2002/03*: The ESS Data Archive, Ed. 5.1, **http://ess.usd. nib. no/2003/ESS1DataDocReporte041.pdf**.

European Social Survey (2004b). *Source Questionnaire*. Round 2, 2004/05. Final Version, **http://ess.nsd.uib.no/index.jsp**.

European Social Survey (2006). *ESS2-2004 Documentation Report*: The ESS Data Archive, Ed. 2.0, **http://ess.nsd.uib.no/index.jsp**.

Fahey T., Whelan, C. and Maitre, B.. *First European Quality of Life Survey: Income Inequalities and Deprivation*, Luxembourg: Office for Official Publications of the European Communities, 2005

Ganzeboom, H. B. and Treiman, D. J., *Three Internationally Standardised Measures for Comparative Research on Occupational Status*, in Hoffmeyer-Zlotnik, J. H. and Wolf, C. (Eds.) Advances in Cross National Comparison. A European Working Book for Demographic and Socio-Economic Variables, New York: Kluwer Academic, 2003, pp. 159-193.

Harkness, J. A. (2006). Round 3 ESS Translation Guidelines, **http://naticent02.uuhost.uk.uu.net/ess_docs/R3/Methodology/r3_ translati%on_guidelines. pdf**.

Harkness, J. A. and Schoua-Glusberg, A.. *Questionnaires in Translation*. Zuma Nachrichten Spezial, 3, 1998, 87-126.

Hoffmeyer-Zlotnik, J. H. and Wolf, C. (Eds.), *Advances in Cross-national Comparison*. A European Working Book for Demographic and Socio-Economic Variables, New York: Kluwer Academic/Plenum Publishers, 2003

Keane, C., *Fear of Crime in Canada: An Examination of Concrete and Formless fear of Victimization*. Canadian Journal of Criminology, 1992, 215-224.

Kish, L. *Survey Sampling*, New York: Wiley, 1965

Kish, L. ,. *Multipopulation Survey Designs: Five Types with Seven Shared Aspects*. International Statistical Review, 62, 1994, 167-186.

Kohler, U., *Internal Criteria for Representativeness for European Comparative Surveys*. Submitted to. Survey Research Methods, 2006a

Kohler, U., *Technical Assistance in Preparing second European Quality of Life Survey. Recommendations for the Methodology of EQLS-2*., Report to the European Foundation for the Improvement of Working and Living Conditions, 2006b

Kohler, U., *Quality Assessment of European Surveys. Towards an Open Method of Coordination for Survey Data*., in Alber, J., Fahey, T. and Sara-ceno, C. (Eds.) Social Conditions in the Enlarged Version, Routledge, 2007

Kreuter, F., *Kriminalitatsfurcht: Messung und methodische Probleme*, Opladen: Leske und Budrich, 2002

Michel, P. and Jaak, B. (). Work package 7. *Data-based Quality Assessment in the ESS*. Part 2: Non-response and Fieldwork Efforts in the ESS: Results from the Analysis of Call Record Data, **http://naticent02.uuhost.uk.uu.net/methodology/nonresponsefieldworkefforts.pdf**.

Nauenburg, R. and Mertel, B., *European Quality of Life Survey*. Technical Report. Draft version., paper distributed by the Wissenschaftzentrum Berlin, 2004

Saraceno, C., Olagnero, M. and Torrioni, P., *First European Quality of Life Survey: Families, Work and Social Networks*, Luxembourg: Office for Official Publications of the European Communities, 2005

Schnell, R., *Die Homogenitat sozialer Kategorien als Voraussetzung für '"Reprasentativitat"' und Gewichtungsverfahren. Zeitschrift fiir Soziologie*, 22, 1993, 16-32.

Schnell, R., Hill, P. B. and Esser, E., *Methoden der empirischen Sozial-forschung*., 7 edn., Miinchen u. Wien: Oldenbourg, 2005

Schnell, R. and Kreuter, F. *Separating Interviewer and Sampling-Point Effects*. Journal of Official Statistics, 3, 2005, 389-410.

Schnell, R. and Trappmann, M., *The Effect of the Refusal Avoidance Training Experiment on Final Disposition Codes in German ESS-2*, working paper 3/2006 of the Center for Quantitative Methods and Survey Research University of Konstanz, Germany, 2006

UNESCO, *International Standard Classification of Education, ISCED 1997*, in Hoffmeyer-Zlodnik, J. H. (Ed.) Advances in Cross-National Comparison. A European Working Book for Demographic and Socio-Economic Variables, New York: Kluwer Academic, 2003, pp. 195-220.

Warner, U. and Hoffmeyer-Zlotnik, J. H.,. *How to measure income,* in Hoffmeyer-Zlotnik, J. H. and Wolf, C. (Eds.) Advances in cross-national comparisons. A European Working Book for Demographic and Socio-Economic Variables., New York: Kluwer Academic, 2003, pp. 307-323.

EF/07/24/EN

