**SPLIT QUESTIONNAIRE DESIGNS FOR LONG SURVEYS**

by Simon Moss

|  |
| --- |
| **Introduction** |

**Drawbacks of long surveys**

In many research projects, researchers administer surveys or questionnaires. Sometimes, these surveys are very long and demand, for example, more than 45 minutes or so to complete. Consequently, several problems are likely to unfold. The following table outlines these problems.

|  |  |
| --- | --- |
| Problem | Consequence of problem |
| Many participants will withdraw from the survey prematurely | Answers to the later questions do not include participants who are not as persistent—and, therefore, these data are biased  |
| Participants may skim the questions rather than read the questions properly | The answers tend to be unreliable or invalid. For example, participants are more inclined to specify the same answer for every question |
| The costs may be steeper | For example, to attract enough participants, researchers might need to pay each participant more money  |

**Benefits of long surveys**

 These problems imply that researchers should administer shorter surveys. However, longer surveys actually generate many benefits as well. The following table outlines these benefits.

|  |  |
| --- | --- |
| Benefit | Consequence of benefit |
| The same participants answer a greater diversity of questions | Researchers can thus explore correlations and relationships between questions that have not been asked in the same questionnaire before  |
| Logistically, one long survey can be more efficient than several shorter surveys | For example, you can recruit fewer participants |

**Solution**

 To enjoy the benefits of longer surveys, while circumventing the problems, some researchers utilize an approach called a split-questionnaire design. This design is feasible whenever the survey is administered online or over the telephone. This document outlines this design. In essence

* All participant receive a specific subset of questions, called the core component
* Participants also receive random subsets of other questions, called subcomponents
* Appropriate statistical analyses are conducted to replace the missing data with suitable estimates

|  |
| --- |
| **1 Divide questions into subsets** |

First, you need to divide the questions in your survey into subsets, as the following figure illustrates. As a rough guide, often

* each subset of questions may correspond to one measure, one scale, or one topic, for example
* a survey that comprises 100 questions might, for example, be divided into 5 to 10 subsets

Decide which of these subsets should be core and presented to everyone. Nobody has developed definitive criteria to decide which questions should be core. Nevertheless

* often, the key outcome measures—the main topic or emphasis—is core, because you want to be able to model these measures as accurately as possible
* measures that are not strongly associated with other measures are often core as well

|  |  |
| --- | --- |
| Core components | Subcomponents |
| Main measure | Unrelated to other measures | Predictor 1 | Predictor 2 | Predictor 3 | Control variable |
| * Question 1
* Question 2
* Question 3
 | * Question 4
* Question 5
* Question 6
 | * Qn 7
* Qn 8
* Qn 9
 | * Qn 10
* Qn 11
* Qn 12
 | * Qn 13
* Qn 14
* Qn 15
 | * Qn 16
* Qn 17
* Qn 18
* Qn 19
 |

|  |
| --- |
| **2 Assign each component to a distinct block or page**  |

 Second, if the survey is administered online, each core component or subcomponent should appear in a separate page or block. For example, if using Qualtrics, you could click “Add block”—an option that appears towards the top and middle in the figure below—and then construct one subcomponent on this block. For more information about Qualtrics, see the document “Administering surveys: Introduction to Qualtrics”.



|  |
| --- |
| **3 Randomly allocate a subset of the subcomponents to each participant**  |

 Third, you need to randomly allocate a particular number of these subsets to each participant. For example, as the following table shows

* Participant 1 might receive the two core components and the first two subcomponents
* Participant 2 might receie the two core component and the third and fourth subcomponent

|  |  |
| --- | --- |
| Core components | Subcomponents |
| Main measure | Unrelated to other measures | Predictor 1 | Predictor 2 | Predictor 3 | Control variable |
| * Question 1
* Question 2
* Question 3
 | * Question 4
* Question 5
* Question 6
 | * Qn 7
* Qn 8
* Qn 9
 | * Qn 10
* Qn 11
* Qn 12
 | * Qn 13
* Qn 14
* Qn 15
 | * Qn 16
* Qn 17
* Qn 18
* Qn 19
 |
|  | Image result for silhouette personPerson 1 |  |  |  |  |

|  |  |
| --- | --- |
| Core components | Subcomponents |
| Main measure | Unrelated to other measures | Predictor 1 | Predictor 2 | Predictor 3 | Control variable |
| * Question 1
* Question 2
* Question 3
 | * Question 4
* Question 5
* Question 6
 | * Qn 7
* Qn 8
* Qn 9
 | * Qn 10
* Qn 11
* Qn 12
 | * Qn 13
* Qn 14
* Qn 15
 | * Qn 16
* Qn 17
* Qn 18
* Qn 19
 |
|  | Image result for silhouette personPerson 2 |

 To achieve this goal in Qualtrics

* click *Survey flow*—a button that appears towards the top of Qualtrics; the following screen will appear
* press *Add below* alongside the final core component of this survey
* a series of buttons will appear; press the button *Randomizer*
* the screen will now resemble the following screenshot



* Press *Move* alongside the first subcomponent of questions
* Drag this block to the phrase *Add a New Element Here*
* Repeat with the all the other subcomponents of questions to generate the following screen
* Enter a numeral in the box after *Randomly present…*This numeral is the number of subcomponents or optional blocks you want participants to receive
* Finally press *Save Flow*



|  |
| --- |
| **Impute missing data** |

 After you administer this survey, you will receive a data file that includes lots of missing data. Researchers can apply a range of strategies to address these missing data. One effective technique is called expectation maximization (for discussions, see Dempster, Laird, & Rubin, 1977; Schafer & Olsen, 1998; Scheuren, 2005). For example, to execute this technique with SPSS

* Choose *Missing Value Analysis* from the *Analyse* menu.
* Transfer all the relevant numerical variables—the columns with numbers—into the box labelled Quantitative Variables.
* Transfer all relevant categorical variables—the columns with labels or categories—into the box labelled Categorical Variables
* Select the EM option
* Press the EM button, and select Save completed data.
* Choose *Write a new data file*. Press File and type a filename.
* Open this new file. This file should include the data together with some of the missing data completed.

This technique is especially effective when the data are missing at random—that is, none of the characteristics of participants should affect which data are missing. Because the subcomponents are allocated randomly, this assumption is likely to be fulfilled. For more information, visit [www.sicotests.com/psyarticle.asp?id=267](http://www.sicotests.com/psyarticle.asp?id=267).

|  |
| --- |
| **Evidence of utility** |

 Many studies have verified that, if appropriate techniques are used to address the missing data, split questionnaire designs are valid. The correlations between questions are similar regardless of whether you utilize the split questionnaire designs compared to instances in which participants complete all the questions (for evidence, see Huff at al., 2016; Raghunathan and Grizzle, 1995).

 Admittedly, as Adigüzel and Wedel (2008) discuss, the validity of this design partly depends on the decisions that researchers choose. For example, researchers can

* utilize a variety of techniques to overcome missing data, such as Markov chain Monte Carlo procedures
* apply various algorithms to identify better ways to divide the questions into components, such as the modified Federov algorithm
* randomly allocate questions within each block—rather than entire blocks—to each participant, called a within-block design instead of a between-block design

|  |
| --- |
| **References and future reading**  |

Adigüzel, F., & Wedel, M. (2008). Split questionnaire design for massive surveys. Journal of Marketing Research, 45(5), 608–617.

Dempster, A., Laird, N., & Rubin., D. (1977). Likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 39,1-38

Huff, S. C., Anderson, S. R., & Tambling, R. B. (2016). Testing the clinical implications of planned missing data designs. Journal of Marital and Family Therapy, 42(2), 313–325.

Vriens, M., Wedel, M., & Sandor, Z. (2001). Split-questionnaire designs: A new tool in survey design and panel management, Marketing Research, 13, 14-19.

Raghunathan, T., & Grizzle, J. (1995). A split questionnaire survey design. Journal of the American Statistical Association, 90, 54-63.

Schafer, J. L. & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. Multivariate Behavioral Research, 33, 545-571.

Scheuren, F. (2005). Multiple imputation: How it began and continues. The American Statistician, 59, 315-319.