

Assessing the Demand Level of Survey Questionnaires: A Meta-Analysis of Experiments in Question Form and Wording

Valentina Hlebec¹ and Miha Rozman¹

Abstract

In this paper meta-analyses of experiments in question form and wording are presented. The demand level of survey questionnaires as perceived by respondents (Slovenian Housing Survey, 2005) was measured by a single question, which was varied in its question form and wording. Each respondent answered one version of the question in a split ballot experimental design. Multiple Classification Analysis was used to evaluate the effects of the question wording experiments and of respondent's characteristics (age and education) on the demand level of the survey as perceived by respondents. Results show that the terms "survey" and "survey questions on average" are equivalent to all respondents, and that the term "demanding" is understood and used differently by older respondents than the term "difficult". Formal balance has no effect on the estimated level of difficulty of the survey for respondents; however, labeling of extreme values has a strong effect on the estimated level of difficulty in the Housing survey.

1 Introduction

Each survey question should meet three distinct standards (Groves et al., 2004: 241-242); it should measure what it is intended to measure (content standard); respondents should be able to understand and answer the question (cognitive standard), and with a reasonable degree of effort (usability standard). Different evaluation methods are used to evaluate how particular survey questions and questionnaires meet these standards (ibid.; Snijkers, 2002; Presser et al. 2004; Biemer and Lyberg, 2003). Cognitive and usability standards are sometimes referred to as "respondent burden", and it is considered to be (Biemer and Lyberg, 2003: 107-109) a general concept, which, in individual and household surveys, reflects the degree to which the respondent perceives the survey as demanding and

¹ University of Ljubljana, Faculty of Social Sciences, Kardeljeva ploščad 5, SI-1000 Ljubljana, Slovenia.

time consuming. It includes a variety of components, varying from questionnaire characteristics to the number of survey requests received by the respondent in a certain period. Respondent burden and/or quality standards are addressed at all stages of survey research (development, implementation and postsurvey testing). During the first stage, respondent burden is assessed with expert appraisal and cognitive laboratory methods, which usually produce a qualitative evaluation of the effort received from a respondent in order to answer survey questions (Snijkers, 2002). During pilot study and during implementation, response latency can be measured and may give an indication of problems at all four stages of question and answer exchange (Biemer and Lyberg, 2003: 271-272). Questions that require longer a time to complete can be regarded as more burdensome for a respondent. Response burden is seen as closely related to survey nonresponse, both partial nonresponse and dropout rate (ibid. 107-108) and is assessed in the pilot study and at the implementation stage of the survey. In most of these methods, respondents are not asked directly about "how burdensome the questionnaire is for them personally". Exceptions are focus groups and confidence ratings (a variety of cognitive techniques, see Groves, 2003: 246), where respondents are asked to evaluate specific components of response burden. Both of these techniques are used in pretests, and their results are used to improve the final version of the survey questionnaire (they cannot be used to measure the amount of respondent effort to answer the final version of the survey questionnaire). We suggest using a direct measure of respondent burden as the last question in the survey questionnaire, where respondents themselves are asked directly how difficult it was for them to answer the implemented survey questionnaire. It is impossible to evaluate each survey question separately as this would double the time needed for completion and would affect the question and answer process in a manner similar to concurrent think alouds (Willis, 2004). Using a single indicator to evaluate respondent burden has several disadvantages, since we cannot establish what were the causes of the burden (question content, wording, response options, questionnaire length, mode of administration, etc.), but it can nevertheless give an overall impression of the survey from the respondents' view point. The perception of respondent burden can thus be evaluated directly and later be completed by indirect measures such as response latency or partial nonresponse.

When deciding to use only one indicator of respondent burden, how should this indicator be designed? Should we ask about the survey in general, or should we ask about survey questions on average? Should we ask about how easy or how demanding or difficult the questions were to answer? Does it matter if we formally balance the question? We designed an experiment in which we address the above questions. The meta-analyses of question wording experiments² involving survey questions assessing the demanding level of a survey questionnaire (Slovenian Housing Survey, 2005) are presented. In the following section previous research is

² The exact wording is in the Appendix.

presented, research questions are outlined and the aim of the paper is clearly stated. In Section 3 the experimental design and data are presented, followed by sections with analysis and discussion.

2 Previous research and the aim of the paper

The first systematic approach to study question wording of survey questions was presented by Schuman and Presser (1981: 3-12), who studied the following types of influences of the measurement instrument (survey questions) on measured variables: question format characteristics (open versus closed questions, don't know and middle position, balance of the survey questions, agree-disagree items, attitude strength and crystallization, tone of wording), question content, and question and response order. We would like to assess tone of wording, formal balance of survey question and selection of labels for a five-point ordinal response scale of the indicator of respondent burden in the Slovenian Housing Survey.

Researchers are often faced with the problem of selecting the tone of wording - the right words for one's survey questions. Which words are more widely accepted and understood in the general population, and which are too abstract for respondents? Sometimes, even a slight change in tone of wording affects the measured variables (e.g. Schuman and Presser, 1981; Krosnick, 1989; Rasinski, 1989; Holeman, 1991; Hunter and Myazdick, 2002). The first research question we address in this paper (R1) is as follows: Which tones of wording in our experiments are equivalent ("survey questionnaire" versus "survey questions on average", "demanding" versus "difficult")? Does it make a difference to respondents if we point out that the survey questionnaire comprised many questions, and that they should consider all of them and make an average estimation of the effort they used to answer them? Are the terms "difficult" and "demanding" perceived as equivalent?

Survey questions should, in order to be neutral, be balanced, i.e. both argument and counter argument or agreement and disagreement should be presented to respondents (e.g. Schuman and Presser, 1981, Sudman and Bradburn, 1991). The easiest way to obtain balance is to use formal balance, i.e. in the text of the survey question presented to respondents, both pro and con are explicit (not only in the response categories). However, experiments done by Schuman and Presser (1981) show that formal balance does not always affect the measured variable to any great extent, as opposed to the use of argument and counter argument (i.e. supporting the pro and con with specific reason). Therefore, we would like to find out (R2) whether the formal balance of question assessing the effort respondents made when answering the questionnaire affects the mean value of estimated respondents' burden.

Based on forbid-allow asymmetry (Schuman and Presser, 1981; Holleman, 1999) and finding that labeling of extreme values of response scales has an effect

on measured variables, we assume (R3) that the labels introducing the most extreme values would cause a change in the measured variables. We assume that terms "undemanding" and "not at all demanding" represent equivalent labels for extreme values, while the terms "very easy" and "not at all difficult" are not equivalent; the term "very easy" introduces a new dimension of respondent burden (the estimated average burden should be higher when the label "very easy" is used).

Most of the question wording experiments have been tested against demographic characteristics of respondents (at least age and education), and some of them were related to the age and education of respondents (Schuman and Presser, 1981, Scherpenzeel, 1995b; Kogovšek, 2001). We would like to know whether, apart from the main effects (i.e. we would expect older and less educated respondents to find the Housing survey more demanding), there are any associations of age and education of respondents with the wording experiments presented in this paper (R4).

Meta-analyses of quality of survey instruments (i.e. reliability and validity) suggest similar research questions (see Andrews, 1990; Ferligoj, Leskošek and Kogovšek, 1995; Scherpenzeel, 1995a-c; Krebs, Berger and Andreenkova, 1995; Koeltringer 1995), since they stress that, among the characteristics of measurement instruments (i.e. survey questions), characteristics of response scales (number and labeling of response categories - R2, R3) have the strongest effect on data quality. Among characteristics of respondents, education and age of respondents affected data quality the most (R4).

Table 1: Question wording experiments.

Scale		Tone of wording		
Formal Balance	Unipolar bipolar	Demanding Difficult	Survey questionnaire	Survey questions on average
Yes	U	Undemanding (1) Very demanding (5)	QW2	QW6
	B	Very easy (1) Very difficult (5)	QW4	QW8
No	U	Not at all demanding (1) Very demanding (5)	QW1	QW5
	U	Not at all difficult (1) Very difficult (5)	QW3	QW7

3 Experimental design and data

The experiment in question form and wording was carried out as part of the Slovenian Housing Survey (2005), which will be described in detail in the second part of this section. Altogether there were eight versions of the question assessing the demand level of the Housing Survey. Since this was a split ballot experiment, each respondent answered only one version of the question. Respondents were randomly assigned to the experimental groups; therefore the observed variability in the dependent variable (respondent burden) can be attributed to independent variables (i.e. characteristics of the question wording experiment: two tones of the wording experiments, formal balance and polarity of the scales). The characteristics of the wording experiments are described in the Table 1.

Following the examples given in other meta-analyses for explaining the effects of different characteristics in the measurement instruments on data quality, (Scherpenzeel, 1995a; Hlebec, 1999), Multiple Classification Analysis (MCA, Andrews et al. 1973) was chosen as the meta-analysis technique. The multivariate (MCA) coefficients indicate how much the level of dependent variables deviates from the total mean as a result of a given characteristic of the measurement instrument (e.g. formal balance of survey question), while controlling for the effects of all other characteristics of the measurement instrument and demographic variables (age and education). Two measures of the overall effect of each predictor (i.e. characteristics of measurement instrument and demographic characteristics) are obtained, and in addition the MCA Eta and MCA Beta (the MCA Eta measures the strength of the bivariate relationship between a dependent variable and a predictor; the MCA Beta measures the strength of the relationship, controlled for the other predictor variables in the model). The rank order of the Betas indicates the relative importance of the predictor variables in their explanation of the dependent variable. Finally, the multiple R is estimated indicating the total proportion of variance explained by all predictors together.

Data for this experiment were collected as part of the Housing Survey in Slovenia. The data collection mode was CATI; data was collected between 13. 4. 2005 and 27. 5. 2005 (for details, see Hlebec and Gnidovec, 2006). Altogether, 4009 respondents were interviewed. The sampling unit was a household and not an individual; therefore, the sample was weighted according to the characteristics of households. All data analysis was done on weighted data. Respondents were self-selected, based on their knowledge about housing matters. It is therefore possible that experimental groups differ in the demographic characteristics of respondents, regardless of the random attribution of households to experimental groups. The demographic characteristics of respondents in the total sample and in the experimental groups are presented in Table 2.

Table 2: Demographic characteristics of the sample.

	G1	G2	G3	G4	G5	G6	G7	G8
Gender								
Male	32	33	33	33	36	33	27	32
Female	68	67	67	67	64	67	73	68
Education								
Elementary	13	17	16	16	18	17	10	11
High school	59	55	55	54	59	55	56	56
University	28	28	29	30	24	28	34	33
Age								
- 30	11	14	14	12	16	14	16	13
30-50	44	42	39	42	42	39	47	40
50+	44	44	47	46	42	48	37	47

Based on the χ^2 test, we can say that all groups except group 7 are equivalent in demographic composition. In group 7 there are some statistically significant differences: more women, more respondents in the middle aged group, and more with higher education. Therefore, we have to be careful in interpreting the results of group 7, since some variation can be attributed to the demographic composition of the group.

4 Results

Firstly we present the descriptive statistics of the dependent variables, and then we present the multivariate analyses and answer our research questions.

Table 3: Descriptive statistics of dependent variables.

Dependent Variable	n	Mean	Std. Dev.
QW1	388	2,23	1,215
QW2	358	2,14	1,198
QW3	360	1,90	1,046
QW4	305	2,10	1,042
QW5	307	2,14	1,131
QW6	337	2,27	1,262
QW7	347	1,72	0,954
QW8	370	2,08	0,965

Table 4: Predictive power and effects for the tone of wording experiments and characteristics of respondents on mean level of difficulty of the survey questionnaire.

		N	Sig.	<u>Mean = 2.07</u>		
				Bivariate Eta	Multivariate Beta	Dev'n
TONE OF WORDING 1						
Survey questionnaire		1098				.015
Survey questions on average		981		.021	.014	-.017
TONE OF WORDING 2						
Demanding		1383	***			.116
Difficult		696		.151	.142	-.231
EDUCATION						
Elementary		312				.348
High School		1169				-.015
University		598	***	.153	.137	-.153
AGE						
- 30		293				-.300
30 - 50		876				.030
50 +		910	***	.128	.106	.068
Multiple R ²						.056

As shown in Table 3, there are differences in the observed mean levels of respondent burden. Based on univariate and bivariate tests (Rozman and Hlebec, 2008), there were some significant differences in the level of dependent variables that cannot be attributed to respondents' characteristics but to the characteristics of the measurement instrument, as well. It was shown (ibid.) that we can treat the term "survey questionnaire" as equivalent to the term "survey questions on average", since there were no statistically significant differences between the assessed mean level of respondent burden. It seems that respondents, even if they are reminded to consider all questions in the questionnaire, evaluate the effort they make in answering the survey questionnaire in a similar way. The terms "difficult" and "demanding" cannot be treated as equivalent: there were significant differences in estimated respondent burden. Formal balance made no difference, while the polarity of the scale affected the mean levels of measured respondent burden. None of these results were tested against demographic variables and controlled for interactions among predictor variables. Therefore, the multivariate tests were done with MCA analysis. Altogether, three separate³ MCAs' were needed, since there is an interaction in the experimental design. There were two groups with bipolar scale (QW4 and QW8), which simultaneously included both the term "difficult" and formally balanced survey question wording. Therefore, for

³ The need for three meta-analyses arises because of the complexity of the experimental design. If there are several predictor variables (as in this case), it may happen that higher order interactions are not estimated by the MCA. An appreciated solution (for example, Hlebec, 1999, 2001) is to run several analyses, each time taking a specific combination of predictor variables into account. The researcher then assesses all the analyses at the same time, looking for identical or contradictory findings.

the first MCA (R1, R4), these two groups were excluded from the analysis. When assessing the equality of the terms "survey questionnaire" and "survey questions on average", and the terms "difficult" and "demanding" and controlling for age and education of respondents and interactions between predictor variables, the experimental groups with bipolar scale were excluded from comparison.

Table 5: Predictive power and effects for the formal balance of survey questions and characteristics of respondents on mean level of difficulty of the survey questionnaire.

		N	Sig.	<u>Mean = 2.19</u>		
				Bivariate Eta	Multivariate Beta	Dev'n
FORMAL BALANCE						
	Yes	691			.005	
	No	692		.005	.004	
EDUCATION			***			
	Elementary	220			.394	
	High School	788			-.024	
	University	375		.161	.153	
AGE			***			
	- 30	190			-.426	
	30 - 50	577			.054	
	50 +	616		.156	.141	
Multiple R ²					.046	

When controlling for multivariate interaction, tone of wording (2) experiment, education and age of respondents were significantly related to the mean level of difficulty of the survey questionnaire. These three factors therefore predict the level of difficulty of the survey questionnaire, i.e., less educated respondents and older respondents find the survey more demanding. There is also a significant interaction between age of respondents and tone of wording (2) experiment, indicating that the term "demanding" interacts more strongly⁴ with age of respondents (the older the respondents the more demanding the survey) than the term "difficult". Multivariate analysis produces new findings: namely, the term "survey questionnaire" can be used interchangeably with the term "survey questions on average". These terms are equivalent to all respondents regardless of their age and education. Regardless of tone of wording experiment, older and less educated respondents find the Housing Survey more demanding. In the Slovenian language, one should use the term "difficult" (Slo. "teško") rather than the term "demanding" (Slo. "zahtevno"), since it is more commonly accepted by respondents.

4

Age	Demanding	Difficult	Total
- 30	1.75	1.73	1.74
30 - 50	2.19	1.78	2.05
50 +	2.33	1.90	2.19

In the second MCA (R2, R4) we tested the effects of formal balance in the survey question, age and education of respondents on the mean level of difficulty of the survey questionnaire. To allow formal balance alone to affect the data, only the groups using the term "demanding" were used (QW2, QW6, QW1, and QW5).

Not surprisingly, formal balance has no effect on the mean level of the dependent variable. An important finding is that there are no significant higher order interactions. Therefore the formal balance of these questions is not important to respondents, regardless of their age and education.

In the third MCA (R3, R4) we tested the effect of bipolarity in the scale, education and age of respondents on the mean level of difficulty of the survey question. To allow only polarity of the scales to affect the data, the groups using the term "difficult" were analyzed for multivariate analysis.

Table 6: Predictive power and effects for the polarity of the response scales and characteristics of respondents on mean level of difficulty of the survey questionnaire.

		N	Sig.	Mean = 1.95		
				Bivariate Eta	Multivariate Beta	Dev'n
SCALE			***			
	Unipolar	696				-.129
	Bipolar	675		.131	.130	.133
EDUCATION			***			
	Elementary	180				.254
	High School	753				.016
	University	437		.120	.118	-.132
AGE			*			
	- 30	188				-.152
	30 - 50	575				.032
	50 +	607		.076	.060	.017
Multiple R ²						.035

The effects of all predictor variables are statistically significant, indicating that the estimated level of the measured variable depends significantly on the labeling of extreme values on the five-point ordinal scale, age and education of respondents. There were no significant higher order interactions, indicating that the change in polarity of the scale (bipolar: 1 - »very easy«, 5 - »very difficult« vs. unipolar: 1 - »not at all difficult«, 5 »very difficult«) affects all respondents in the same way.

5 Discussion

Even though these experiments reveal new knowledge about question wording, they are, like most of these experiments, case studies. Both tone of wording experiments are relevant for Slovenian surveys and cannot be generalized beyond

the Slovenian language. As far as formal balance is concerned, we can say that the questions measuring the level of difficulty of the questionnaire belong to that group of questions where formal balance is irrelevant to the measured variable. More widely generalizable is the finding about labels of extreme values on the five-point ordinal scale.

These meta-analyses of question form and wording experiments assessing indicators of respondent burden for the Slovenian Housing Survey show some expected and some unexpected results. Tone of wording experiments suggest that the terms "survey questionnaire" and "survey questions on average" are equivalent for all respondents, regardless of their age and education. Therefore, one can assess the effort the respondent made to answer the survey questionnaire using either term.

The terms "demanding" and "difficult" (Slo. "zahtevno" and "težko") are understood and used in different ways by older respondents. The difference increases with increased age. It seems that the term "difficult" is used similarly by all respondents, which suggests that one use this term in questions assessing the demand level of survey questionnaires. Further, qualitative testing is needed to fully comprehend how older respondents understand and interpret these two terms.

Formal balance did not play a role in the level of the measured variable; what is more, it is used the same way by all respondents, regardless of their age and education. Selection of labels for extreme values of a 5-point ordinal scale is very important. Whereas the terms "not at all demanding" and "undemanding" are equivalent, the terms "very easy" and "not at all difficult" represent the extreme values on the bipolar vs. unipolar scales, respectively.

The fact that older respondents and less educated respondents find the survey more difficult to answer supports the suggestion that this question can be used as an indicator of respondent burden. It is consistent with other quality indicators (Scherpenzeel, 1995b, Kogovšek, 2001) tested against demographic variables. However, further examination of this indicator of respondent burden is required, such as assessing the association with response latency or partial nonresponse, or examining the validity and reliability of data given by respondents who report higher respondent burden.

References

- [1] Andrews, F.M. (1990): Construct validity and error components of survey measures: a structural modeling approach (reprint from *Public Opinion Quarterly*). In Saris, W.E., and van Meurs, A. (Eds.): *Evaluation of Measurement Instruments by Meta-Analysis of Multitrait Multimethod Studies*, 15-51. Amsterdam: North-Holland.
- [2] Andrews, F.M., Morgan, J.N., Sonquist, J.A., and Klem, L. (1973): *Multiple Classifications Analysis*. Ann Arbor: Institute for Social Research.

- [3] Biemer, P.P. and Lyberg, L.E. (2003): *Introduction to Survey Quality*. New Jersey: Wiley.
- [4] Ferligoj, A., Leskošek, K., and Kogovšek, T. (1995): *Zanesljivost in veljavnost merjenja*. Ljubljana: Založba FDV.
- [5] Hlebec, V. (1999): *Evaluation of Survey Measurement Instruments for Measuring Social Networks*. Ljubljana: University of Ljubljana, Faculty of Social Sciences.
- [6] Hlebec, V. (2002): Meta-analiza zanesljivosti anketnega merjenja socialne opore v populnih omrežjih. *Teorija in Praksa*, **1**, 63-76.
- [7] Hlebec, V. and Gnidovec, M. (2006): Metodologija raziskave. In Mandič, S. and Cirman, A. (Eds.): *Stanovanje v Sloveniji 2005*, 9-53. Ljubljana: Fakulteta za družbene vede.
- [8] Groves, R.M., Fowler, F.J., Couper, J.L., Lepkowski, J.M., Singer, E., and Tourangeau, R. (2004): *Survey Methodology*. New Jersey, Wiley.
- [9] Kogovšek, T. (2001): *Ocenjevanje zanesljivosti in veljavnosti merjenja značilnosti egocentričnih socialnih omrežij*. Ljubljana: University of Ljubljana, Faculty of Social Sciences.
- [10] Holleman, B. (1999): The Nature of the Forbid/Allow Asymmetry. *Sociological Methods & Research*, **2**, 209-244.
- [11] Hunter, G. and Miazdyck, D. (2002): *Question wording and Public Opinion about Social Assistance*. Social Policy Research Unit. Regina: Faculty of Social Work (May 9, 2006).
<http://www.ualberta.ca/parkland/research/occasionalpapers/opinwelfare.pdf>
- [12] Koeltringer, R. (1995): Measurement quality in Austrian personal interview surveys. In Saris, W.E., and Münnich, A. (Eds.): *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*, 207-224. Budapest: Eötvös University Press.
- [13] Krebs, D., Berger, M., and Andreenkova, A. (1995): Political efficacy in Russia - An MTMM two wave panel model. In Saris, W.E., and Münnich, A. (Eds.): *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*, 145-154. Budapest: Eötvös University Press.
- [14] Krosnick, J. (1989): The polls - a review: Question wording and reports of survey results: The case of Louis Harris and associates and AETNA life and casualty. *Public Opinion Quarterly*, **1**, 107-113.
- [15] Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., and Singer, E. (2004): *Methods for Testing and Evaluating Survey Questionnaires*. New Jersey, Wiley.
- [16] Rasinski, K. (1989): The effect of question wording on public support for government spending. *Public Opinion Quarterly*, **3**, 388-395.
- [17] Rozman, M. and Hlebec, V. (2008): The effect of question wording on the estimation of difficulty in a survey. Submitted to Družboslovne razprave.

- [18] Scherpenzeel, A. (1995a): Design of meta analysis. In Saris, W.E., and Münnich, A. (Eds.): *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*, 187-206. Budapest: Eötvös University Press.
- [19] Scherpenzeel, A. (1995b): Meta analysis of a European comparative study. In Saris, W.E., and Münnich, A. (Eds.): *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*, 225-242. Budapest: Eötvös University Press.
- [20] Scherpenzeel, A. (1995c): *A Question of Quality: Evaluating Survey Questions by Multitrait-Multimethod Studies*. Amsterdam: Royal PTT Nederland NV.
- [21] Schuman, H. and Presser, S. (1981): *Questions & Answers in Attitude Surveys*. London: Sage Publications.
- [22] Snijkers, G.M.J.E. (2002): *Cognitive Laboratory Experiences: On Pre-testing Computerised Questionnaires and Data Quality*. Utrecht: University of Utrecht.
- [23] Sudman, S. and Bradburn, N. (1991): *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey - Bass Publishers.
- [24] Willis, G.B. (2004): *Cognitive Interviews: A "How to" guide*. Research Triangle Institute.

Appendix: Question wording of dependent variable

- QW Assessing the demand level of survey questionnaires
- QW1 Finally, we would like to know how **demanding** answering the **survey** seems to you on a scale from 1-**not at all demanding** to 5-**very demanding**.
- QW2 Finally, we would like to know how **undemanding** or **demanding** answering the **survey** seems to you on a scale from 1-**undemanding** to 5-**very demanding**.
- QW3 Finally, we would like to know how **difficult** answering the **survey** seems to you on a scale from 1-**not at all difficult** to 5-**very difficult**.
- QW4 Finally, we would like to know how **easy** or **difficult** answering the **survey** seems to you on a scale from 1-**very easy** to 5-**very difficult**.
- QW5 Finally, we would like to know how **demanding** answering the **survey questions on average** seems to you on a scale from 1-**not at all demanding** to 5-**very demanding**.
- QW6 Finally, we would like to know how **undemanding** or **demanding** answering the **survey questions on average** seems to you on a scale from 1-**undemanding** to 5-**very demanding**.
- QW7 Finally, we would like to know how **difficult** answering the **survey questions on average** seems to you on a scale from 1-**not at all difficult** to 5-**very difficult**.
- QW8 Finally, we would like to know how **easy** or **difficult** answering the **survey questions on average** seems to you on a scale from 1-**very easy** to 5-**very difficult**.