Nonreponses by design? Implementation of Long Surveys in Mixed Mode Environment*

Alexandre Pollien[1], Patricia Milbert[1], Michael Ochnser[1], Michèle Ernst Stähli[1] and Dominique Joye[1]

[1]FORS Lausanne
[2]ISS-SSP, University of Lausanne and NCCR Lives

August 10, 2017

1 Introduction

The discussion here will be concerned before all by the idea of "matrix design", looking for reducing the length of a survey and to adapt it to a different setting. This could be seen as a little bit provocative in a Nonresponse workshop context: most often the discussion is about reducing the number of nonresponse while this one is about introducing nonresponse! In fact we will try to show that there is no contradiction between these approaches. Some arguments in this line:

- About burden: we know that the response rate is, among others, linked to the length of a survey, but also, in the same line, to the interest in the survey for the respondent, however it is measured.
- About design: is it possible, and under which conditions, to shorten a survey while keeping as much as possible of the information? This could be done by selecting the most informative question but also by imputing, as far as possible, missing information.
- About efficiency: in a perspective to the total survey error paradigm, some designs could improve the "cost/information" ratio of the survey, even at the price of a growing total number of non-response.

These elements have to be discussed not only in the context of a single survey but also in a comparative frame: how far the principles governing this search for the most useful information are the same in different contexts?

And, to include one more aspect, how to include, and to train, the users of a dataset in the problems related to non-response. The argument here is twofold: if it is true that it is a challenge to impute information if non-responses are produced by design, this is also perhaps easier than to deal with "non-response by error", resulting of difficulties in the survey process. We can even argue

*Paper prepared for the Nonresponse Workshop, Utrecht, August 2017

that "non-response by design" tends to decrease "non-response by error", at least to a certain extent, which is linked to the redundancy of the information contained in an questionnaire, and which makes possible to impute the missing information. It is therefore a question of finding this equilibrium point which minimizes the "non-response by error" thanks to the "non-response by design".

The next edition of EVS[1] will allow some countries to test a matrix design in order to see, in a comparative perspective, the consequences of changing mode in one of the oldest repeated transversal survey. Of course, this could give useful information on adaptation of log term established surveys to new constraints without necessarily losing comparability in the time.

2 A Short History of Matrix Design

Matrix design, or the idea to ask only a subset of questions to selected groups of respondents, is not a new idea. It was called "Multiple Matrix Design" by Shoemaker in the beginning of the seventies already. A development referring to the multiple imputation by Ragunathan and Grizzle was published in 1995 and discuss the choice of the variables in order to build the blocks, suggesting that a random selection is better. There was also implementation of such technique for psychometric scale or research in education (Kaplan & Su 2016) as well as integration in surveys like consumer survey (Gonzalez & Eltinge 2007 and 2008). This strategy is also now discussed in the survey literature like Survey Methodology (Merkouris 2015) or JOS (Chipperfield and Steel 2009).

It is also used in one of the major project of the social sciences, the European Social Survey (ESS), even if it is probably not totally obvious for most users: by the way, the strategy proposed by Saris for the last questions of the ESS, and used in a MTMM context for the quality test of questions, basis of SQP, is also a matrix design. And many other examples exist, even if they were not always presented explicitly nor their methodological consequences discussed in details.[2]

These small set of references show a little contrasted result: even if presented in the literature since nearly fifty years, even if used in a number of context without so many discussion, such a strategy was not adopted, at least to our knowledge, to important surveys, even more in a comparative context. Nevertheless the preoccupation to reduce burden is always a main worry.[3]

3 Burden and Interest or Some Rules for Selecting Questions to Integrate in Different Blocks

One of the argument mentioned refers to the length of the survey in order to reduce the burden. If a lot of studies have established a relation between length, burden and data quality, there are still a lot of variations according the studies, the discipline and the context in order to define the best standards. Two elements are important to recall in this line.

[1] European Value Study, for details see www.europeanvaluesstudy.eu

[2] For example, in the Swiss case, the "Tous égaux survey" (Levy et al., 1997), fielded in 1991, was already using such a schema.

[3] For example see Plewes (2016)

- There is a common agreement linking the "acceptable length" to mode. Traditionally, a face to face survey is designed with one hour interview in mind while telephone, mail or web surveys where designed with 25 minutes in mind. Converting this into items, a face-to-face survey could correspond to a little bit more than 240 "ticks" while mail survey is typically designed for 100 "ticks" in mind.[4]

- The length is always related to the interest of the survey for the respondents. This interest is not easy to measure, and can vary a lot according the settings and the social characteristics of the respondent, but we believe that a survey is before all a social interaction.[5]

  In considering how to cut the questionnaire, we noticed that the "meaning" for the researcher could lead to the creation of questionnaire that do not make sense to the respondent (going from one topic to another, from a particular question to a general question, absurd precision, and so on). Thus, the overall approach of matrix design consists also in attempting to combine the "meaning" of research (the survey must make sense in term of volume and precision of information collected, of relation between scientific topics) with meaning for the respondent (the questionnaire must make sense as the interview must be not too long, it must be a pleasant moment on interesting topics).

These elements give us some guidelines for designing surveys or their adaptation to other modes, like the EVS in our case. But an other point has to be underlined: when selecting items in order to build a survey in a matrix design frame, each edition has to keep a guiding threat to help the respondent the understand what it is expected from his side.

4 An Evolution of International Comparative Face-to-Face Surveys?

There is a tradition of face to face in most international comparative surveys. ESS, EVS, SHARE and most "Barometers" surveys are done using this mode. The only notable exception is probably the ISSP, but this is probably due to its origin as an addition to national surveys. Face to face was seen as the simplest mode in a comparative perspective because it does not ask for special capabilities like reading and writing and could be implemented in all societies. However, these last years, a lot of things have changed, with surveys more difficult, or more expensive, to realize while keeping the same level of quality. In fact, one can observe the continuation of the process describes by Tönnies 130 years ago about the move from the *Gemeinschaft* to the *Geselschaft*. With the growth of commuters, individualization of lifestyles, we are less and less often in the situation of an interviewer going into a neighbourhood made of inter-knowledge and long-standing contacts. Interviewers face to closed doors, distrust, fear or blindness of the neighbourhood. This has often be mentioned in the survey literature by reference to, among others, the following points.

[4] By the way, an ISSP survey is 60 "ticks" meaning a survey of a little more than 100 "ticks" including the "background variables".

[5] And the literature on incentives is clear on this, insisting on the idea of social exchange rather economic value.

- Technical reasons: as access limited by codes in many cities
- More professional or external activities for the different members of an household, meaning more difficulty to contact the peoples.
- Change of acceptation of interviewers in a context of changing survey climate.

The result of this is an increase in cost of the face to face, just in order to maintain the same quality. This ask of course questions for the sustainability of this mode in the long run and aboiut the possibility to change the model. In fact this is not at all a revolution: since the beginning of the survey industry, there was discussion about the mode of choice. For example, even if it is probably no more true, telephone was seen as an alternative to face-to-face in some countries for quality surveys and when ESS was designed this aspect was also discussed.[6] This is also an illustration of the principle saying that, in a changing environment, the strategy not to change method is not a solution to maintain comparability.[7]

This change in the environment imply perhaps an adaptation of the survey instruments, even more in a comparative perspective. This is one more argument to design a real size experiment.

5 The design of an Experiment

For the 2017 edition, the EVS was designed as requiring a 1'200 persons (effective size) in face-to face. However, the rules allow countries to chose an alternative design with at least 600 interviews in face-to-face and at least 600 by web or web/mail. Part of the reasoning was to test an adaptation to new modes while keeping comparability with the previous editions.

Finland, Germany, Iceland, Netherlands and Switzerland were interested to participate to this experiment. However, the length of the questionnaire, more than 250 "ticks" was clearly excessively long for a web survey. Some of these countries are considering to implement a similar matrix design complemented by a "mini-panel" component, meaning to come back to the same respondents some weeks or month later for a second "session". This will introduce a very interesting design allowing comparison of effects in different contexts.

In the next paragraph, we will describe the Swiss design, knowing that probably the other countries will use a similar one.

5.1 The Main Design in Switzerland

The main design decision was to take the option of a matrix design dividing the questionnaire in 4 blocks and a "core". Each respondent has to answer to two blocks in addition to a "core" part. This ensure some overlaps either between

[6] This was particularly true in Switzerland where survey institute as well as the Statistical office have invested a lot in this model of survey.

[7] This is also true when thinking about the content of some question referring to contextual element like inflation for example, at different level in the seventies or nowadays in most western countries. In the same line, the work in translation about keeping the meaning rather the form could be interesting to keep in mind when trying to maintain equivalent measurement tools.

the core and the blocks, either between blocks in order to be in the best possible situation to impute the content.

Some respondents will be asked to answer to the full questionnaire. In the design we have:

- CAWI, matrix part: 4800 addresses, 6 blocks and 50% response rate, means 400 respondents by block (240 for the second session if 60% of the respondents accept to continue).
- CAWI in one step, 2000 addresses and 30% response rate means 600 respondents.
- CAPI, 1400 addresses, 44% response rate, means 616 respondents.

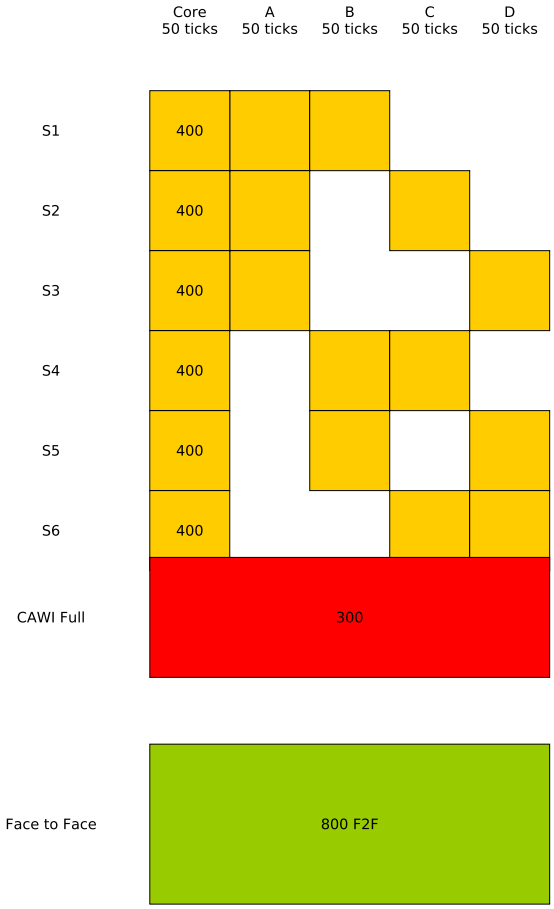

Figure 1: The Swiss Design

As mentioned, the matrix design is completed by a "mini-panel" device. Respondents of this second questionnaire will be asked to answer to the parts that were not filled in the first run, some weeks or months later. Even if we can expect a significant attrition, perhaps as high as one third, this allow to

play with different way to "rebuild" the missing cells, even if the second session induce a change of the context of interviewing difficult to control, this provides us with different measures to "rebuild" the missing cells. But that means also a careful examination of the way to build the blocks and to take into account the content of each.

5.2 Building Blocks

The construction of blocks is perhaps one of the more challenging operation. Basically there are 2 options, presented here according extreme positions:

- The "Statistical option" where the choice of questions to be included is random or subject to particular algorithms.[8] The advantage is not to have any systematic effect, particularly if the number of blocks is not limited. The drawback is twofold:
  - The difficulty to manage a questionnaire individualized by respondent;
  - The possibility to have a very disconcerting questionnaire for the respondent
- The "Sociological option" where the choice of question to be included is carefully made by the team. Two constraints and an ambition at least in this situation, showing potentially a tension between the logic of the researcher trying to optimize the his conception of the measurement and the one of the respondent:
  - A choice of items that make sense for the respondents
  - The need to keep nevertheless a diversity of subjects and not to change too much the context effect of the reference survey.
  - To minimize information loss of missing questions as far as possible or, in other words, to distribute the redundancy of questions over the different splits. EVS 2008 and the research made on these data can be used to improve the heuristic power of each split.

Two more elements of design have to be considered in such a situation:

- The content and size of the core. Once again two elements:
  - The content: one idea could be to use mainly the socio-demographic part that we need for every respondent. However, another constraint is to have variables highly correlated with the variables present in the blocks. This is not necessarily the characteristics of the only

[8] For example Adigüzel & Wedel (2008) are writing "Instead of the heuristic randomization methods to design split questionnaires that are currently used in applied and academic research, we develop a methodology to design the split questionnaire to minimize information loss using estimates from a first wave or pilot study. Because the number of possible questionnaire designs is exponential in the number of questions, we apply the Modified Federov algorithm, using Kullback Leibler Distance as a design criterion, to find the optimal splits. We use Markov chain Monte Carlo procedures to impute missing values that result from the design.

At another level of abstraction, we will also discuss later the importance of the structure of correlation in previous studies when searching the way to build blocks inside a questionnaire.

socio-demographic variables. This is why we have decided to exclude some socio-demographic variable of the core, for example concerning the partner, and to include "opinion" variables supposed to be more correlated with values in the core.

  - The greater the core, the easier to include a large set of variables, at the same time the gain in length will be smaller if the core is bigger. We chose to stay with a core more or less of the same size as a block (2 blocks taken within 4 and 1 core of nearly the same size, means 60% of the original size).

- The number of blocks: in statistical terms, the higher the better. At the same time, we have to keep in mind that we have to be coherent in the recalls as well as for the "mini-panel" part later. This is even more important and complicated in a country where the survey is fielded in three different languages: French, German and Italian. Furthermore, we have the constraint to keep some logic in the way that the survey is experienced by the respondents. We have therefore chosen to stay with 4 blocks meaning, in combination 2x2, 6 situations.

More precisely, we have followed the following rules, the details of the reasoning can be found in additional documents.

- Each questionnaire has to be meaningful for the respondents.

  Creating meaningful questionnaires for the respondents not only lowers their burden but they will also be more likely to give reasoned answers, are less likely to interrupt their participation or to drop-out, and will be more willing to answer also the follow-up questionnaire, answering therefore the whole questionnaire.

- The questions will be asked strictly in the same order as in the source questionnaire.

  Questions or group of questions can be dropped for certain respondent groups, but cannot be moved. This means for example that questions 22 will always be asked after question 10, never the inverse, but that for some respondent groups there will be no question in between because these question in between have been defined as belonging to another block.

  Keeping the order of the source questionnaire optimizes comparability across modes and respondent groups and with data of previous waves. From the point of view of the whole design, including the follow up, therefore, the order of the questions is not respected, which makes the context of the second session less strictly comparable to the first session.

- If variables are subject to change meaning when isolated from other adjacent variables, they are placed in the same block.

  E.g. it does not mean the same thing to say “I feel close to Europe” if the previous variable is “feel close to my city” rather than "feel close to my country".

  Changing as little as possible the context of the questions is important to insure equivalence of meaning for the respondents. As changes cannot be totally avoided when splitting a questionnaire, special attention is given to potential context effects.

- If variables are known to be often analysed together, they are placed in the same block. It is here that the meaning of the questionnaire for the respondent and the meaning of the questions for the sociologist try to join. Each split must constitute a survey in its own right. The whole information collected must be coherent and meaningful from the respondent's point of view as well as from the researcher's point of view.

  However, keeping variables that are often analyzed jointly in the same block is not the leading principle, because new ways of analyzing the data should not be discouraged. We allow for the splitting of batteries in two different blocks (but not more) under the condition that the split batteries make sense for the respondents and do not alter the meaning of the variables. 5/6 of the respondents will then answer half of the battery, and 1/6 will answer the whole battery (when split into 3 blocks, nobody would answer the whole battery).

- The block of background variables includes the principal socio-demographic information but also some very broadly used substantial variables. Less used socio-demographic variables are placed in one of the thematic blocks.

  Adding substantial variables (such as life satisfaction) to the background variables block will improve the analytical power for most researchers. It also allows for better analyzing drop-outs (and attrition for the follow-up).

An appendix is joined to this paper showing part of the work done in terms of use of variables by researchers and correlation between items while preparing the structure of these blocks and the final design.

6 Conclusion and Open Questions

Based on the literature, the design described here seems possible to implement. Concretely, once the decision of a design of this type is taken, there are not so many room for interpretation: number and length of blocks is limited and we need as many respondents as possible in each of the condition.

One important point nevertheless could be the acceptability for the users.[9] One interest if to use multiple imputation of course but not all user of an international survey has an extensive knowledge of such a technique.[10] This means that the scientific community has to provide guidelines for the users, and the potential reviewers of scientific paper, and the way to use this kind of data. This is one more challenge while the documentation of more standard surveys seems yet to be not exploited at his full potential...

Another point is the cost/benefit ratio of such a design. In the case, combining 800 face-to-face interviews, a "mini-panel" structure and 2'000 cases in the Internet part, without counting the ones to which the full survey will be proposed does not mean a great saving by comparison to a 1'200 respondents face-to-face survey. But the main interest is perhaps to test in real condition the way to push long face to face surveys in other modes and to envisage what

[9] This was already addressed in the discussion of "First Workshop of the Consumer Expenditure Surveys Redesign Paper, Split Questionnaire"

[10] This is also one more reason to use a "mini-panel" structure and overlapping blocks.

could be, perhaps, a future for comparative analysis and a way of doing large comparative survey in CAWI mode.

A last point could be interesting to discuss more deeply. From one side, the TSE paradigm is a way to think to the best possible design for a survey. However, this optimum depends of course of the point of view of the observer. One interest of this kind of design is perhaps to reintroduce other actors than survey designers, like users of the survey, probably mainly social scientist in the case, and also respondents, which are often the key element when thinking about survey quality.

7 References

Adigüzel F. & Wedel M. (2008) "Split Questionnaire Design for Massive Surveys", *Journal of Marketing Research*, October 2008, Vol. 45, No. 5, pp. 608-617.

Chipperfield J. & Steel D. (2009) "Design and Estimation for Split Questionnaire Surveys", *Journal of Official Statistics*, Vol. 25, No 2, pp. 227-244.

Coelho Rodrigues N.M. & Iwakami Beltrão K. (2005) "A Split Questionnaire Survey Design applied to Brazilian Census", retrieved from `iussp2005.princeton.edu/papers/51631` accessed the 10.8.2017.

Gonzalez J.M. & Eltinge J.L. (2007) "Multiple Matrix Sampling: A Review" *Proceeding of the Section on Survey Research Methods, American Statistical Association*, pp. 3069-3075.

Gonzalez J.M. & Eltinge J.L. (2008) "Adaptive Matrix Sampling for the Consumer Quarterly Interview Survey", Washington, retrieved from `https://www.bls.gov/osmr/pdf/st080060.pdf` accessed the 9.8.2017

Kaplan D. & Sue D. (2016) "On Matrix Sampling and Imputation of Context Questionnaires With Implications for the Generation of Plausible Values in Large-Scale Assesments" *Journal of Educational and Behavioral Statistics*, Vol41, No 1, pp. 57-80.

Levy R., Joye D., Kaufmann V. & Guye O. (1997) *Tous égaux?*, Seismo, Zurich.

Merkouris T. (2015) "An efficient estimation method for matrix survey sampling" *Survey Methodology* 41-1, pp. 237-262.

Plewes T. (2016) *Reducing Response Burden in the American Community Survey: Proceeding of a Workshop*, National Academies Press, `http://www.nap.edu/23639` accessed the 9.8.2017.

Ragunathan T.E. & Grizzle J.E. (1995) "A Split Questionnaire Survey Design" *Journal of the American Statistical Association*, 90, pp. 54-63.

Rässler, Susanne, Koller, Florian and Mäenpää, Christine, (2002), *A split questionnaire survey design applied to German media and consumer surveys*, No 42b/2002, Discussion Papers, Friedrich-Alexander University Erlangen-Nuremberg, Chair of Statistics and Econometrics, `http://EconPapers.repec.org/RePEc:zbw:faucse:42b2002` accessed the 10.8.2017.

Shoemaker D.M. (1973) *Principles and Procedures of Multiple Matrix Sampling*, Ballinger, Cambidge MA.