

4.1 GENERATION OF VIGNETTE TEXTS & RANDOM VIGNETTE SAMPLES

[Accompanying material for: Katrin Auspurg & Thomas Hinz (2015): Factorial Survey Experiments. Sage Series: Quantitative Applications in the Social Sciences No. 175. Thousand Oaks, CA: Sage]

(§4_questionnaires.pdf and the material in the folder “Questionnaires” show how to insert the resulting vignette texts into questionnaires using mail-merge functions.)

Material that is provided along with this introduction (in §4_material.zip):

- **Folder “Stata”:** several Stata do-files to generate random vignette samples, to produce vignette texts, to prepare random orders of vignettes for each respondent, and to prepare data and vignettes for their use in questionnaires and data analyses:
 - **1_master.do:** overview of all do-files; can be used to run all files in the correct order.
 - **2_random.do:** generation of random samples of vignettes (with the exclusion of illogical or implausible cases), random allocation of vignettes to decks, and checks of sample features (level balance and uncorrelatedness of dimensions).
 - **3_labeling.do:** renaming, recoding, and labeling of vignette dimensions and generation of labels for all levels to prepare their use in data analyses and to obtain a better overview of the sample.
 - **4_texts.do:** translation of vignette dimensions and levels into single phrases to be shown in text vignettes; reshaping data from long to wide format to prepare the use of vignettes in questionnaires.
 - **5_sorting.do:** duplicating decks so that there are questionnaire versions for each respondent; arranging idiosyncratic, random vignette orders for each respondent and preparing vignettes for their use in questionnaires (e.g. reshaping the vignette data from long to wide format).
 - **Fractionalized120.dta:** a fractionalized, *D*-efficient sample of 120 vignettes as it is produced by the SAS job in §3_material.
- **Folder “PDF”:** One might use other software packages to create the vignettes as well. To make the do-files and single steps readable for users who do not work with Stata, we provide all files additionally as pdfs.

4.1.1 Starting Point: *D*-Efficient Sample

One might use any statistical software package to translate the numerical codes of the dimension levels into the string phrases that build up text vignettes. When using *D*-efficient sampling techniques (as we recommend) the resulting data matrix (like the **fractionalized120.dta** data we provide) is a matrix that contains all dimensions as different variables (x_1, x_2, \dots, x_p) that have the different levels numerically coded (1, 2, etc.). There are as many lines as there are vignettes in the sample (i.e. the data are in the “long format”; see Chapter 5 for data formats) and there are an additional two variables that indicate the different decks (“blocks”) and the vignette position within the decks

(variable “run” when working with the SAS macros, e.g. the level “1” of the variable “block” in combination with the level “1” for the variable “run” indicates that this is the first vignette within the first deck). Figure 4-1 shows the data matrix of our fairness of earnings example.

Figure 4-1 Data Matrix of a *D*-Efficient Sample of 120 Vignettes, Blocked to 12 Decks

block	run	x1	x2	x3	x4	x5
1	1	1	1	2	1	1
1	2	1	2	2	2	4
1	3	1	6	3	4	9
1	4	1	8	2	4	7
1	5	1	8	3	3	8
1	6	2	1	3	5	2
1	7	2	3	3	5	10
1	8	2	4	1	2	6
1	9	2	5	1	3	3
1	10	2	7	2	6	5
2	1	1	1	1	5	7
2	2	1	2	2	4	8
2	3	1	3	2	2	1
2	4	1	5	3	3	2

4.1.1 Random Sampling

An alternative way (which is, however, less efficient in a statistical sense and also at risk of confounding meaningful parameters) is to use **random samples**. In the provided zip folder there is also one do-file showing how to create a random sample for our fairness of earnings sample with Stata and how to block this sample randomly to the different decks (**2_random.do**). The result is a similar data matrix to the one shown in Figure 4-1.

4.1.2 Labeling and Generation of Text Phrases

Next, the **variables should be labeled**. In addition, one should ensure that there is a random order of vignettes (see the sample file **3_labeling.do**). After that, the **numeric codes have to be translated into text phrases**. This is done by generating additional string variables and using several if-commands. In what follows, we show an example for the Stata programming code (the whole code can be found in **4_texts.do**).

The dimensions or vignette variables in the following example are named *age*, *sex*, *educ*, and *child*; age is a metric variable, while sex is a categorical variable with the level “1” indicating the male vignette persons and “2” the female ones.

In the first command lines, with the commands `generate` and `replace` a new string variable (with 30 characters) is generated, and filled with the correct phrases to represent the different combinations of levels for the first two dimensions.

```
generate str30 phrase1 = "A 30-year-old man"    if age == 30 & sex == 1
replace phrase1 = "A 30-year-old woman"        if age == 30 & sex == 2
replace phrase1 = "A 40-year-old man"          if age == 45 & sex == 1
```

.
. .
.

Similarly, the text phrases for the other dimensions are built up, e.g.

```
generate str30 phrase2 = " with no vocational training"    if educ == 1
replace phrase2 = " with vocational training"              if educ == 2

.
.
.

generate str30 phrase3 = " has no children."              if child < 3
replace phrase3 = " has one child."                        if child == 3
```

Next, the text phrases are combined to the single vignette texts, and the variables with the single phrases are deleted (dropped; for doing this. One could use the following Stata code to produce the first sentence (stored to the variable `vigA`):

```
gen str300 vigA = phrase1 + phrase2 + phrase3
drop phrase*
```

The asterisk after `phrase` (`phrase*`) is simply used as a wildcard to indicate that all phrases, independently of the number they end with, should be deleted. What results is a data set that still contains the numeric coded vignette dimensions, but also string variables for the different vignette sentences. After some additional data management these texts can be inserted into questionnaires. For further instructions, see **4_texts.do**.

4.1.3 Arranging Random Vignette Orders per Respondent

To prevent order effects it is recommended to arrange idiosyncratic, random orders of vignettes for all respondents. To do so, one has to prepare as many different questionnaire versions as there might be respondents, and then shuffle the vignettes within all questionnaire versions in a random order. An example of how to manage this with Stata is provided in **5_sorting.do**.

4.1.4 Final Remarks: Keep Numeric Codes and Identifiers!

In all cases it is **important to keep not only the text phrases but also the variables containing the numeric codes for the vignette dimensions**, as these will be needed for data analyses (see Chapter 5 and the sample analyses files provided with the §5_material). Moreover, **one needs clear identifiers for the decks, vignettes, or questionnaire versions** allocated to the different respondents. Instructions on how to arrange this are given in the Stata do-files.

4.2 GENERATION OF PAPI QUESTIONNAIRES USING MAIL-MERGE FUNCTIONS

(§4_vignetttexts.pdf and the material in the folder “Stata” show how to produce the vignette texts that are needed in this step.)

Material that is provided along with this introduction (§4_material.zip):

- **Folder “Questionnaires”**
 - **Questionnaire_template.doc:** A Word document that can be used as a template to produce one’s own PAPI questionnaire or practice the mail-merge functions.
 - **Vignettetexts.xls:** An Excel sheet that contains in the different columns the vignette sentences for 12 decks (stored in separate lines) containing 10 vignettes each for the fairness of earnings example. The different column names (in letters; A, B, and C) indicate the different sentences, while the numbers indicate the different vignette positions within the decks. For instance, the first line of column vigA1 contains the first sentence of vignette1 in deck1; vigB1 (vigC1) would be the second (third) sentence of this vignette; while vigA2 contains the first sentence of the second vignette. The second (nth) line shows this information for the second (nth) deck. How to produce such an Excel sheet is shown with the materials in the Stata folder and in particular within the **Stata do-file 4_texts.do**.

4.2.1 General Remarks

As is explained in detail in the textbook (see Chapter 4), it is strongly recommended to use some kind of mail-merge function to insert the vignettes into PAPI or computer-assisted questionnaires. One might use any office program to arrange this for paper questionnaires. In the following we illustrate the single steps for Microsoft Word 2010. The same general principle of using wildcards for the vignettes in the questionnaire and linking these wildcards with information on the vignette texts provided in the setup database (in our case the Excel sheet “vignettetexts.xls”) can be used for computer-assisted questionnaires—see the information in Chapter 4 in our textbook.

4.2.2 Single Steps to be Done

Before explaining the mail-merge functions in detail (4.2.3), we provide an overview of all of the single steps that are needed to set up the questionnaire.

1) Prepare the vignette data

Sometimes you have to edit the vignette texts first—for example, there might be double spaces. Or you want to highlight some levels of the dimensions in the vignettes. Both can be achieved by means of the find-and-replace function in Microsoft Excel (*Find & Select > Replace*). Click *Options > Format* if you want to change the format of some characters. For instance, you might highlight the vignette levels in bold or italic (Note: you can edit the texts as well later on after inserting them into your PAPI questionnaires). Save the resulting data file (e.g. as “vignettes_wide_edited.xls”).

2) Prepare the questionnaire

Even when the vignettes are your core questionnaire module, you will have to add some further questions and modules. You should ask some standard questions before the vignette module. It is not recommendable to start a questionnaire with a very complex task and you will need some introductory text for the questionnaire and vignettes. In addition, you will be interested in the socio-demographic background or other characteristics of respondents. Thus, simply set up your questionnaire; you might use our template for this (**Questionnaire_template.doc**).

3) Insert the vignette module

Now add the FS module via the following steps:

- a) Implement the introductory text for the vignette module.
- b) Insert numbers for the N single vignettes per questionnaire (e.g. V1, V2, ... ,VN).
- c) Insert the answer scales for the vignettes.
- d) Use the mail-merge function (see the following instructions) to insert:
 - the single vignette texts; and
 - the ID of the vignette deck or the ID of the single questionnaire version when you are working with respondent-specific vignette orders (this variable can be inserted at the front page or in the header or footer of all pages; in any case you should guarantee that you are able to identify the vignettes that were presented to all respondents!).

4) Print and allocate the questionnaires to respondents

See 4.2.3 for an instruction on how to print questionnaires when using the mail-merge function. However, in addition you should also note the following two aspects:

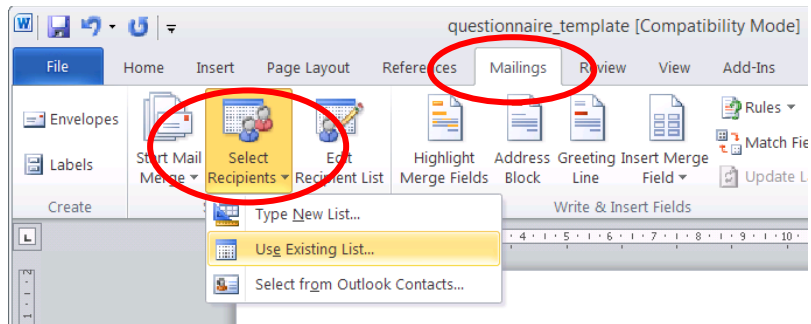
- a) **When you are using the same questionnaire version (deck) for several respondents** (no random vignette order per respondent), you should **shuffle the questionnaires to achieve a random allocation of decks to respondents**. By doing this, you might take care that the single decks are balanced (for instance, shuffle separately staples of questionnaires that contain each deck only once; you can achieve this also by means of statistical software (for an example see §4_material → Stata → 5_ordering.do). **Print on all questionnaires the deck number!** (for a description on how to arrange this with the mail-merge function, see 4.2.3).
- b) **When you are employing a respondent-specific order of vignettes for all respondents and hence idiosyncratic questionnaire versions for all respondents** you should arrange these different questionnaire versions by means of statistical software packages like Stata (see §4_material → Stata → 5_ordering.do). In this case, and when everything worked well (which should be carefully checked), the questionnaires are already in a random order and hence do not have to be manually shuffled. **Print on all questionnaires the unique ID of the questionnaire!** (for a description on how to arrange this with the mail-merge function, see 4.2.3).

4.2.3 Inserting Vignette Texts by Means of the Mail-Merge Function

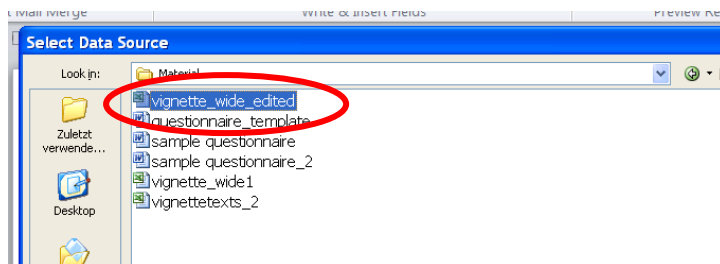
Just follow the following steps.

1) Link your Word questionnaire document with the Excel file that contains the vignette texts:

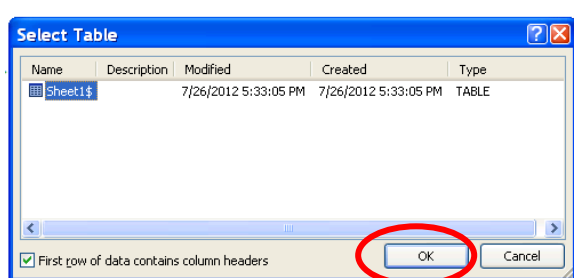
a) Mailings > Select Recipients > Use Existing List.



b) Select the Excel file with your edited vignette texts.

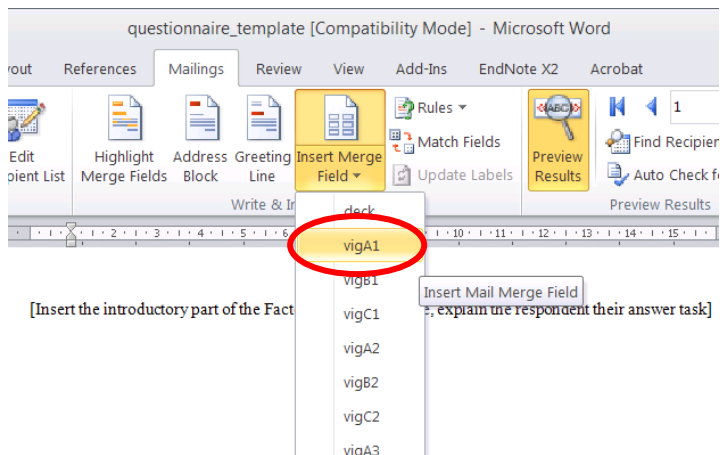


c) Click OK.

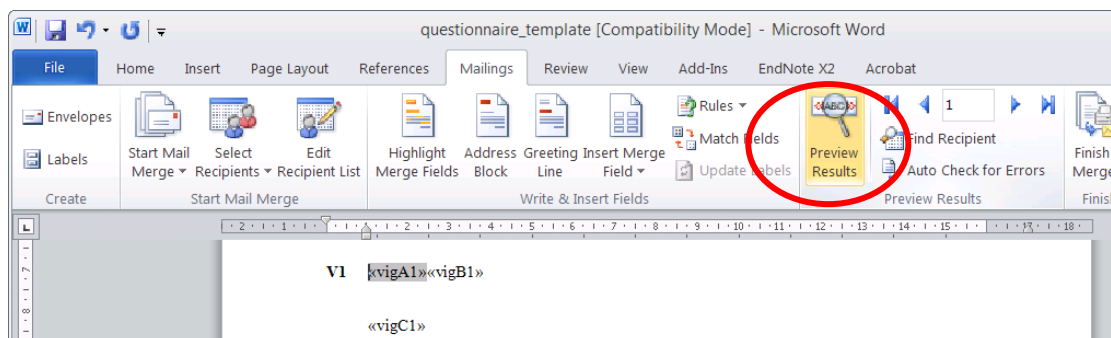


2) Insert merge fields that correspond to the single vignette phrases

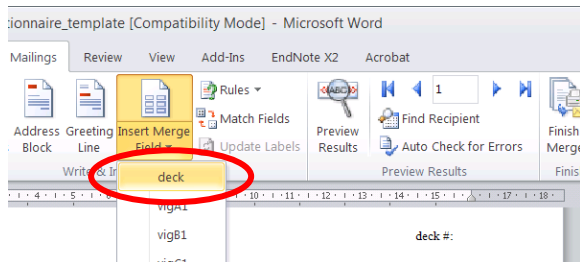
- a) Click *Insert Merge Field* and select the correct merge field; for instance, in our example we would have to insert vigA1, vigB1, and cigC1 to represent the first vignette, and vigA2, vigB2, and vigC2 for the second vignette and so on.



- b) You may add any spaces or paragraphs between the single vignette phrases or change their format by means of the standard Word functions.
- c) When only the wildcards but not the vignette texts are displayed, click "Preview Results".

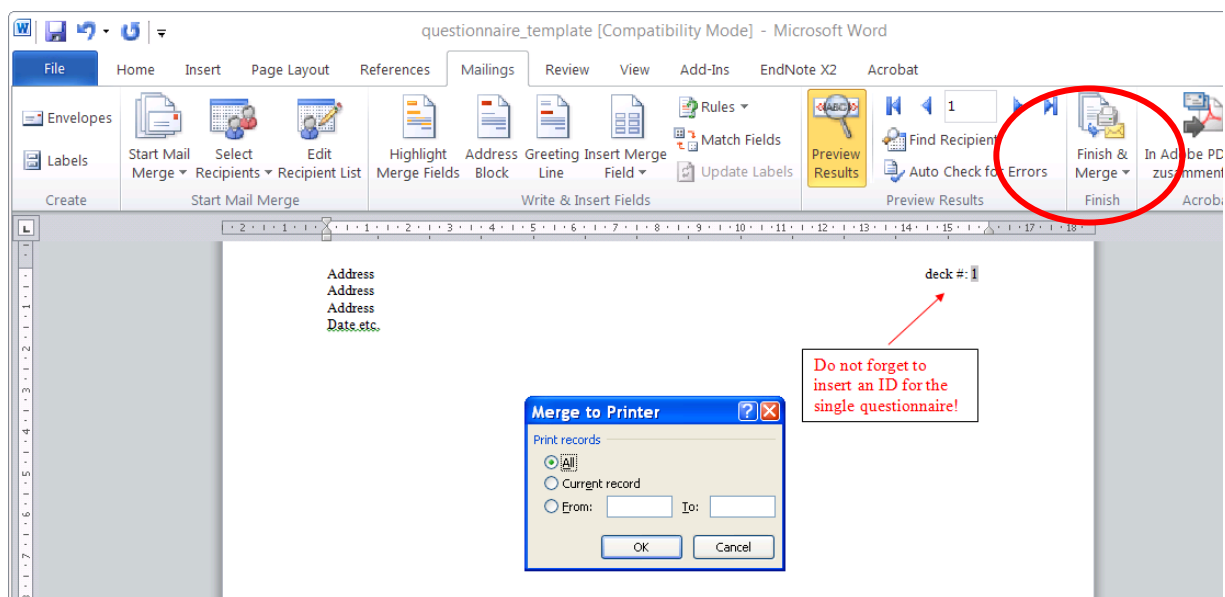


- d) Do not forget to **insert the merge field showing the ID of the vignette deck or the idiosyncratic questionnaire version** in case you are working with random vignette orders per respondent! For instance, you might insert this variable on the front page of the questionnaire or in the footer of each page.



3) Print the questionnaires

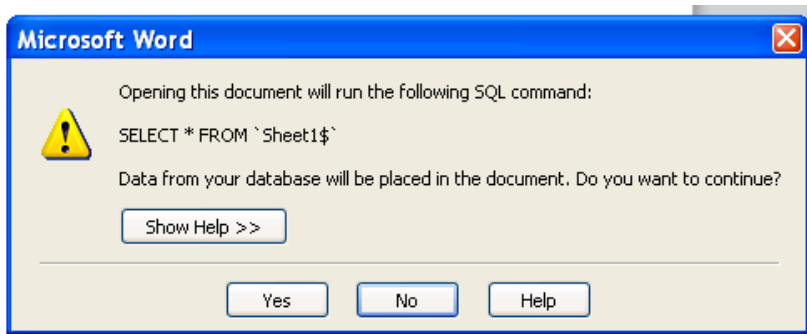
- a) Click *Finish and Merge*. You are then asked if you want to print all questionnaires or only a selection.



- b) You might print the questionnaires first as PDF to check that everything has worked well.

4) Additional Remarks

- a) You will be asked if you want to insert the merge fields each time you re-open the questionnaire:



Just click "Yes"!

- b) Before going to field, you should check if you are really able to identify the vignette information shown to respondents in the different questionnaire versions. Do the vignette texts displayed in your questionnaires correspond to the information in your setup data? For instance, does the first vignette of a questionnaire with ID X really show a 25-year-old man when the variables in the Stata setup data indicate the information for this vignette case (age == 25, sex == 1)? You should check at least some vignettes and questionnaire versions to make sure that everything was correctly done.
-