Creating a good question: How to use cumulative experience

Melanie Revilla
Diana Zavala
Willem Saris

RECSM, Universitat Pompeu Fabra

Introduction

Creating a good question for survey research is a difficult task given the many decisions that are involved. The fundamental part of a question is what we will call the request for an answer. This part always has to be available. However, besides the request for an answer, many different components can be added such as an introduction, a motivation statement, extra information regarding the content or definitions, instructions to the respondents or the interviewer, a show card and finally answer categories. All these components can be formulated in many different ways.

The effects that the wording of survey questions can have on the responses have been studied in the tradition of survey research, for example by Sudman & Bradburn (1983), Schuman & Presser (1981), Andrews (1984), Alwin & Krosnick (1991), Költringer (1993), Scherpenzeel & Saris (1997) and Saris & Gallhofer (2007). On the contrary, little attention has been given to the problem of translating the concepts one wants to measure into the basic component of a question, the request for an answer (De Groot & Medendorp 1986; Hox 1997). However, if this step goes wrong, little improvements can be made afterwards. Therefore, we first discuss how to design theoretically valid requests for an answer. This can be achieved following the three-step procedure proposed by Saris & Gallhofer (2014):

1. Specification of the concept-by-postulation (complex concepts like attitudes) in concepts-by-intuition (simpler concepts for which questions can be directly formulated)

2. Transformation of concepts-by-intuition in assertions indicating the requested concepts

3. Transformation of assertions into requests for an answer

The three-step procedure, if properly applied, should lead to a measurement instrument that measures what is supposed to be measured and nothing else.

However, there are much more decisions to take while designing questions that can affect the results. There are many possible formulations of questions that are theoretically valid for a given concept. But not all of them are equally good. Therefore, it makes sense to study the quality of the question. This quality is defined as the strength of the relationship between the latent variable of interest (the concept-by-intuition) and the observed answers to a specific formulation of the question. A quality of 1 would mean that there are no measurement errors at all: the question would perfectly measure the concept-by-intuition. Nevertheless, in practice, the quality is always lower than one.

Indeed, the quality of a question is determined by the amount of random and systematic errors which are estimated respectively as 1-reliability of the question and the variance in the observed responses explained by the method used (Saris & Andrews, 1991). Over the last 25 years, a lot of experiments have been done to estimate the random and systematic errors for many questions (Andrews, 1984; Költringer, 1993; Scherpenzeel & Saris, 1997; Saris et al., 2011; Saris & Gallhofer, 2014). This cumulative knowledge has led to the development of the program SQP 2.0 that can be used to predict the quality of questions on the basis of the characteristics of these questions.

In this chapter, we start by discussing the design of theoretically valid questions. Then, we give an overview of the other decisions that have to be taken in order to get the final formulation of survey questions and we show how the SQP program, based on a meta-analysis of thousands of quality estimates, can help researchers in evaluating if their questions are good enough. SQP also provides suggestions of improvements. We propose a concrete illustration of how to code a question in SQP in order to check its quality and improve it. Finally, we explain how SQP can also be used in cross-national research to create good questions over different languages.

1. How to design theoretically valid requests for an answer

In the first step, we should determine if we consider that the concept of interest is a complex concept that require more than one indicator to be measured, or if it is a simple concept about which we can ask directly. If our concept is a concepts-by-postulation, then, it has to be operationalized in terms of several simpler concepts (concepts-by-intuition), using reflective or formative indicators (for more details, we refer to Saris & Gallhofer 2014). In this chapter, we concentrate on the second and third steps of the 3-step procedure, have determined the different concept-of-intuition needed..

The transformation of concepts-by-intuition into questions is not so simple if we want to provide a procedure that leads to a question representing the concept-by-intuition with a very high probability. One of the reasons for this is that so many different concepts exist that one cannot make rules for each of them. Another reason is that there are many possible ways of formulating questions.

We try to simplify the task by classifying the different concepts used in social sciences in general classes of basic concepts. For these basic concepts, we can formulate valid questions. First, assertions can be formulated that represent them quite certainly. After that, these assertions can be transformed in questions.

1.1 Basic concepts and concepts-by-intuition

There is a nearly endless list of possible concepts in social sciences. We cannot specify how to formulate questions for each of them. However, if we can reduce the number of concepts by classifying them into a limited number of classes of basic concepts, then this problem may be solved.

Table 16.1 illustrates that by presenting a list of concepts-by-intuition measured in round 1 of the European Social Survey (ESS) and indicating the classes of basic concepts to which they belong.

Table 16.1: The classification of concepts-by-intuition from the ESS into classes of basic concepts of the social sciences

Question ID	ESS concepts	Basic concept
B33	evaluation of services	evaluation
C7	health	evaluation
A8	social trust	feeling
B7	political trust	feeling
B27	satisfaction with ….	feeling
C1	Happiness	feeling
B1	political interest	importance
S	value benevolence	importance
B2	political efficacy	judgment
B28	left right placement	judgment
C5	victimization of crimes	evaluative belief
C16	discrimination	evaluative belief
B44	income equality	policy
B46	freedom of lifestyle	policy
C9	religious identity	similarity or association
C10	religious affiliation	similarity or association
C14	church attendance	behavior
C15	Praying	behavior
B13-B19	political action	behavior
A1	media use	frequency or amount
C2	social contacts	frequency or amount
C20	country of origin	demographic
F1	household composition	demographic
F2	Age	demographic

Table 16.1 does not give examples of all possible basic concepts. Instead, it shows that there are many different concepts-by-intuition that can be seen as specific cases of basic concepts. For example, evaluation of services and evaluation of one's own health have in common that both are evaluations (good or bad) while the subject in this case is very different. If we know how to formulate sentences that express an evaluation, we can apply this rule to both concepts to formulate questions that measure what we want to measure. Other examples can be interpreted in the same way. It illustrates that even if the number of concepts-by-intuition is nearly unlimited, they can be classified in a limited number of basic concepts. For each basic concept, an assertion can be formulated.

1.2 The basic elements of assertions

In linguistics, a simple assertion can be decomposed into main components. A first relevant basic structure for assertions expressing evaluations, feelings, importance, demographic variables, values, and cognitive judgments is Structure 1.
Structure 1: Subject + Predicator + Subject Complement.

For example:

Q1a- *Clinton* *was* *a good president.*
Subject + Predicator + Subject Complement.

"Clinton" functions as the *subject* that indicates what is being discussed in the sentence. The *predicator* or verb "was" connects the subject with the remaining part of the sentence, which expresses what the subject *is* and is therefore called a *subject complement.* It contains a noun ("president") with an adjective ("good") as modifier of the noun. . Predicators that indicate what a subject is/was or becomes/became are called *link verbs.* .

A second relevant linguistic structure used to formulate e.g. relations, preferences, duties, rights, actions, expectations, feelings, and behaviour, is Structure 2.

Structure 2: Subject + predicator + direct object.

It is illustrated in example Q1b:
Q1b- *The president* *likes* *his job*.
Subject + predicator + direct object.

This example has a subject ("the president"), the predicator "likes," and a direct object "his job." Koning & Van der Voort (1997: 52) define a *direct object* as the person, thing, or animal that is "affected" by the action or state expressed by the predicator.

By changing the predicator in this structure, we change the concept-by-intuition the assertion refers to.

There is a third linguistic structure relevant to present behaviors, behavioral intentions, and past and future events.

Structure 3: Subject + predicator
For example:
Q1c- *The position of the president* *has changed*.
Subject + Predicator.

Example Q1c has a subject ("the position of the president") and a predicator ("has changed"). In linguistics, these verbs which are not followed by a direct object are called *intransitive*.

The meaning of the sentences is easily changed by modifying the predicator as in structure 2. However, the number of possibilities is much more limited because of the reduced number of intransitive verbs.

The basic components of these three possible linguistic structures of assertions can be extended with other components (Saris & Gallhofer 2014).
These structures are not only used in English but in most European languages. We will treat further the topic of cross-national questions in section 4.

1.3. Basic concepts-by-intuition

Now, we will describe how assertions that are characteristic of the basic concepts-by-intuition employed in survey research can be generated. For more details about the different points discussed in this section, we refer to Saris & Gallhofer (2014).

We distinguish between subjective and objective variables. By subjective variables, we understand variables for which the information can only be obtained from a respondent because the information exists only in his/her mind. The following basic concepts-by-intuition are discussed: evaluations, importance judgments, feelings, cognitive judgments, perceived relationships between the X and Y variables, evaluative beliefs, preferences, norms, policies, rights, action tendencies and expectations of future events.

By *objective variables,* we mean non-attitudinal variables, for which, in principle, information can also be obtained from a source other than the respondent[1]. Commonly these variables concern factual information such as behaviors, events, time, places, quantities, procedures, demographic variables, and knowledge.

Table 16.2 presents the assertions for the different basic subjective and objective variables.

Table 16.2: The basic structures of assertions

Basic concepts	Structure 1 xIsc	Structure 2 xPy	Structure 3 xP
Subjective Variables			
Evaluation (e)	**xIe**	**—**	**—**
Importance (i)	**xIi**		
Values (v)	**vIi**	**—**	**—**
Feelings (f)	**xIf**	**xFy or xP f**	**—**
Cognitive judgment (c)	**xIc**	**—**	**—**
Causal relationship (ca)	**xIca**	**xCy**	**—**
Similarity relationship (s)	**xIs**	**xSy**	**—**
Preference (pr)	**xIpr**	**xPR y (z...)**	**—**
Norms	**—**	**oH (+I) y**	**oH(+I)**
Policies	**—**	**gH (+I) y**	**gH(+I)**
Rights (ri)	**xIri**	**xHRy**	**—**
Action tendencies	**—**	**rFDy**	**rFD**
Expectations of Future events	**— xFDy**	**xFD**	
Evaluative belief	**—**	**xP_e y or xPy_e**	**xP_e**
Objective Variables			
Behavior	**—**	**rDy**	**rD**
Events	**—**	**xDy**	**xD**
Demographics (d)	**xId**	**—**	**—**
Knowledge	**xIsc**	**xPy**	**xP**
Time	**—**	**—**	**xDti**
Place	**—**	**—**	**xDpl**
Quantities	**—**	**xDqu**	**—**
Procedures	**—**	**—**	**xDpl, pro**

Notations: x denotes the grammatical subject; **P**, the predicator; **I** the link verb. Frequently occurring subjects are the government (**g)**; anyone (**o)**; and the respondent himself (**r)**. Frequently employed lexical verbs for predicators are: **C**, that indicates relationships where the subject causes the object; **D**, that indicates deeds; **E**, that indicates expectations; **F**, that specifies feelings; **FD**, that indicates a predicator referring to future deeds; **H(+I)**, that specifies a predicator which contains words like "has to" or "should" followed by an infinitive; **HR**, that specifies predicators like "has the right to"; **J**, that specifies a judgment; **PR**, that indicates predicators referring to preferences; **S**, that indicates relationships of similarity or difference between the subject and the object.

[1]Although we refer to *objective variables*, that does not imply that they are exempted from measurement error. For instance, administrative records may have low quality as measurement instruments if their categories are poorly defined.

Imagine that we want to know the degree of importance that the respondents place on the value “honesty”. Then, Table 16.2 indicates that the structure we should use is structure 1 (vIi), where ‘v’ refers to the value of interest (honesty), ‘I’ is the link verb and ‘i' refers to the basic concept of importance. Following this structure, we can formulate several assertions:

Q2.1a- *Honesty is very important.*
Q2.1b- *Honesty is important.*
Q2.1c- *Honesty is unimportant.*
Q2.1d- *Honesty is very unimportant.*

For some concepts, different structures can be used to formulate the assertions.

Until now, we have focused on the basic structure of assertions. However, in reality, assertions have a lot of variations. They are expressed in sentences much longer than have been studied so far. Often indirect objects, modifiers, or adverbials are added to the simple sentences.

We have indicated how the most commonly applied basic concepts-by-intuition in survey research can be expressed in assertions specifying these concepts. These rules are summarized in Table 16.2. This table can be used to specify an assertion for a certain type of concept according to the criteria specified there. For example, if we want to specify an evaluation about immigrants, the structure of the sentence recommended is (xIe). Therefore, we can formulate a statement such as “immigrants are good people.” If we want a feeling (xIf), we can write “immigrants are in general friendly.” If we want a cognitive judgment (xIc), the statement can be: “immigrants are hard-working.” If we want to formulate a cognition concerning the reasons why immigrants come here, the structure is (xRy), and a possible assertion would be “Problems in their own country cause emigration to Europe.” In the same way, assertions can be formulated for any other concept.

Now that standard assertions have been specified for the basic concepts of the social sciences, the task of the researcher is to determine what type of basic concept his/her specific concept-by-intuition is. If that is done, sentences that represent with very high probability this concept can be formulated.

1.4 From assertion to request for an answer

The term "request for an answer" is employed, because the social science research practice and the linguistic literature (Harris 1978; Givon 1984; Weber 1993; Graesser, et al. 1994; Huddlestone 1994; Ginzburg 1996; Groenendijk & Stokhof 1997; Tourangeau et al. 2000) indicate that requests for an answer are formulated not only as requests (interrogative form) but also as orders or instructions (imperative form), as well as assertions (declarative form) that require an answer. Even in the case where no request is made and an instruction is given or a statement is made, the text implies that the respondent is expected to give an answer. Thus, the common feature of the formulation is that an answer is expected.

If an assertion is specified for a concept, a simple way to transform it into a request for an answer is to add a pre-request in front of the assertion. This procedure can be applied to any concept and assertion. Using Q2.1a-Q2.1d, to make a request from these assertions, pre-requests can be added in front of them, for example:

Q2.2a- *Do you think that honesty is very important?*
Q2.2b- *Do you think that honesty is important?*
Q2.2c- *Do you think that honesty is unimportant?*
Q2.2d- *Do you think that honesty is very unimportant?*

Using such a pre-request followed by the conjunction "that" and the original assertion creates a request called an *indirect request*. The choice of one of these possible requests for a questionnaire seems rather arbitrary as this specific choice of the request can lead the respondent in that direction. Therefore, a more balanced approach can be used:

> Q2.2e- *Do you think that honesty is very important, important, unimportant or very unimportant?*

In order to avoid such a sentence with too many adjectives, it is advisable to substitute them with a so called “WH word” like “how”:
Q2.2f- *How important, do you think, is honesty?*

This is also an indirect request with a pre-request and a sub-clause that starts with a WH word and allows for all the assertions Q2.1a-Q2.1d as an answer and other variations thereof.

Instead of indirect requests, *direct requests* can also be used. The most common form is an interrogative sentence. In this case the request can be created from an assertion by the inversion of the (auxiliary) verb with the subject component. The construction of direct requests by the inversion of the verb and subject component is quite common in many languages but also other forms can be used.

Let us illustrate this by the same example using only two of the four assertions mentioned:
Q2.1b- *Honesty is important.*
Q2.1c- *Honesty is unimportant.*

One can transform these assertions into direct requests by inverting the auxiliary verb and the subject:

Q2.3b- *Is honesty important?*
2.3c- *Is honesty unimportant?*

Here, the requests can be seen as “leading” or “unbalanced” because they have only one possible answer option. It could be expected that a high percentage of respondents would choose this option for this reason. Therefore, the requests can be reformulated as follows:
Q2.3e- *Is honesty important or unimportant?*

Thus, two basic choices have to be made for formulating a request for an answer: the use of direct or indirect requests and whether to use WH words. The combination of these two choices leads to four different types of requests.

Besides the interrogative form, the second grammatical form of a request for an answer possible is the imperative form. In its basic form the request consists only of an instruction to the respondent, as for example:

Q2.4 *Indicate how important honesty is for you:*

Example 2.4, colloquially known as an instruction or in grammatical terms as an "imperative", is another example of a direct request for an answer.

The third grammatical form, a declarative request, is only possible as an indirect request. Illustrations are examples Q2.5 and Q2.6. Both examples have a declarative pre-request, followed by a WH word and an embedded interrogative query:

Q2.5 *I would like to ask you how important honesty is for you.*

Q2.6 *Next we ask you how important honesty is for you.*

Although these are statements from a grammatical perspective, it is commonly understood that an answer to the embedded interrogative part of the sentence is required.

To conclude, different assertions and even more different requests for an answer can be formulated to measure concepts like "the importance of the value of honesty", as Figure 16.1 summarizes. However, it is important to note that whatever the request form used, all these requests measure what they are supposed to. Therefore, there is no real difficulty with making an appropriate request for a concept if the assertions represent the concept of interest well. For more details, we refer to Saris & Gallhofer (2014).

Figure 16.1: The different steps applied to the importance of the value honesty

Concept by intuition
Concept = importance. Value = honesty

Step 2: Assertions, structure 1 (vli)
Honesty is very important. Honesty is important. Honesty is unimportant. Honesty is very unimportant.

Step 3: Requests for an answer
Do you think that honesty is important? Do you agree with the following statement: “honesty is unimportant” How important do you think honesty is? Is honesty important? I would like to ask you how important honesty is.

2. How to deal with all the other decisions?

2.1 Many more decisions to complete the questions

So far we have discussed the most fundamental decisions that have to be made in order to create valid request for an answer, i.e. in order to be as sure as possible that the request formulated measures the concept-by-intuition that it is supposed to measure and nothing else.

There are, however, many more decisions to be made to create a complete survey question, about the introduction of other components of a question and the way they can be formulated. All these decisions lead to different forms of the final question and potentially to different results. Table 16.3 provides an overview of the characteristics of the questions that have to be taken into account.

Table 16.3: The characteristics of the questions to be taken into account

Group	Specific characteristic
The trait	Domain
	Concept
Associated to the trait	social desirability
	centrality of the topic
	time specification
Formulation of the request for an answer	trait requested indirectly, direct or no request and presence of stimulus (battery)
	WH word and what type of WH word
	Type of the request (interrogative, Imperative question-instruction, declarative or none (batteries).
	Gradation
	Balance of request or not
	Encouragement to answer
	Emphasis on subjective opinion
	Information about the opinion of other people
	Absolute or a comparative judgment
Characteristics of the response scale	Categories; yes/no answer scale; frequencies; magnitude estimation; line production and, more steps procedures.
	Amount or the number of categories
	full or partial labels
	labels with long or short text
	Order of labels
	Correspondence between labels and numbers
	theoretical range of scales (bipolar or unipolar)
	Range of scales used
	Fixed reference points
	Don't know option
Instructions	Respondent instructions
	Interviewer instructions
Additional information about the topic	Additional definitions, information or motivation
Introduction	Introduction and if request is in the introduction
Linguistic complexity	Number of sentences
	Number of subordinated clauses
	Number of words
	Number of nouns
	Number of abstract nouns
	Number of syllables
Method of data collection	
Language of the survey	
Characteristics of the show cards	Categories in horizontal or vertical layout
	Text is clearly connected to categories or if there is overlap
	Numbers or letters shown before answer categories
	Numbers in boxes
	Start of the response sentence shown on the show card
	Question on the show card
	Picture provided.

The table shows that a considerable number of decisions are made in the formulation of one question. Going on with the example of the importance of the value honesty, we can formulate quite different complete questions, even if we start from the same assertion "honesty is important":

Q3a- To what extent do you think that honesty is important?

Not important				Very important	Don't Know
1	2	3	4	5	

Q3b- Now we want to ask you questions about different values. Do you agree or disagree with the following statement: “honesty is important”. Please use this card to answer.

1 Agree strongly
2 Agree
3 Neither agree nor disagree
4 Disagree
5 Disagree strongly
Don't Know

Both questions differ at many different levels: presence of an introduction, of an instruction, formulation of the request, scale characteristics, etc. Because of the huge number of decisions and the potential interactions between them, one cannot easily evaluate the consequences of all these decisions on the quality of the question. In order to evaluate how good a question is, we need a more adapted tool: this is the program Survey Quality Predictor (Saris et al, 2011; Saris & Gallhofer, 2014) available for free at: http://sqp.upf.edu/.

2.2 The program SQP 2.0

What is behind SQP? On which knowledge is the program based?

One of the most common procedures to evaluate the quality of questions is the MultiTrait-MultiMethod (MTMM) approach. Indeed, Lance et al. (2009) searching a citation database of the seminal article by Campbell and Fiske (1959) found up to 4,338

citations spanning several disciplines with special emphasis in psychology They conducted a literature review of applied research that used the MTMM approach which included among others personality traits, labor studies, organizational research, mental and social health, social psychology, and organizational research. Campbell and Fiske (1959) proposed to repeat several “traits” (i.e. questions) using different “methods” (in our case it will be formulations using different characteristics) in order to study the convergent and discriminant validity. Later, the approach has been developed and formalized by using Confirmatory Factor Analysis models to analyze the MTMM correlation matrices (Werts & Linn, 1970; Jöreskog, 1970, 1971; Althauser, Heberlein & Scott, 1971; Andrews, 1984). Alternative models have been proposed and compared, with the conclusion (Corten et al., 2002; Saris & Aalbers, 2003) that the model of Alwin (1974) and the one of Saris & Andrews (1991) are the two that fit the best to several data sets.

SQP is based on thousands of quality estimates of questions obtained in more than 30 European countries and languages by MTMM analyses using the True Score model by Saris & Andrews (1991). This True Score model allows decomposing the quality in validity and reliability. Most of these MTMM experiments have been done in the ESS. In each ESS round, four to six MTMM experiments are included in almost all the participating countries. In each experiment, three traits are measured using three or four different methods. All respondents get a common method in the main ESS questionnaire and a repetition with a different method at the end in a supplementary questionnaire (split-ballot MTMM design as proposed by Saris, Satorra & Coenders, 2004). More than 25 minutes of similar questions separate the first question from its repetition in order to avoid potential memory effects (van Meurs & Saris, 1990). Reliability, validity and quality have been estimated for all these traits, methods and countries.

However, even if thousands of quality estimates are available (you can consult them in the SQP program), there are still much more questions for which MTMM experiments have not been implemented. It is not possible to repeat all the questions of every questionnaire twice. The costs and cognitive burden will be too high. Only a small subset of questions can be repeated each time that a survey is done. Therefore, the MTMM model can be used to estimate the quality of only a limited number of

questions. What about the quality of all the other questions that were not part of an MTMM experiment? What about new questions?

In order to get information about the quality of all possible questions, Saris & Gallhofer (2007) proposed to do a meta-analysis over the thousands of quality estimates available from all the MTMM experiments done in the past. The idea is to explain the quality estimates as a function of all the characteristics presented in Table 16.3. This has been done originally using a regression model (Saris & Gallhofer, 2007) but later the Random Forrest approach (Breiman, 2001) for prediction has been applied on the existing data (Oberski et al., 2011; Saris & Gallhofer, 2014).

Once the impact of these characteristics on the reliability and validity is known and is good enough (R^2 around .8), this information can be used to produce a prediction of the quality of new questions. This is what the program SQP does in a user-friendly way as illustrated in the next section.

What can you do with SQP?

The program SQP makes predictions of the quality of questions on the basis of information about the choices that have been made with respect to the questions characteristics mentioned in Table 16.3.

If you have to create new questions, in order to evaluate how good the questions you proposed are and how you could improve them, you can go to the SQP program. There, you have to code within the program the questions with respect to all the characteristics mentioned in Table 16.3. When you finish, the program provides a prediction of the reliability, validity[2], and quality of the questions, together with confidence intervals for the predictions. Besides, it provides suggestions about how changing one or another of the characteristics would affect the quality. In that way, questions can be improved.

[2] Note that in this case the validity is defined as 1-method variance. We assume that the question measured what is supposed to measure. So, here we just look at the invalidity due to the effect of the reaction of the people to the chosen method on the observed responses.

Each time that a user codes a question and obtains a quality prediction, this information is stored and accessible for other users. Therefore, SQP will provide a growing database of survey questions with quality information over the years. If one does not trust the prediction by another user, one can recode the question to get a new prediction. But if one trusts the prediction, he/she could use it directly.

The predictions of reliability, validity and quality can then be used to improve the questions before going to the field, but also to control the comparability of questions translated in different languages in a cross-national or cross-cultural setting as will be discussed in section 4. Finally, they can be used to correct for measurement errors in studies of relationships between variables (Saris & Revilla, 2013; De Castellarnau & Saris, 2014).

Limits of the scope of the program

SQP 2.0 covers a large number of countries (more than 25). However, the program is mainly based on data from European countries, collected by face-to-face or self-completion, and about attitudinal questions. Thus, it is more appropriate to use the SQP predictions when one is interested in these countries or countries culturally similar (e.g. SQP may not give a good prediction for Asiatic countries), this mode of data collection, and kinds of questions. Besides, the program considers only some aspects of the context (level of social desirability and sensibility), so if one is interested in very specific contexts, this might not be taken into account in the prediction. Nevertheless, we should notice than the context affects usually more the answers themselves than the quality of these answers.

3. Getting quality predictions of your questions using SQP: an illustration

At the end of the 3-step procedure, we had formulated theoretically valid request for answers. Then, we had to make decisions about how to complete these requests in order to get complete questions. Using the example of the importance of the value honesty, we proposed two forms of complete questions that should be theoretically valid. However, in order to know if they are good questions, we need to estimate their quality.

In this section, we illustrate how this can be done using the program SQP to predict their quality. Both Q3a and Q3b are measuring the same concept-by-intuition. Table 16.4 repeats the two forms in a way that underlines the main differences across them: for instance, Q3b has an introduction but Q3a has not, Q3b has an instruction for the respondent whereas Q3a has not, Q3a has an horizontal scale whereas Q3b has a vertical one, the scale in Q3a is partially labeled with whereas in Q3b it is fully labeled. Besides, in Q3a the scale is specific to the item measured whereas in Q3b it is an Agree-Disagree scale. Overall, these two formulations differ on many characteristics that may affect the way respondents answer the question and therefore the quality.

Table 16.4: Two survey questions for a concept-by-intuition

Question	Introduction	*Request for Answer*	*Answer options*
Q3a		To what extent do you think that honesty is important?	Not important 1 2 3 4 5 Very important; Don't Know
Q3b	Now we want to ask you questions about different values.	Do you agree or disagree with the following statement: "honesty is important". Please use this card to answer.	1 Agree strongly 2 Agree 3 Neither agree nor disagree 4 Disagree 5 Disagree strongly Don't Know

SQP asks the coder information about all the characteristics listed in Table 16.3. Appendix 16.1 provides the code for the two questions. Once all the characteristics are coded, SQP shows the quality predictions and information about the uncertainty of the predictions (standard errors and interquartile ranges). English is used for this example but the program allows users to get quality predictions of survey questions in more than 20 languages. Table 16.5 gives the prediction for Q3a and Q3b.

Table 16.5: Quality predictions in SQP

	Q3a		**Q3b**				
	Prediction	**(IR)**	**Prediction**	**(IR)**	**Total quality**	**Q3a**	**Q3b**
Reliability Coeff.(r)	0.805	(0.734, 0.866)	0.798	(0.718, 0.850)	**Reliability (r^2)**	0.649	0.636
Validity Coeff. (v)	0.964	(0.897, 0.990)	0.962	(0.906, 0.989)	**Validity (v^2)**	0.928	0.925
Quality Coeff. (q)	0.776	(0.656, 0.817)	0.767	(0.647, 0.808)	**Quality (q^2)**	0.602	0.588

Note: IR: Interquartile range; Coeff.: coefficient

This table shows that even if both questions vary on many characteristics, the overall quality is very similar. It is slightly higher in Q3a but if we consider the interquartile ranges, we see that the incertitude about the prediction estimates is much larger than the observed difference between Q3a and Q3b. In this illustration, both questions are therefore similarly good.

However, it also shows that questions Q3a and Q3b are not perfect: the measurement errors are still quite large (quality much lower than one). Therefore, we can try to improve it. For that, SQP also provides suggestions of potential improvements of the questions that will increase the measurement quality.

In SQP, users get suggestions over 20 characteristics that have had the largest impact in the quality predictions over all questions in the database. It is also possible to ask the program to evaluate not only these first 20 variables but all the 60 characteristics that are included in the model to predict measurement quality.

If we concentrate on suggestions for Q3a, the variable that potentially leads to a largest change in quality is the number of categories (from 7 categories onward, the increase in quality is 0.049) and the number of fixed reference points. The program also suggests avoiding showing explicitly the “Don’t know” option. Following these suggestions, we propose an improved question, presented in Table 16.6.

Table 16.6: An improved question for the same concept-by-intuition

Question	Introduction	*Request for Answer*	*Response scale*
Q3a-bis		To what extent do you think that honesty is important?	Not at all important 0 1 2 3 4 5 6 7 8 9 Extremely important 10

However one should realize that each suggestion is done while all the other characteristics remain the same. This is in general not true. Changing characteristics may affect the linguistic complexity of the items and this will have an effect on the quality prediction. For instance, by introducing here fixed reference points using “extremely”, the number of syllables for the labels changes, and this increase in syllables can have a negative effect on the quality. In order to check if by introducing one suggested improvement the quality of the new question really increases, one has to

reformulate the question and code the new question in SQP in the same way as before. The results are shown in Table 16.7.

Table 16.7: Quality predictions for Q3a and Q3a-bis

	Q3a	Q3a-bis
Reliability (r^2)	0.649	0.707
Validity (v^2)	0.928	0.932
Quality (q^2)	0.602	0.659

The quality of Q3a-bis is higher than the one of Q3a by 0.057. The improvement is mainly coming from an increase in the reliability. This shows that the suggestions proposed by the program in the response scale helped to formulate a better question. However, we did not manage to get a very high quality neither. Therefore, correction for measurement errors after the data collection will be really crucial to draw correct conclusions about standardized relationships across variables.

4. Good questions for cross-national research

In order to create a good question, we started by formulating theoretically valid requests for an answer. Then, we made complete questions based on these simple requests for an answer and use SQP to evaluate how good different complete questions were. We got to the point where we have a theoretical valid question with acceptable quality and we have improved it as much as we could using SQP. But in the frame of cross-national research, a good question needs also to be comparable across countries. Therefore, there is one more level to consider in this case. This is the object of this last section.

Cross-national research requires that the questions used are functionally equivalent across countries, meaning that the message embedded in a text is received by the receptor in the same way as it would be received in the source language (Nida, 1964). Survey questions are not comparable with respect to quality if formal characteristics of the questions are different. Functional equivalence in cross-cultural survey research is only confirmed by formally testing invariance. Several studies have identified translation deviations as a source of non-equivalence in assessments of survey data (Hambleton et al., 2005; Harkness et al., 2010a; Mallinckrodt & Wang, 2004; Oberski et al., 2007; Saris & Gallhofer, 2007; Van de Vijver & Leung, 1997; Zavala Rojas &

Dorer, 2013). Unfortunately non-invariance was detected once data was collected and survey organizations had already spent a lot of resources in data collection. Empirical methods are mostly used for detecting flaws once data is already collected (Horn & McArdle, 1992; Meredith, 1993; Steenkampt & Baumgarter, 1998; Vandenberg & Lance, 2000; Braun & Johnson, 2010; Byrne & Van de Vijver, 2010; Saris & Gallhofer, 2014).

Translation guidelines suggest that a good translation aiming at functional equivalence would avoid changing deliberately other semantic components than those necessary because of language differences. Questions in different languages should keep the concepts the same; preserve the item characteristics and maintain the intended psychometric properties (Harkness, 2003; Harkness et al., 2003; Harkness et al., 2010b).

In order to preserve the same meaning of concepts across languages, the state-of-the-art suggests a multi-step committee approach to translate survey items (Harkness et al., 2003; Harkness et al., 2010b). However, it is difficult to compare systematically that throughout a questionnaire item characteristics in different languages are the same (such as domain, concept, wording, response scale, polarity, labeling, symmetry, balance of the request, introduction, instructions, and linguistic complexity). As these characteristics define the measurement properties of survey items, a way to prevent non-equivalence is to compare them systematically before data collection. As SQP asks users to code a large set of properties of a survey item, we propose a procedure aiming to detect deviations in translations by comparing the codes of a source questionnaire and target languages. This procedure is complementary to the translation committee approach (which focuses in semantic similarity). As one cannot be familiar with all languages participating in a cross-cultural project, the coding scheme in SQP allows that trained individuals in survey research and proficient in the respective languages provide information about item characteristics. t Once characteristics of the source and translated versions are coded, comparing the codes allows to detect in a systematic way deviations across language versions.

Characteristics of translated items can be compared using SQP in a five-step procedure.

Step 1: Introducing questions in SQP

Each question in the source and target languages should be introduced into the program SQP. This can be done by any user at no cost after signing up and logging in the program at sqp.upf.edu webpage. When coding, the program displays a help option on each screen indicated by a yellow box, which defines each item characteristic asked and gives examples.

Step 2: Coding the source questionnaire

The information regarding the item characteristics of the source questionnaire must be accurate because target versions will be compared against it. It should be coded independently by two individuals with deep knowledge about questionnaire design; differences should be reconciled in collaboration with a third individual which plays the role of a reviewer.

Step 3: Coding a target questionnaire

The translated questionnaire should be coded by a proficient speaker of the target language, preferably someone involved in the translation process.

Step 4: Comparison of measurement properties

The codes of the characteristics of source items should be compared with those in the target language. Any difference should be clarified with coders, first, to rule out coding errors in the target questionnaire. True differences in the codes should be reported to the translation team.

Step 5: Interpretation of deviations and actions taken in the target text

The translation team should clarify any difference in the codes in terms of the definition of the features. In other words, it should justify the reasons behind a deviation in the item characteristics.

The differences may fall into one of three categories shown in Table 16.8. For each category, an action is suggested for the translated text.

Table 16.8: Categories for differences in the SQP codes for two languages

Type of deviations found (source vs. translation)	Action taken
A) A difference that cannot be warranted, for instance a different number of response categories, leaving out a "don't know" option or/and an instruction for the respondent.	The translation should be amended
B) A difference that may or may not be warranted e.g. use of complete sentences in the scales instead of short texts. In some languages it is necessary, in some others this may be a fact of stylistic choice	Amendments in the translation are recommended to keep the principle of functional equivalence in translation if the language structure allows keeping the item characteristic the same.
C) A difference in the linguistic characteristics that may be warranted e.g. different number of words, syllables. Also, a difference in the codes of linguistic characteristics that may not be warranted e.g. different number of sentences, nouns, extreme deviations in the number of words.	Amendments in the translation are recommended to keep the principle of functional equivalence in translation if the language structure allows it. If the differences are unavoidable due to linguistic characteristics, no change is recommended.

Source: Zavala-Rojas, D., & Saris, W. (2014). "A procedure to prevent differences in translated survey items using SQP."

This procedure was tested for the first time in the ESS Round 5 and has become part of the specifications of the survey design from Round 6 onwards. This process has helped to prevent deviations such as: different formulations for a *repetition*, *m*issing introductions, *e*xtra explanations making the item more complex, *m*issing definitions of the scale, *devi**ations in the scale formulation*, *f*ixed reference points, *l*abels *about* bipolar *concepts* defined as unipolar *concepts, c*omplete sentences instead of short texts for labels, among many others.

5. Conclusion

To conclude, creating good questions requires a lot of efforts. First, the researchers have to make sure that the questions really measure the concepts of interest, i.e. that the questions are theoretically valid. This can be done by following the three-step procedure proposed by Saris & Gallhofer (2014). In this way, requests for answers that measure the concepts of interest can be formulated for any concept.

However, there are many more decisions to take in order to formulate a complete question: using or not an introduction or an instruction, defining the answers categories,

etc. Each of these decisions can influence the quality of the question in different ways and it is difficult to evaluate really how good a complete question is depending on the choices made without an appropriate tool. However, it is also essential to do so, before the data collection, in order to collect the highest quality data possible, and after data collection, in order to have the information to correct for the remaining measurement errors.

The tool we propose to evaluate the quality of survey questions is the program SQP. This program is based on a meta-analysis of thousands of quality estimates obtained by estimating MTMM models. It uses the cumulative information of the past MMM research in order to predict the quality estimates of new questions depending on their characteristics.

This program is available for free and gives researcher a powerful tool in order to evaluate how good their questions are. It also gives suggestions to improve the questions.

Finally, in the frame of cross-national research, a good question is also a question that can be used in different languages and give comparable answers, such that observed differences indicate true differences across countries and not differences in the questions. One point to take into account to facilitate measurement equivalence across different languages is that no unnecessary changes are made during the translation process for characteristics that are not directly language related. For instance, if fixed-reference points are used in a language, they should also be used in the other ones. The SQP program can be and has been used to compare for a given question the characteristics in different languages and detect potential deviations introduced during the translation process.

One limit of this approach is that not all existing languages are covered so far. Still, the program covers around 20 languages currently, including almost all the European ones. Therefore, it is a powerful tool to create good questions. But it has to be used only once the theoretically valid requests for answers have been formulated.

References

Althauser R.P., T.A. Heberlein, & R.A. Scott 1971. A causal assessment of validity: The augmented mulitraitmultimethod matrix. In H. M. Blalock Jr. (ed.), *Causal Models in the Social Sciences*. Chicago: Aldine, 151–169.

Alwin D.F. (1974). An analytic comparison of four approaches to the interpretation of relationships in the multitraitmultimehod matrix. In H. L. Costner (ed.), *Sociological Methodology*, San Francisco: Jossey Bass, 79–105.

Alwin D.F., & J.A. Krosnick (1991). The reliability of survey attitude measurement. The influence of question and respondent attributes". *Sociologic Methods and Research*, 20:139–181.

Andrews, F. (1984). Construct validity and error components of survey measures: A structural modeling approach. *Public Opinion Quarterly*, 46:409–442.

Bradburn N.M., & S. Sudman (1988). *Polls and Surveys. Understanding what they tell us.* San Francisco: Jossey Bass.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1): 5–32.

Braun, M., & Johnson, T.P. (2010). An Illustrative Review of Techniques for Detecting Inequivalences. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp:73–393). John Wiley & Sons, Inc. doi:10.1002/9780470609927.ch20

Byrne, B.M., & van De Vijver, F.J.R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, 10(2), 107–132.

Campbell, D.T., & D.W. Fiske (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, 6:81–105.

Corten, I.W., Saris, W.E., Coenders, G.M., van der Veld, W., Aalberts, C.E., & C. Kornelis (2002). Fit of different models for multitrait–multimethod experiments. *Structural Equation Modeling: A Multidisciplinary Journal*, 9:213–232.

De Castellarnau, A., & W.E. Saris (2014). A simple way to correct for measurement errors. European Social Survey Education Net (ESS EduNet). Available at: http://essedunet.nsd.uib.no/cms/topics/measurement/

De Groot, A.D., & F.L. Medendorp (1986). *Term, Begrip, Theorie: Inleiding tot Signifische Begripsanalyse*. Meppel: Boom.

Ginzburg J. (1996). Interrogatives: Questions, facts and dialogue. In S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*. Cambridge MA: Blackwell, 385–421.

Givon T. (1984). Syntax. *A Functional– Typological Introduction* Vol. I–II. Amsterdam: J. Benjamin.

Graesser A.C., C.L. McMahen, & B.K. Johnson (1994). Question asking and answering. In M. Gernsbacher (ed.), *Handbook of Psycholinguistics*, San Diego, CA: Academic Press, 517–538.

Groenendijk, J., & M. Stokhof (1997). Questions. In J. van Benthem and A. ter Meulen (eds.), *Handbook of Logic and Language*, Amsterdam: Elsevier, 1055–1124.

Hambleton, R.K., Merenda, P.F., & Spielberger, C.D. (2005). Adapting educational and psychological tests for cross-cultural assessment. *Psychology Press*.

Harkness, J.A. (2003). Questionnaire translation. In J. A. Harkness, F.J.R. van de Vijver, & P.P. Mohler (Eds.), *Cross-cultural survey methods*. Hoboken: Wiley & Sons.

Harkness, J.A., Edwards, B., Hansen, S.E., Miller, D.R., & Villar, A. (2010). *Designing Questionnaires for Multipopulation Research. In Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 31–57). John Wiley & Sons, Inc. doi:10.1002/9780470609927.ch3

Harkness, J.A., Vijver, F.J. R. van de, & Mohler, P.P. (2003). *Cross-cultural survey methods*. Hoboken: Wiley & Sons.

Harkness, J.A., Villar, A., & Edwards, B. (2010). Translation, Adaptation, and Design. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp:115–140). John Wiley & Sons, Inc. doi:10.1002/9780470609927.ch7

Harris, Z. (1978). The interrogative in a syntactic framework . In Hiz H. (ed.), *Questions*. Dordrecht: Reidel, 37–89.

Horn, J.L., & McArdle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3-4):17–44. doi:10.1080/03610739208253916

Hox J.J. (1997). From theoretical concept to survey questions. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.), *Survey Measurement and Process Quality*. New York: Wiley, 47–70.

Huddleston, R. (1994). The contrast between interrogatives and questions. *Journal of Linguistics* 30, 411–439.

Jöreskog, K.G. (1970). A general method for the analysis of covariance structures. *Biometrika*, 57:239–251.

Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36:109–133.

Koltringer R. (1993). *Gültigkeit von Umfragedaten*. Wien: Bohlau.

Koning P.L., & P.J. van der Voort (1997). *Sentence Analysis*. Groningen: Wolters–Noordhoff.

Lance, C. E., Baranik, L. E., Lau, A. R., & Scharlau, E. A. (2009). "If it ain't trait it must be method". In Charles E. Lance, Robert J. Vandenberg (ed.) *Statistical and methodological myths and urban legends*, Taylor & Francis, p. 337-60.

Mallinckrodt, B., & Wang, C.-C. (2004). *Quantitative Methods for Verifying Semantic Equivalence of Translated Research Instruments: A Chinese Version of the Experiences in Close Relationships Scale.*

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4):525–543. doi:10.1007/BF02294825

Nida, E.A. (1964). *Toward a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating* (p:331). Brill Archive.

Oberski D., T. Grüner & W.E. Saris (2011). The prediction procedure the quality of the questions based on the present data base of questions. In Saris W.E, Oberski, D., Revilla, M., Zavalla, D., Lilleoja, L., Gallhofer, I. and T. Grüner (2011). "The development of the Program SQP 2.0 for the prediction of the quality of survey questions". *RECSM Working paper 24. (Chapter 6)* Available at: http://www.upf.edu/survey/_pdf/RECSM_wp024.pdf

Oberski, D., Saris, W. E., & Hagenaars, J. (2007). Why are there differences in measurement quality across countries. Measuring Meaningful Data in Social Research. Acco, Leuven.

Oppenheim, A.N. (1966). *Questionnaire Design and Attitude Measurement*. London: Heinemann.

Richards J., C.J. Platt, & H. Platt (1993). *Dictionary of Language Teaching and Applied Linguistics*. Harlow: Longman.

Saris, W.E., & C. Aalberts (2003). Different explanations for correlated disturbance terms in MTMM studies. *Structural Equation Modeling: A Multidisciplinary Journal*, 10:193–213.

Saris, W.E., & F.M. Andrews (1991). Evaluation of measurement instruments using a structural modeling approach. In P. P. Biemer, R. M. Groves, L. Lyberg, N.

Mathiowetz, & S. Sudman (Eds.), *Measurement errors in surveys* (pp.575–597). New York, NY: Wiley.

Saris, W.E. & I. Gallhofer (2007). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. New York:Wiley. doi:10.1111/j.1751-5823.2008.00054_20.x

Saris, W.E. & I. Gallhofer (2014). *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. New York:Wiley (second edition)

Saris W.E, Oberski, D., Revilla, M., Zavalla, D., Lilleoja, L., Gallhofer, I. & T. Grüner (2011). The development of the Program SQP 2.0 for the prediction of the quality of survey questions. *RECSM Working paper 24.* Available at: http://www.upf.edu/survey/_pdf/RECSM_wp024.pdf

Saris, W.E., & M. Revilla (2013). Correction for measurement errors in survey research: necessary and possible. *RECSM Working Paper 31*. http://www.upf.edu/survey/_pdf/RECSM_wp031.pdf

Saris, W.E., Satorra, A., & G. Coenders (2004). A new approach to evaluating the quality of measurement instruments: The split-ballot MTMM design. *Sociological Methodology*, 34:311–347.

Scherpenzeel, A.C., & W.E. Saris (1997). The validity and reliability of survey questions: A meta-analysis of MTMM studies. *Sociological Methods and Research*. 25:341–383.

Schuman, H., & S. Presser (1981). *Questions and Answers in Attitude Survey: Experiments on Question Form, Wording and Context*. New York: Academic Press.

Smith, T.W. (1987). The art of asking questions 1936-1985. *Public Opinion Quarterly,* 51, 95–108.

Steenkamp, J.-B. E.M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research,* 25(1):78–107.

Sudman, S., & N.M. Bradburn (1983). *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey Bass.

Tourangeau R., L.J. Rips, & K. Rasinski (2000). *The Psychology of Survey Response*. Cambridge MA: Cambridge University Press.

Vandenberg, R.J., & Lance, C.E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, 3(1):4–70. doi:10.1177/109442810031002

Van Meurs, L., & W.E. Saris (1990). Memory effects in MTMM studies. In W.E. Saris & L. van Meurs (Eds.), *Evaluation of measurement instruments by meta-analysis of multitrait–multimethod studies*. Amsterdam, Netherlands: North Holland.

Vijver, F.J.R. van de, Leung, K., Fons J.R., van de V., & Kwok, L. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA, US: Thousand Oaks, CA, US: Sage Publications, Inc.

Weber, E.G. (1993). *Varieties of Questionsin English Conversation*. Amsterdam: J. Benjamin.

Werts, C.E., & R.L. Linn (1970). Path analysis: Psychological examples. *Psychological Bulletin*, 74:194–212.

Zavala-Rojas, D., & Saris, W.E. (2014). A procedure to prevent differences in translated survey items using SQP. Available at: http://www.upf.edu/survey/_pdf/RECSM_wp038.pdf

Zavala Rojas, D. & Dorer, B. (2013). A Mixed Method Approach to Link Translation Deviations and Measurement Quality. *Mimeo*.

Appendix 16.1: SQP prediction codes

Characteristic	Choice Q3a	Choice Q3b
Domain	Other beliefs	Other beliefs
Domain: other beliefs	Yourself	Yourself
Concept	Importance of something	Importance of something
Social Desirability	A lot	A lot
Centrality	Rather central	Rather central
Reference period	Present	Present
Formulation of the request for an answer: basic choice	Indirect requests	Indirect requests
WH word used in the request	WH word used	Request without WH word
'WH' word	How (extremity)	-------
Request for an answer type	Interrogative	Interrogative
Use of gradation	Gradation used	Gradation used
Balance of the request	Balanced or not applicable	Balanced or not applicable
Presence of encouragement to answer	No particular encouragement present	No particular encouragement present
Emphasis on subjective opinion in request	Emphasis on opinion present	No emphasis on opinion present
Information about the opinion of other people	No information about opinions of others	No information about opinions of others
Use of stimulus or statement in the request	No stimulus or statement	Stimulus or statement is present
Absolute or comparative judgment	An absolute judgment	An absolute judgment
Response scale: basic choice	Categories	Categories
Number of categories	5	5
Labels of categories	Partially labeled	Fully labeled
Labels with long or short text	Short text	Short text
Order of the labels	First label negative or not applicable	First label positive
Correspondence between labels and numbers of the scale	High correspondence	Low correspondence
Theoretical range of the scale bipolar/unipolar	Theoretically unipolar	Theoretically unipolar

Number of fixed reference points	1	1
Don't know option	DK option present	DK option present
Interviewer instruction	Absent	Absent
Respondent instruction	Absent	Present
Extra motivation, info or definition available?	Absent	Absent
Introduction available?	Not available	Available
Number of sentences in introduction	-------	1
Number of words in introduction	-------	10
Number of subordinated clauses in introduction	-------	0
Request present in the introduction	-------	Request not present
Number of sentences in the request	1	3
Number of words in request	10	18
Total number of nouns in request for an answer	2	3
Total number of abstract nouns in request for an answer	2	1
Total number of syllables in request	14	26
Number of subordinate clauses in request	1	0
Number of syllables in answer scale	8	13
Total number of nouns in answer scale	0	0
Total number of abstract nouns in answer scale	0	0
Show card used	Yes	Yes
Horizontal or vertical scale	Horizontal	Vertical
Overlap of text and categories?	Text clearly connected to category	Text clearly connected to category
Numbers or letters before the answer categories	Numbers	Numbers
Scale with numbers or numbers in boxes	Only numbers	Only numbers
Start of the response sentence on the showcard	No	No

Question on the showcard	No	No
Picture on the card provided?	No picture provided	No picture provided
Computer assisted	Yes	Yes
Interviewer	Yes	Yes
Visual presentation	Oral	Oral
Position	15	15