The Belgian Risk Perception Barometer Risk Perception Measuring Instruments Revisited

Koen Van Aeken
University of Antwerp
and PISA, SCK-CEN
BELGIUM
http://www.sckcen.be/pisa

Benny Carlé & Frank Hardeman
PISA, SCK-CEN
BELGIUM

1. Setting the scene

From the late 1980s on, societal perspectives have claimed – and have been awarded – recognition in the field of risk assessment. As a scholarly discipline devoted to the identification and quantification of risk [1], risk assessment had originally developed itself as an autonomous science which was dominated by probabilistic theory and confinement to all things numerical [2]. With the majority of hazards rooted in a sphere of industrial applications, like nuclear energy or chemical engineering, it is easy to understand the original, 'hard science' approach to risk assessment. The engineers, natural scientists and mathematics opened the same toolbox to evaluate the risks as the one they used to create the applications from which these risks emerged.

The majority of people, however, use intuitive judgment to estimate the characteristics of a risk [3]. Experts use a language that could as well be the speech of aliens, and with their arguments, they are often found to be lined up against the ordinary citizen [4]. These citizens, furthermore, are not 'atomistic', isolated entities evaluating a risk on a purely individual basis; on the contrary, the assessment of risks takes place in a larger framework that Kasperson and other researchers dubbed the social amplification of risk [5]. Apart from the inclusion of subjective and social factors in the analysis of risk, the concept of risk itself emancipated from the industrial and technological sphere to the social sciences, emphasized by the publication of Ulrich Beck's seminal Risk Society [6]. These factors not only contributed to the appreciation of the societal component in the study of risk – encompassing, amongst other, social, psychological and economic elements – but it also boosted the popularity of the discipline. Risk assessment is a fashionable activity these days. Its popularity is not only illustrated by a tremendous rise in scholarly publications, but its appeal also becomes apparent through the interest of various governments, lobbying groups, stakeholders, NGOs and the likes in the scope of applications of risk assessment.

The recognition of the societal dimension of risk assessment has been at the cradle of the opinion research on risks. Since risk estimates are not fixed by experts anymore, but are considered to show variation across a diverse population, the people themselves must be asked how they experience the risks.

The social sciences have formalized the 'asking about' in various ways. One way is to ask a limited number of people about their views, which generates in-depth, detailed and very rich data. So-called focus groups are a common method of this qualitative approach. The richness of the data, however, is countered by very limited representativity. The other main approach, the quantitative research method, mobilizes armies of interviewers which pose simple questions to hordes of people. A well known example is the Eurobarometer, this year celebrating its 30th birthday. It measures on a regular basis the public opinions on a wide range of topics in the 25 European member states [7]. While we definitely acknowledge the merits of the qualitative method, this contribution is concerned with the latter approach.

Following the rise in popularity of risk assessment and the recognition of its 'human' dimension, the demand for public opinion surveys on risks has been increasing at a fast pace. Unfortunately, this high demand sees some negative consequences. First, surveys are frequently conducted by people lacking even a minimal knowledge of survey methodology. In this respect, we might think of a journal or a newspaper trying to impress their readers with the definitive public opinion poll about the latest issue in vogue. Second, time pressure causes experienced or trained researchers to lower themselves to 'quick & dirty' work.
While methodologically flawed opinion research might not be something to worry about when appearing in the amusement press, concern is due if the results of a survey inspire policy development. Indeed, when public opinion research is conceived as an instrument to support rational, evidence based public policy, the strictest methodological standards should be applied, even if it is clear that scientific research will never substitute political reasoning [8].
This contribution deals with the safeguarding and enhancing of the quality of large scale surveys focusing on risk perception and related issues. This attention is relevant, not only for the reason that methodological standards may be flawed due to the immense popularity of the opinion poll, but also because the results of opinion surveys may have far-reaching policy consequences.

2. Pitfalls and challenges of large scale surveys of risk perception
Social science students, generally minimally interested in mathematics, are often horrified by the numerous courses of statistics and research methods that are obligatory in the curriculum. It makes them realize that sociology and other human sciences may well study 'soft' themes, but the research tools itself may end up to be as 'hard' as they are in the context of engineering or natural sciences. This contribution does not aim to deliver a condensed textbook in the methodology of social science data collection and analysis. The number of volumes devoted to this matter is already gargantuan. Instead, this text intends to highlight some major points of attention that are often overlooked in the practice of risk perception research by means of statistical surveys.
Once the main pitfalls and challenges are identified, a number of solutions and alternative approaches are offered. These challenges and subsequent solutions are predominantly inspired by the Belgian Risk Perception Barometer studies by the

Centre for the Study of Nuclear Energy (SCK-CEN) [9], occasionally referring to the Eurobarometer or other public opinion measuring instruments. It is important to notice that the Belgian barometer is constructed by means of a CAPI, short for Computer Assisted Personal Interviewing, better known as face-to-face interviews, whereby the answers are directly stored in a portable computer [10].

2.1. Social desirability

Some risks refer to activities that are of a private, intimate nature. Questions about risks regarding sexual behaviour or the use of illicit drugs may be answered in line with the informal social control that stipulates what is allowed or reasonable. In the mind of the respondent – and why not, in reality – deviance from these informal rules is frowned upon by the interviewer, possibly causing respondents to provide incorrect answers. While sex and drugs are exotic illustrations, the social desirability phenomenon may also play in the field of the perception of risks involved with nuclear energy and environmental issues, especially when is polled for the willingness to pay for a cleaner environment. It does not sound too good if one is not willing to pay 15$ a month to reduce emission of greenhouse gasses [11] This effect is a typical attribute to CAPI, but also exists in CATI. It is ruled out almost completely in CASI, short for Computer Assisted Self Interviewing (the digital analogue for a written survey that is filled in on the internet or returned by e-mail), but this method suffers from low response.

2.3. Interviewer bias

Personal interviewing implies social interaction. The way a question is asked can easily influence the answer. Variations in voice and tone, could provoke different answers regardless of the content of the question. This bias, which is often of an unconscious nature, generally reflects the preferences of the interviewer. A pronuclear interviewer could present the answering alternatives in such a way that an option favouring nuclear would appear the most desirable. Applying the theory of social desirability to a microsocial interaction between two persons, the respondent may then become inclined to answer what is deemed proper. While this effect only takes place if certain psychological conditions are fulfilled, it is not uncommon, certainly if the respondent is not familiar with the topics questioned.

2.4. Interview rush

Being only human, interviewers sometimes 'rush through' the interview. While the average duration of an interview for the Belgian risk barometer of 2002 amounted to 35 minutes with a standard deviation of 18 minutes, a handful of collaborators of the research agency recorded interview times of a mere four to nine minutes – leaving no blanks in the questionnaire. This was, of course, impossible. We suspect that some interviewers 'invented' a respondent and typed in the answers themselves, unaware of or not frightened by the fact that the interviewing software recorded the beginning and end times of each interview. While this is clearly fraud, a more common problem is the fact that interviewers may rush through the interview, since they are paid per interview and not per hour. Rushing through provokes unreliable answers, since the conditions for the respondent to calmly

Interviews with young women lasted at least 30 minutes, stretching to over an hour in the case of girls in their early 20s. For older female respondents or males, the interview time did hardly exceed ten minutes. It is self-evident that the obtained results should be treated with much skepticism.

2.5. Respondent reliability
Long interviews may cause the respondent to get distracted or bored, undermining the reliability of the answers. Difficult topics are another source of unreliable answers, since respondents might guess so as not to lose their face in front of the interviewer. Questions that might reveal a deviant position of the interviewer *vis-à-vis* informal social rules can sometimes yield false answers (see social desirability).

2.6. Lack of hypotheses
Social science research can be roughly divided into exploratory or confirmatory studies. The former 'explore' a part of the world without having specific goals or questions in mind. Just as an explorer who sails the world looking for things unknown, the researcher collects data and analyses them in the hope of finding meaningful correlations, patterns, relations and so on. Confirmatory research departs from well-defined research hypotheses, aiming at the verification or falsification of these presumptions. Exploration is primarily connected with qualitative analysis, whereas hypothesis-based research is a characteristic of the quantitative approach. Opinion surveys which do not depart from hypotheses risk ending up in a conflict between volume and nature of the data. While the amount of figures is abundant, their applications are often limited to mere description.

2.7. Single items
Many, not to say most barometers are based on items that consist of single questions. In the development of the questionnaire for the 2006 barometer, we discussed the issue of measuring the acceptance of legal norms in the field of food safety. One way of asking would be to present the following statement: 'Legal norms safeguard the quality of our food', which should be answered to on a five points scale. However, the sociology of law teaches us that the acceptance of legal norms is a multidimensional variable. Acceptance depends on the content of the norm (do people think the limits make sense?), the perceived enforcement (profound inspections secure the working of the norm) and the perception of the legitimacy of the norm giver (how much authority does the norm giver possess?). The single item above would thus yield information that is limited to merely one dimension of acceptance. Another pitfall with single issues consists in the varying interpretations that people adhere to terms or concepts. Suppose one asks to judge the risk of 'nuclear terrorism'. For some, nuclear terrorism would involve attacking an installation, like a power plant or research centre. Others would conceive terrorism as an act in which a dirty bomb is deployed that is composed out of stolen nuclear material, such as waste or plutonium from a research institute. Still others

would think of terrorists thieving nuclear weapons and posing a global threat, inspired by Hollywood fiction. It is self-evident that the risk perception of these scenarios varies according to the interpretation of the risk in question.

A third difficulty with single items consists in the limited number of statistical techniques that can be applied to the data. Indeed, in the common case of five points scales, single items yield variables of the ordinal measurement level. With answers ranging from one to five, the minimal prerequisite for metric (interval) measurement level is not fulfilled, since the rule of thumb prescribes at least seven categories to approximate the metric interval. This severely limits the scope of analysis. The use of multiple regression analysis, analysis of variance, or structural equation modeling, to name but a few of the principal methods, is not justified since one of the main assumptions is violated [12].

2.8. Wording
A common sin appearing in survey interviews is the use of suggestive wordings, difficult terminology, nesting of two questions in one phrase, or the lack of variation in positive and negative statements leading to so-called response set. The latter problem appears when, e.g., a long list of statements is offered whereby all statements are put in a positive (or negative) way, causing respondents to give the same answer (e.g. 4) on the scale for each and every following topic.

An associated problem exists when the interviewer provides a virtual context in which the respondent's risk perception should take place. Suppose you want to test whether norm-acceptance regarding the safety of food decreases after a radioactive contamination. If the interviewer has to explain the context of contamination, chances are real that the sketch of the context will be quite different from one interviewer to the other, depending on his or her level of knowledge regarding such phenomenon, his or her position towards radiation and nuclear and the likes. This variation causes undoubtedly unjustified variation in the responses.

3. Maneuvering out of the minefield: avoiding the pitfalls
In the preceding section, we identified a number of common challenges to the design of large scale CAPI-surveys on risk perception. Luckily, the observance of some rules of thumb suffices to avoid the majority of pitfalls. What is particularly important, however, is that the majority of these rules are observed *before* the interviews are conducted. Avoiding the main pitfalls is thus a question of devoting time, means and energy to the preparatory works for the survey. In actual practice, the importance of this phase in opinion research is often overlooked, since one wants to jump as fast as possible to conclusions. The bulk of methodology handbooks and experience suggest that one third of the project time is spent on the preparation of the research design, which include a desk study of relevant literature, the elaboration of the research hypotheses and variables and the actual construction of the questionnaire. While the long duration of the preparatory phase may inflict some frustration in researchers and project managers, the patience and prudence pay well off in the end.

So what is to be done in the run to the actual interviews? The following paragraphs provide an overview of activities that may help us to avoid the numerous pitfalls

listed. As in the previous section, we presuppose that is opted for CAPI. If the research budget is adequate, CAPI is superior, since it stands for good representativity through good response levels, and allows the questions to be more difficult than in case of CATI or CASI. A fresh argument is that CAPI allows the inclusion of multimedia in the interviewing, a creative direction that is discussed under heading 3.2.

3.1. The backbone: research hypotheses and summated scale variables
3.1.1. Formulating research hypotheses
Brainstorming sessions with a multidisciplinary team, combined with a desk study of relevant literature provide the answer to what is probably the most important question in the construction of the research design: what do we want to study? The answer is preferably formulated under the flag of a number of research hypotheses. We will illustrate this point with a hypothesis stemming from our familiar source, the risk perception barometer that is undertaken by SCK-CEN : 'public acceptance of a legal norm assuring the safety of foodstuffs is lower after a contamination incident'. Working with research hypotheses has many advantages over the blindfold exploratory approach: it facilitates referring to the scientific literature, it reminds us to slot in the necessary variables (items) in the questionnaire; it gives a hint to which statistical techniques are appropriate to verify or falsify the hypothesis, so that the required nature and measurement level of the variables can be taken into account when operationalizing the variables; and, not in the least, it offers a sense of direction throughout the research process.

3.1.2. Constructing summated scale variables
Elaborating hypotheses brings along another central element of decent statistical survey research: it helps to identify and construct the necessary research and context variables. The hypothesis mentioned above requires the construction of one variable, the public acceptance of a legal norm (concerning foodstuffs), and one so-called context variable, i.c. the contamination. As explained above, a variable is more than just one question. The illustrating variable, public acceptance has three dimensions (see 2.7). Each and every dimension should be taken into account when the concept of public acceptance has to be operationalized and later on incorporated in a bigger theoretical framework. Single items, fragile and little robust as they are, do seldom suffice.

The first step to construct a variable is to see whether it contains more than one dimension. For each dimension, a subvariable should be created. One-dimensional variables or subvariables are then operationalized by composing numerous items which are conceived as Likert scales. A typical question using a *Likert Scale* might pose a statement and ask the respondent whether they strongly agree - agree - undecided - disagree or strongly disagree. By way of illustration, we consider the dimension 'enforcement' of the multidimensional variable 'legal acceptance' (which is, in fact, the sum of three (sub) variables). Many researchers would be comfortable with measuring perception of enforcement through one single item, like: 'The government punishes violators of the norms as it should be'. Warned against the shortcomings of single items, we present a selection of items which do

all measure – more or less – the same concept. A second item could be: 'penalties are not severe enough in case of a violation of the legal norm'. All items are answered to on a scale from 1 to 5. Summation of the individual scores yields a so-called summated scale variable (which is subsequently standardized). We have now a variable that is of the (pseudo)interval level and is resistant to different interpretations of one or more of the single consisting items by respondents. Of course, there is a need for formal tests to confirm whether the particular items for one variable measure the same construct. Cronbach's Alpha – a calculation of intracorrelation between the items – and a simple factor analysis are ideal tests. If time and money are plentiful, the test could be applied to a test sample before the 'real' survey is conducted. Of course, not all variables must be constructed. Most of the independent variables, like sex or age, do not need any operationalization at all; they are translated directly into questions. This is, however, not so simple if one wishes, e.g., to study the impact of emotions on risk perception. This could, e.g., require mapping of emotional states of the respondents. To do so, advice must be sought from psychologists studying the matter. Multidisciplinarity is thus more than a buzzword.

3.2. The interface: the questionnaire

Once the hypotheses and the variables are elaborated, it is time to construct the questionnaire. Thoughtful wording, shunning suggestion, alternation between positive and negative statements, avoidance of jargon and rotation of themes, items and nominal answering categories are some classic points of attention.

An original approach that is brought forward in the new SCK barometer involves the use of video clips during the CAPI. One of the stated research hypotheses that is used as an example throughout this contribution deals with a contamination, a context that must be communicated to the respondent. We expect that this message will influence the perceived acceptance of legal norms. We decided not to let the interviewer explain the contamination incident because this might provoke too much variation in the way the 'news' was brought: it concerns a message of at least two minutes. Furthermore, we wanted to approximate the conditions of a real accident as close as possible, since we suppose that risk perception is dependent of the way in which the information regarding the risk is communicated. The result was a two minutes' newsflash brought in a media format compatible with the interviewers' notebook's operating system, imitating the format of the most popular television news. We will report in another article on this highly experimental method to convey dedicated context information in a video format in the course of an interview.

3.3. The weakest link: the interviewers and the interviewed

In section 2, we listed some threats to the reliability of the answers. Dealing with these threats is a form of quality assurance. Of utmost importance is the briefing of the interviewers: they should be told in the clearest terms how the interview should be conducted and how long it should last; furthermore, during this briefing, it could not be stressed enough that quality assurance will be applied thoroughly. Especially the 'rushing through' phenomenon should be countered in advance.

Special attention should also go out to the effects of social desirability: during the interviews, interviewers should take a stance as neutral as possible. Besides, they should repeatedly emphasize in the course of the interview that the anonymity of the participants is guaranteed in every stage of the research project.

While bias during the interviews is unavoidable, a thorough briefing of the interviewers can help. It also helps to select an opinion research bureau that has a good reputation, even though this will be traded off in the price tag.

Another bothersome element is respondent reliability. Long or difficult interviews cause fatigue amongst respondents, who will answer anything as long as it helps to get the interview over and done with. Questions should be interesting and not too difficult, nor too easy, to begin with. But of principal importance is the duration of the interview: 40 minutes is a maximum for reliable face-to-face interviews. In case CATI is preferred, the timeframe becomes even more stringent.

References

1.   Slovic P. Perception of Risk, Science, Vol. 236, April 1987
2.   Callon M. Lascoumes P. Barthe Y. Agir dans un monde incertain. Essai sur la démocratie technique, Editions du Seuil, Paris, 2001
3.   Slovic P. Perception of Risk, Science, Vol. 236, April 1987
4.   Martin B. Strip the Experts, Freedom Press, London, 1991
5.   Kasperson R.E. et.al., The Social Amplification of Risk: a Conceptual Framework, Risk Analysis, 8(2), June 1988
6.   Beck U. Risk Society, Towards a New Modernity, Translated from German to English by Mark Ritter, Sage Publications, London, 1992 [originally publ. 1986]
7.   http://europa.eu.int/comm/public_opinion/index_en.htm
8.   Van Houtte J. Aanvaarding van de rechtsnorm, Antwerpen/Utrecht, De Nederlandsche Boekhandel, 1973
9.   SCK·CEN. Risk Perception Barometer, Mol, 2002. Report available at http://www.sckcen.be/pisa. A new barometer is under construction.
10.   A widely used alternative to CAPI is CATI (Computer Assisted Telephone Interviewing). It costs less than CAPI, but the nonresponse is much higher and representativity is jeopardized since mobile phone owners without a fixed line are (at least in Belgium) not listed in telephone directories. Since this is often the case with young people, CATI tends to over represent the older generations, and people who are at home quite often, like immobile, ill or unemployed people. Moreover, questions should be very simple, as the lack of direct communication prevents further explanations and the threat of the breaking-up of the interview is acute.
11.   http://faculty.msb.edu/brewert/documents/USPublicOpinionupdatefor2005.doc
12.   Unfortunately, violation of assumptions does happen quite often in research in the humanities. Poor questionnaires are one of the main reasons for this abuse, since they yield variables that are of too low a measurement level.