



Guidelines for Tracer Studies and Handbook of Survey Data Management in TVET Institutes

Publisher:

Vietnamese-German Development Cooperation
Programme Reform of Technical and Vocational Education and Training in Viet Nam

General Directorate for Vocational Training (GDVT)
37 B Nguyen Binh Khiem Street
Hanoi, Viet Nam

Tel. +84 4 397 45 207 (Department of Administration and International Affairs)
Fax +84 4 397 40 339

Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH
2nd Floor, No. 1, 17 Alley, Ta Quang Buu Street
Hanoi, Vietnam

Tel: +84 4 397 46 571/-2
Fax: +84 4 397 46 570
Website: www.tvet-vietnam.org

Author: Christoph Ehlert and Prof. Dr. Jochen Kluge (RWI/on behalf of GIZ)

Contributors: Nguyen Hoang Nguyen (NIVT/GDVT)
Beate Dippmar, Le T. Hong Nhi, Nguyen T. Bich Ngoc (GIZ)

Photo (cover page): Ralf Bäcker (Berlin)

Cover page layout: Mariette Junk (Berlin), Nguyen Cong Trang (GIZ)

Place and year of publishing: Hanoi, 2011

This Handbook was written to assist Vietnamese vocational training institutes in conducting tracer studies. It represents the output of several training workshops intended to build up knowledge about the steps of implementation of tracer-studies and about data-management and evaluation.

Contents

- 1 Relevance and objectives of tracer studies 3
- 2 Selected aspects of questionnaire design 3
 - 2.1 Defining the objective of a survey 3
 - 2.2 Content and structure of a questionnaire 3
 - 2.3 Response rate and type of question 4
- 3 Steps of implementation 5
- 4 Introduction into Excel as a data management tool..... 7
 - 4.1 Starting with Excel..... 7
 - 4.2 Inserting, copying and formatting 8
 - 4.3 Functions and formats 11
 - 4.4 Importing & Printing..... 13
- 5 Sampling..... 18
 - 5.1 Introduction into sampling 18
 - 5.2 Sampling with Excel..... 19
- 6 Coding questionnaires with Excel 30
 - 6.1 Structure of the data-entry file 30
 - 6.2 Handling missing data 32
 - 6.3 Defining new option-sets 33
- 7 Automated postal mail survey 37
 - 7.1 Selecting a data-base..... 37
 - 7.2 Using data-fields 41
- 8 Merging data from different data-sets (baseline and follow-up)..... 45
 - 8.1 Importance of a consistent person identifier..... 45
 - 8.2 Creating a new identifier 45
 - 8.3 Merging data 53
- 9 Creating overview statistics 55
 - 9.1 Pivot tables to summarize questionnaire information..... 55
 - 9.2 Recoding of data 66
 - 9.3 Creating averages and other statistical information..... 71
- 10 Reporting and interpretation of the results..... 75

- Appendix 1 - Baseline questionnaire 77
- Appendix 2 - Follow-up questionnaire..... 80

1 Relevance and objectives of tracer studies

Tracer studies can be an effective tool for following-up the situation of graduates from any educational institution (professional secondary school, college, university). They allow collecting information on labor market success of graduates, on the effectiveness, adequateness, quality of the training, and on many other aspects which makes them relevant for the steering of a market-oriented educational and vocational training system as well as the on-going TVET reform process. Tracer studies typically consist of a baseline and at least one follow-up survey. While the objective of a baseline is to create awareness of the study and to collect pre-graduation information, the follow-up has the objective to collect post-graduate information on labor market status (employed/not employed), job characteristics, relevance of the training and on other data of interest. By means of the collected information, tracer studies not only allow a comparison of training institutions with respect to their training effectiveness, but also an adjustment of training curricula to market needs. Furthermore they can give direct feedback to the training institutions on the quality of training by means of subjective evaluation but also by means of hard facts, such as the number of students that found a job in a given period of time.

2 Selected aspects of questionnaire design

2.1 Defining the objective of a survey

Before starting the design of the questionnaire, the aim of the survey has to be made clear. This is important for two reasons: (i) A clearly defined aim allows asking the right questions and (ii) it allows sticking to the smallest possible set of questions. To achieve this, a baseline survey is done in advance. It has the objective to raise awareness for the study and to collect contact information to ensure that the student will receive the tracer questionnaire. The tracer survey aims at collecting information on the quality of training and on the labor market status. In the following part, we will give some insights into structuring of the questionnaires especially TVET monitoring tracer questionnaire and into some important issues connected to the way questions are surveyed.

2.2 Content and structure of a questionnaire

Depending on the subject surveyed, a questionnaire consists of several sections that are thematically congruent and simplify orientation. The design of a baseline questionnaire is usually simpler than that of a follow-up questionnaire. The TVET baseline questionnaire (see appendix) can be divided into several parts: A) Demographical information (age, gender, and educational background), B) future plans (personal expectation for future study or employment) and C) contact information. In case of the TVET survey, part C) deserves special attention as it helps to ensure smooth communication for the tracer part of the survey.

The tracer questionnaire of the TVET monitoring survey consists of four sections, namely a section (A) on personal information, (B) on employment and training, (C) on non-employment and (D) on training evaluation.

Guidelines for Tracer Studies

Section (A) on personal information would usually also comprise questions on education and, if necessary, on family background and other demographic questions. However, the former topic is already part of the baseline questionnaire and therefore not asked again in the tracer questionnaire. Although the same is true for the other questions asked in the personal information section, they are surveyed again to allow moderate and easy to answer entry into the questionnaire.

Section B deals with topics related to employment situation and on-the-job training of the graduates. As some employment characteristics such as wage and working hours may also be influenced by firm characteristics, they are also part of this section. In section B, there is one separate part for non-employment to shorten the employment part and to underline that this part is only to be answered by non-employed. To speed up the answering of questionnaire, there can be filter question in this section for graduates who are not working but continue to further study. Although filter questions might speed up the answering of a questionnaire, they at the same time increase its complexity and the likelihood that interviewees give answers to wrong sections.

The section C covers information regarding further education of the graduates who are not looking for job but decided to continue study. .

The section D which is the last section of the questionnaire is reserved for the training evaluation. Most of the questions here work with statements that have to be answered on a five point scale to indicate the degree of agreement with the statement. The use of a “neutral” middle category has some advantages and drawbacks. An advantage is that individuals who are really indifferent can make use of this category. Without the middle option he might skip the question. Therefore, the middle category may be a way to increase item-response. An example for the drawback would be a student who does not agree with the statement that the training was good but at the same time does not want to reveal such a judgment, she/he can also select this option. Therefore, it may reduce the accuracy of the assessment of training quality.

2.3 Response rate and type of question

As the response rate is an important issue in mail surveys, we designed the questionnaire accordingly. The response rate is likely to be higher with a short questionnaire and questions that are easy to answer, because the interviewee has to invest only a limited amount of time. Therefore, we opted for a short questionnaire with multiple choice questions. When using multiple choice questions, it is very important to know all important possible answers or at least to have a residual category which allows respondents to indicate that they paid attention to the question but found no appropriate answer category. Additionally, a multiple choice question can have an open component to allow for other answers when the given answer categories do not apply (see for example question 7 in the follow up questionnaire). This kind of multiple choice questions can also be useful when asking for sensible

information. Question 8 is such an example. For individuals who do not want to reveal the exact amount of their income, we have a multiple choice closed question component, while the others can answer the open part of the question.

Apart from the use of multiple choice questions, the interview time can be reduced by using filter questions. These questions guide respondents through the questionnaire and allow them to skip certain questions or whole parts of the questionnaire (question 14 in the follow-up questionnaire is an example for such filter questions). This is another reason why the organization of a questionnaire into thematically congruent sections is important.

3 Steps of implementation

The following implementation guideline requires that the questionnaire is already at hands. It assumes the use of a baseline, which is very important to increase awareness of the tracer study. This will help to improve the response rate to the tracer questionnaire and will also allow to build-up or update the necessary address data-base.

- 1) The first decision to make is whether you want to survey the total population of graduates (1) or only a representative part of them graduates (2). This decision largely depends on the number of graduates you have in a given year, your demand of information and on your capabilities of surveying a large number of students. In case of (1), all students in the year of graduation are part of the sample and enter into the survey. Otherwise (2), you have to randomly draw students which are in the year of graduation (see chapter 5).
- 2) When the decision regarding the size of the sample is made, the next step depends on whether the students already graduated and left school or whether they are still at school at the point in time you want to implement the baseline survey. When the students are still in school, you may organize a collective information meeting to create awareness about the tracer study and to fill out the baseline-questionnaire. Given that the students have already left the school at the point in time you want to implement the baseline survey, you may either use alumni meetings to distribute the baseline questionnaire or utilize a combined tracer questionnaire that also comprises information of the baseline questionnaire. Please note that it is VERY IMPORTANT at the baseline stage of the survey to create a unique identification number for each student that you can use in the follow-up part of the tracer study again. A name or year of birth are not sufficient, as several names are not unique and matching students by name and birthday is error-prone and time-consuming. The identification code may have the following form: YYLXXXX, where the first two digits (YY) reflect the year of graduation, the third digit (L) the level of training courses (1=college level and 2=secondary level level) and last 4 digits (XXXX) the personal identification number of a student. An example of such a number would be 1120003 for student number 0003 who graduated in 2011 on secondary level (2).
- 3) After the baseline has been created, please record the reason of non-participation for those who were selected into but did not participate in the baseline survey. Reasons might be refusal, no-show, or unable to contact. Subsequently, enter the

Guidelines for Tracer Studies

baseline-questionnaire data into the Excel sheet provided with the survey. If no Excel template was provided, you will find the necessary information for creation of such a template in the later chapters of this handbook. Finally it is very important to keep the students' addresses up to date by means of email-contact or address change notification forms handed out to the students.

- 4) Five months after graduation, please send out emails to those email-addresses you collected with the baseline questionnaire to ask for postal address changes and changes in phone numbers.
- 5) Six months after graduation, please send out the tracer questionnaires with postal mail, together with a stamped, self-addressed envelope. Make sure that the unique identification number of each student (created during the baseline survey) is printed on each questionnaire. This is necessary to allow an unambiguous matching of the tracer information with the baseline information collected before. In case of return-to-sender letters flag the corresponding addresses in your data-base as invalid and update these invalid addresses by phone. Finally, send out the tracer questionnaire again to updated addresses.
- 6) Two weeks after tracer questionnaire has been sent out or distributed, you may start a telephone follow up for those students who didn't fill out the tracer questionnaire and its follow up. You may enter the answers into the excel sheet while doing the interview. Make sure that you attach students' unique identification numbers to the data collected in the interviews.
- 7) Analyze your data to monitor and improve training quality

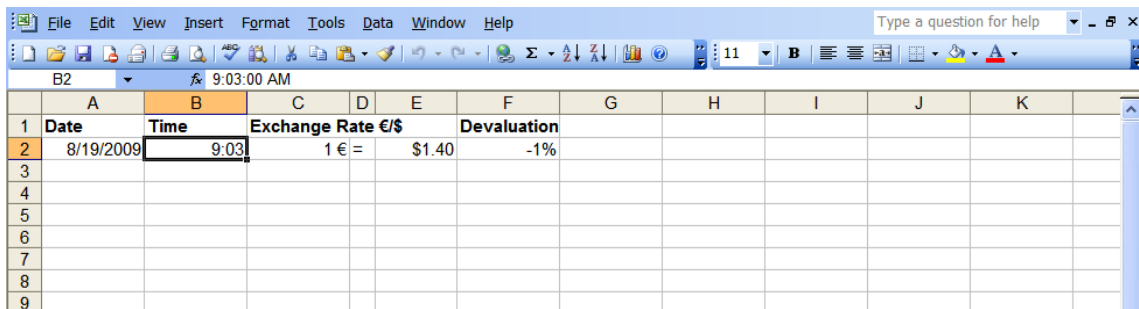
The time-frame implied by steps 4) and 5) refers to a tracer study 6 months after graduation. Longer time-spans between graduation and tracer-study have to be considered with care, because they rather tend to complicate contacting graduates due to problems in tracing students' addresses. Furthermore, commitment to the study may decline as the time span between baseline and tracer survey increases.

4 Introduction into Excel as a data management tool

4.1 Starting with Excel

In the following, we give a short introduction to Excel 2003 – the main data management tool of Tracer study. Excel is selected for this study because it is user-friendly, low cost and easy replicable software for most of TVET institutes. More advanced users might start with section 4.3. To start this tutorial, please open Excel 2003 and maximize the Excel window on your screen. To open the tutorial file, press the “Ctrl” button and the button for the letter “o” simultaneously and select the Examples_Handbook.xls file to follow this tutorial. As you already see on the example of the tutorial file, Excel files have an .xls extension as opposed to a .doc extension for MS Word Documents. Excel files are also referred to as Worksheets.

When you are working with an Excel worksheet, you can activate a particular cell simply by left clicking on it. Active cells are outlined and the corresponding row and column are highlighted at the border of the worksheet. These highlighted cells are the cell coordinates of the activated cell and refer to the row and column intersection in which the cell is found. In this case the intersection occurs at Row 2, Column B.

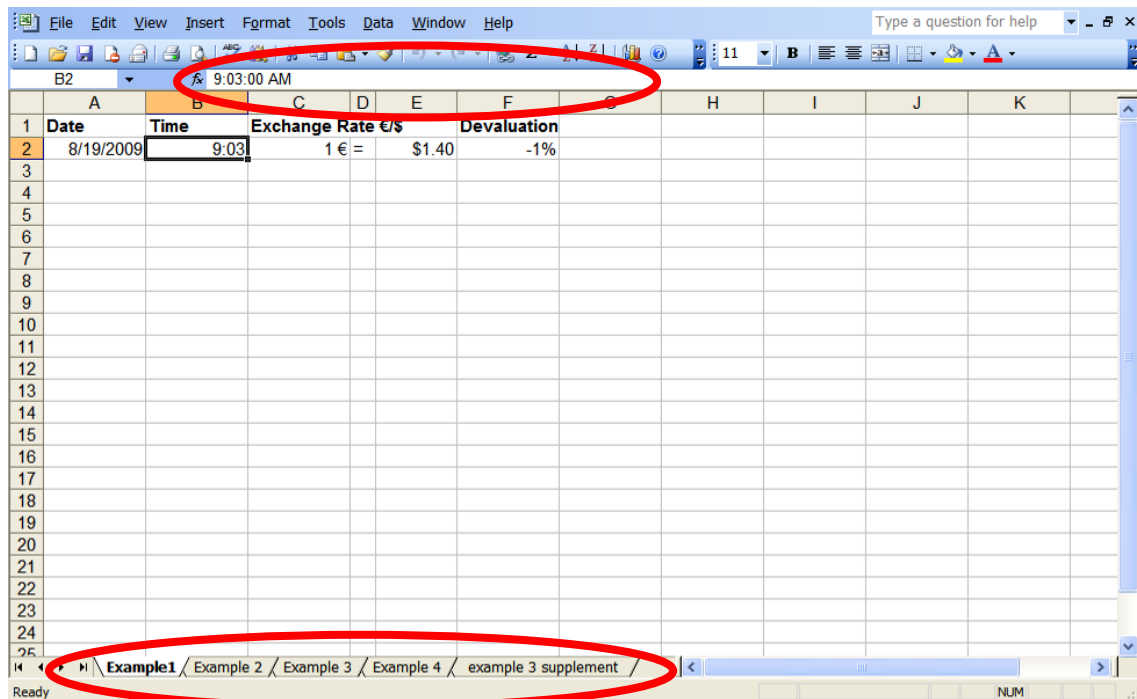


Screenshot 1: Selecting cells and identifying coordinates

Above the column coordinates area (A, B, C, ...), you find the formula bar where all the information entered for the particular cell will appear. Depending on what is entered in the cell, you may see text, numbers or formulas/functions here.

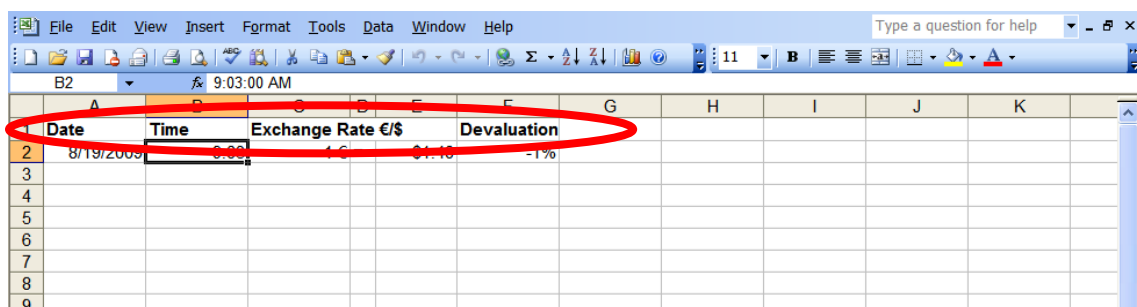
The information stored in an Excel file may be distributed among many different worksheets, which can be accessed via tabs located at the bottom of the window. They enable you to control which worksheet you are working on at a given time and allow you to navigate through all available worksheets.

Survey Data Management



Screenshot 2: The formula bar and worksheet tabs

Apart from the column coordinates, you may want to label the information in your data-set. Labels are typically used as column or row headers or simply to identify whatever you have entered into the cell.



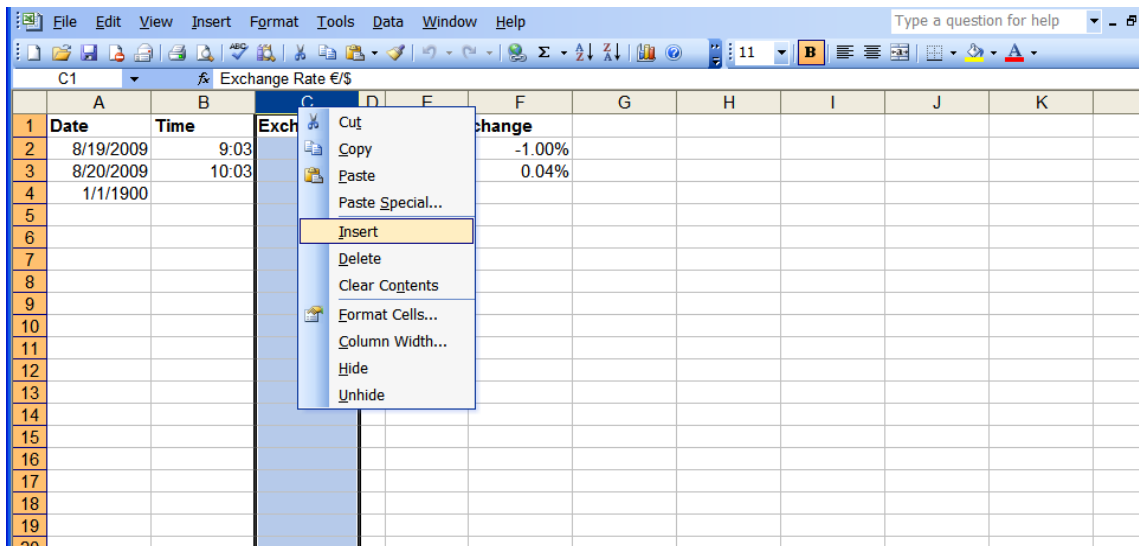
Screenshot 3: Labels

4.2 Inserting, copying and formatting

Whenever you need to add additional columns between the columns in your existing worksheet, you can add them by right clicking onto the column header of the column where you would like to add an additional column. Then click on "Insert" in the menu that appears. The new column will be added to the left of the column you selected.

Rows are inserted the same way as columns, by right clicking onto the row header of the row where you would like to add an additional row. Then click on "Insert" in the menu that appears. The new row will be added above the row you selected.

Survey Data Management



Screenshot 4: Adding columns

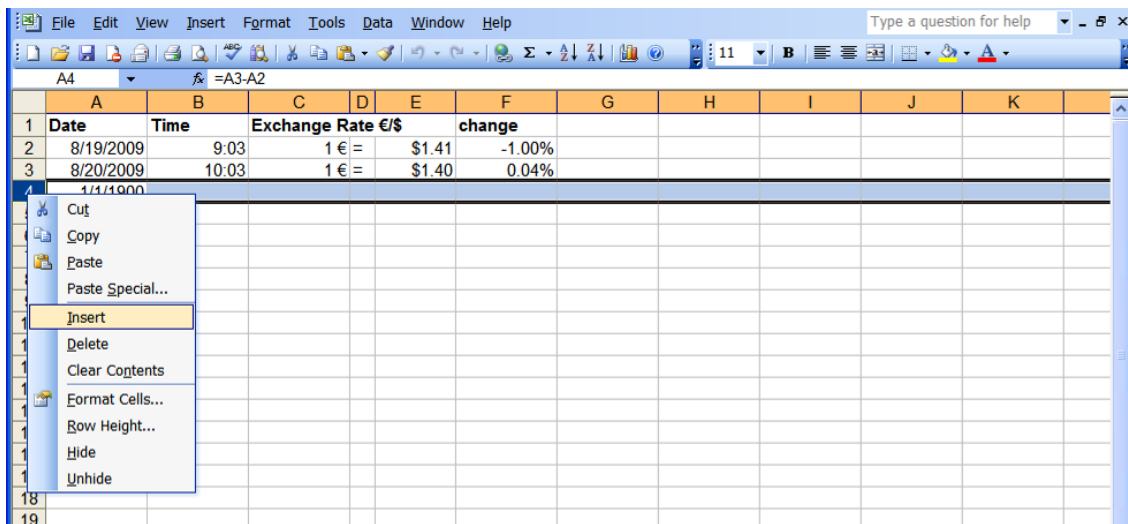
Action Insert Column:

SELECT: Column at whose left you want to insert a new column

PRESS: Right mouse button

SELECT: Insert from the list

PRESS: Left mouse button



Screenshot 5: Adding rows

Action Insert Row:

SELECT: Row above which you want to insert a new **row above**

PRESS: Right mouse button

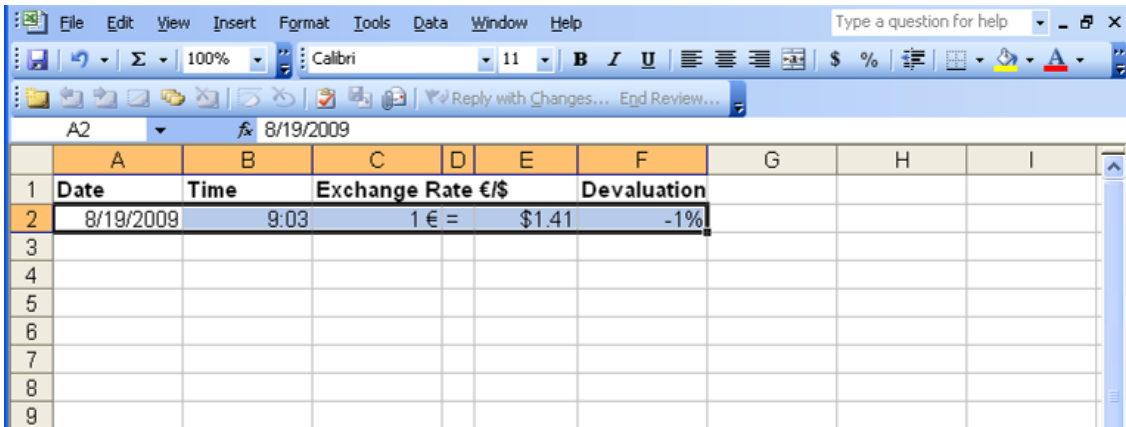
SELECT: Insert from the list

PRESS: Left mouse button

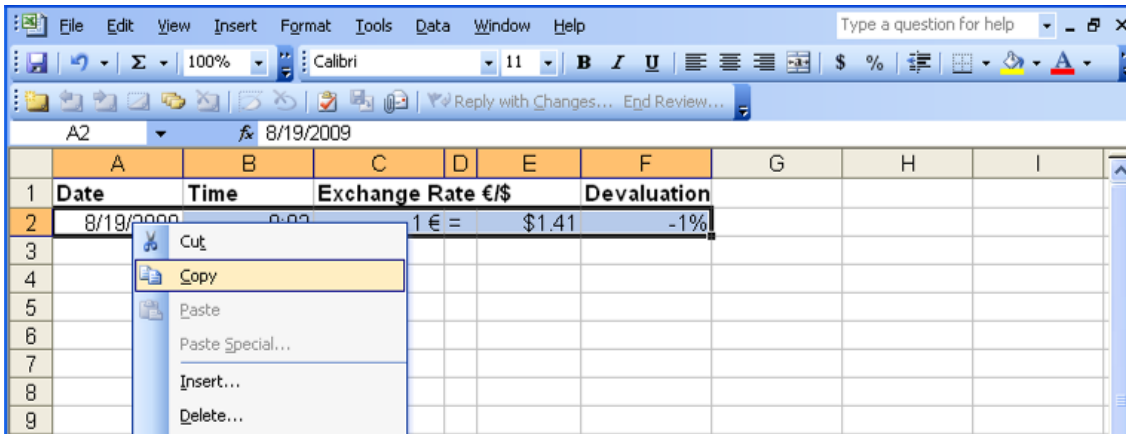
For copying cells, you first have to select the cells that you want to copy by left-clicking onto the first cell you want to select. Keep the left mouse-button pressed and move the mouse to the last cell you want to select, and then release the left mouse-button. After selecting the cells, click right on the selected cells and select “Copy” from the menu. For pasting this content back to the spreadsheet again, select the first cell, where you would like the copied

Survey Data Management

content to start and then press the right mouse-button. Select “Paste” in the menu that appears.



Screenshot 6: Selecting cells



Screenshot 7: Copying content

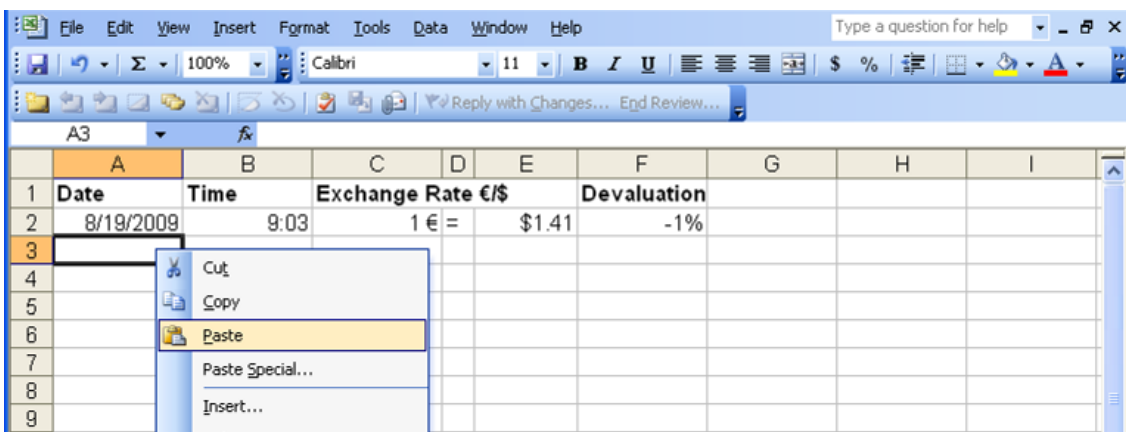
Action Copy:

SELECT: Cell that you want to copy

PRESS: Right mouse button

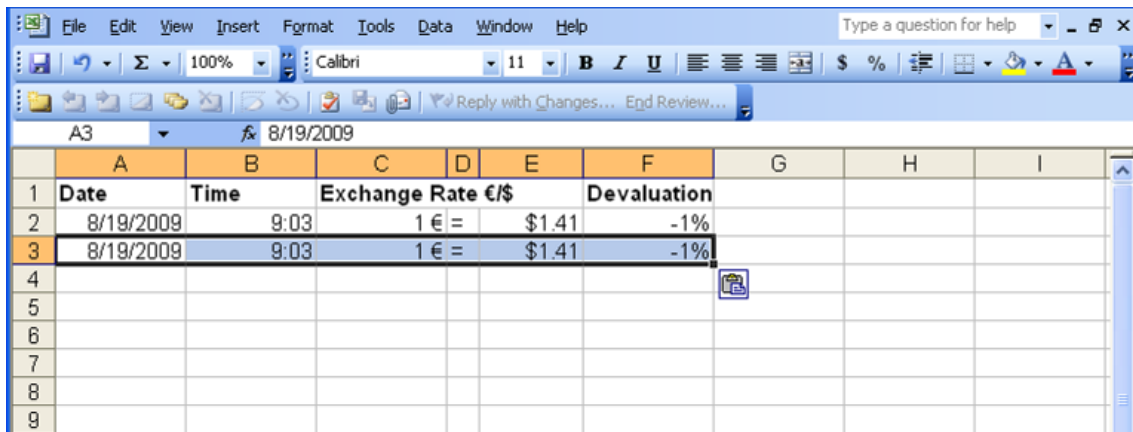
SELECT: Copy from the list

PRESS: Left mouse button



Screenshot 8: Pasting cells

Survey Data Management



Screenshot 9: Pasted cells

Action Paste:

SELECT: Cell that you want to fill with the copied content

PRESS: Right mouse button

SELECT: Paste from the list

PRESS: Left mouse button

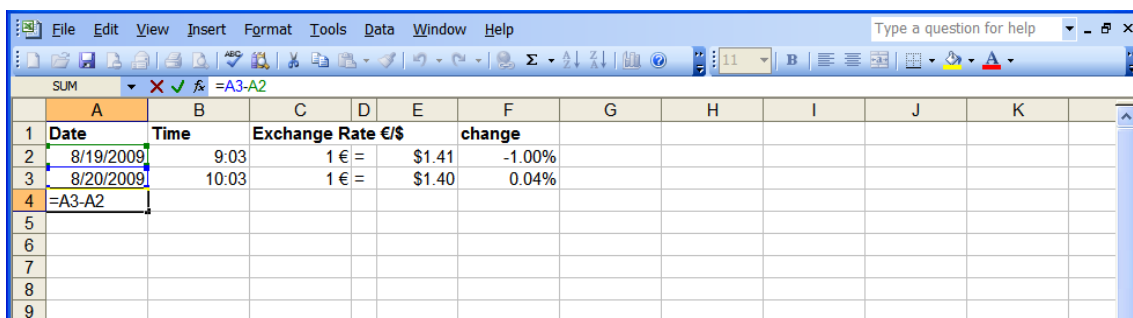
4.3 Functions and formats

Functions and formulas are expressions that you can use to manipulate or analyze information that you have entered into a cell. Functions always begin with an = sign. You can insert functions just like text or numbers. When pressing “enter”, the result of your function will appear.

For the following example, you will have to change the cells just copied and enter the following values into the row “B”:

8/20/2009	10:03	1 €	=	\$1.40	0.04%
-----------	-------	-----	---	--------	-------

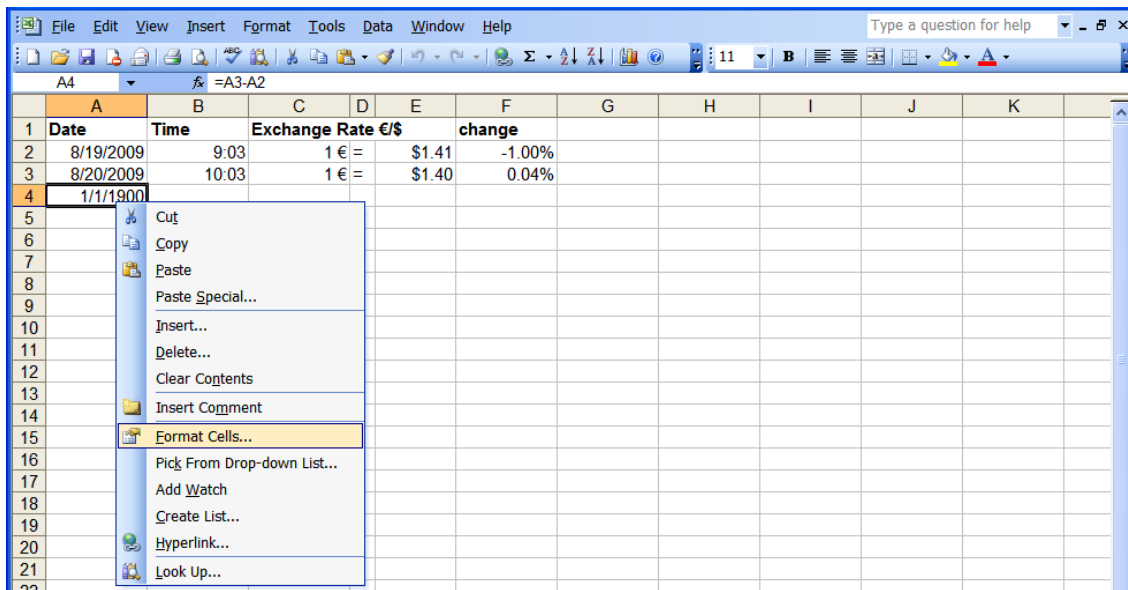
We will now use a simple formula for subtracting the contents of two cells. Please click on cell A4 and subtract cell A3 from A2 by typing: = A3-A2. As you can see, the format makes no sense for the contents anymore, because the difference of 8/20/2009 and 8/19/2009 should be 1 (day) and not 1/1/1900.



Screenshot 10: Formulas

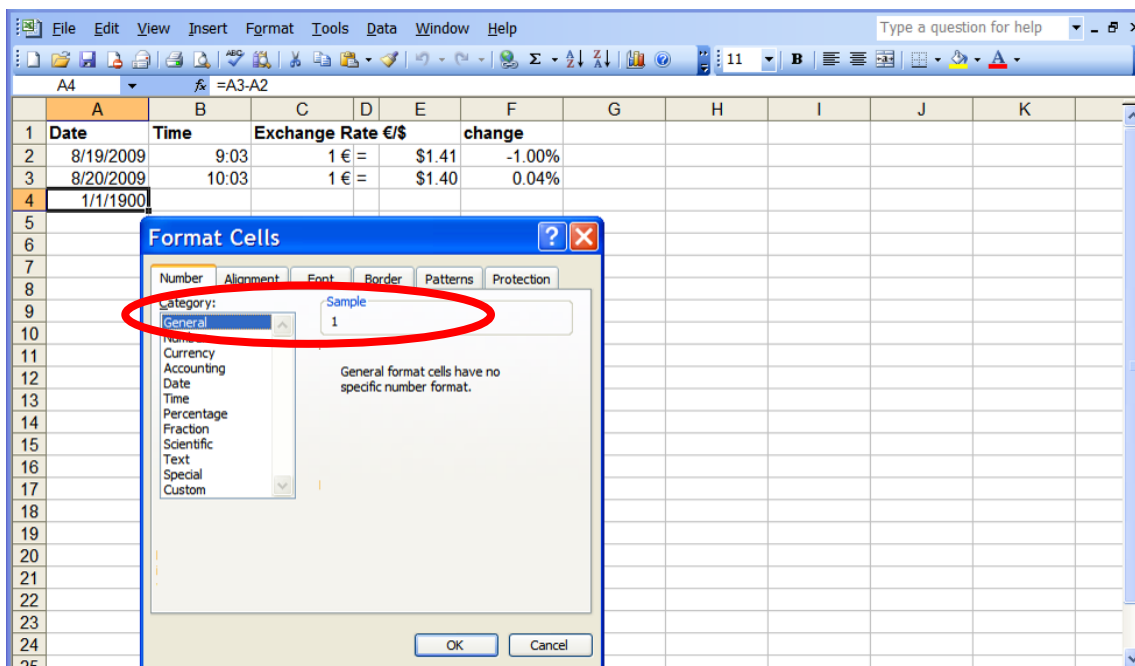
To format this cell correctly, select the cell and click on “Format cells” in the menu, a window with several format options will appear.

Survey Data Management



Screenshot 11: Formatting cells

The different format options in Excel enable you to change the format of cells to display numbers, among others, in a specific currency or in a specific time or date format.



Screenshot 12: Formatting cells menu

In case of our example, we select the format “General” in order to show the difference between the two dates not in a date format, but in a general format. In the “Sample” part of the “Format Cells” window, you see a preview of what will happen after you apply the format. In our case this is perfectly correct – the difference between the two dates will be correctly displayed as “1” after the format change.

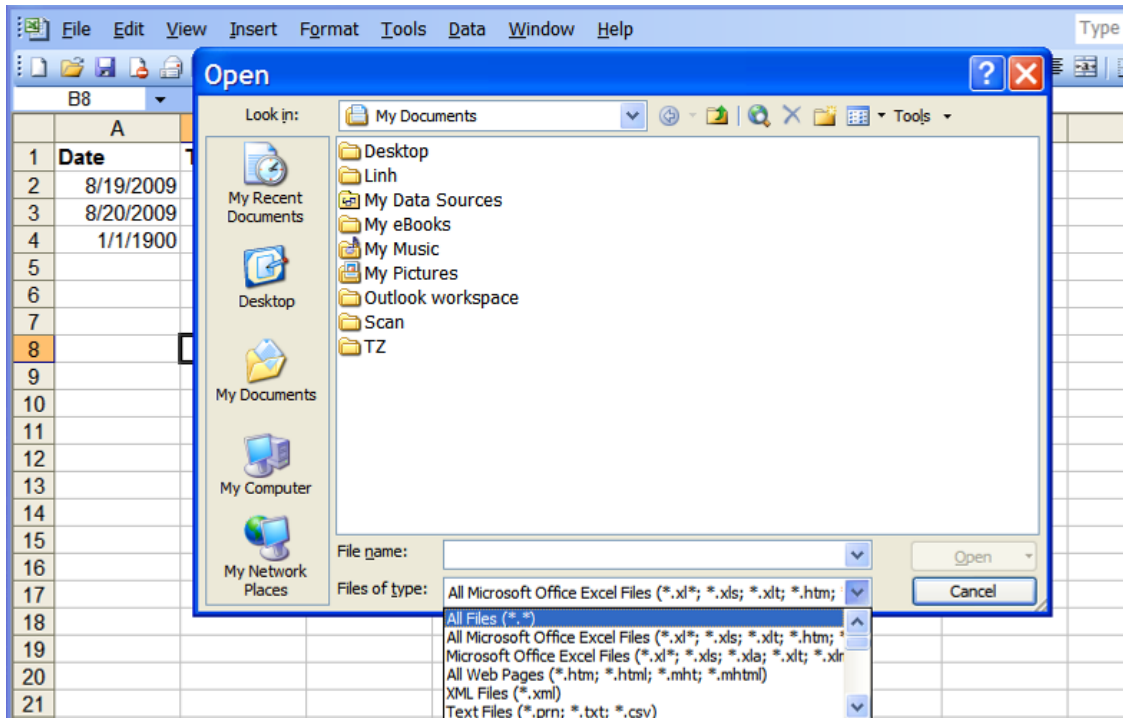
Besides these formats, you have the possibility to change other format options, such as the text appearance (Bold, Italics...) directly by selecting a format in the **Formatting Toolbar**.



Screenshot 13: Formatting toolbar

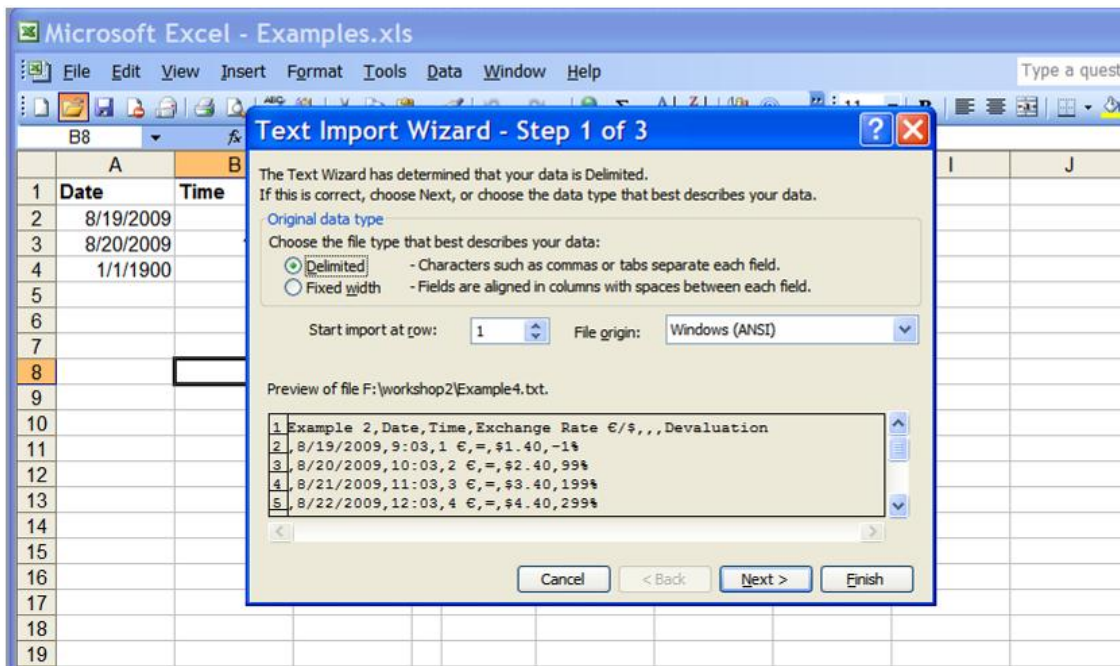
4.4 Importing & Printing

To import data into a spreadsheet, proceed with the same steps as opening an ordinary Excel file, but select "All files" from the file type menu.



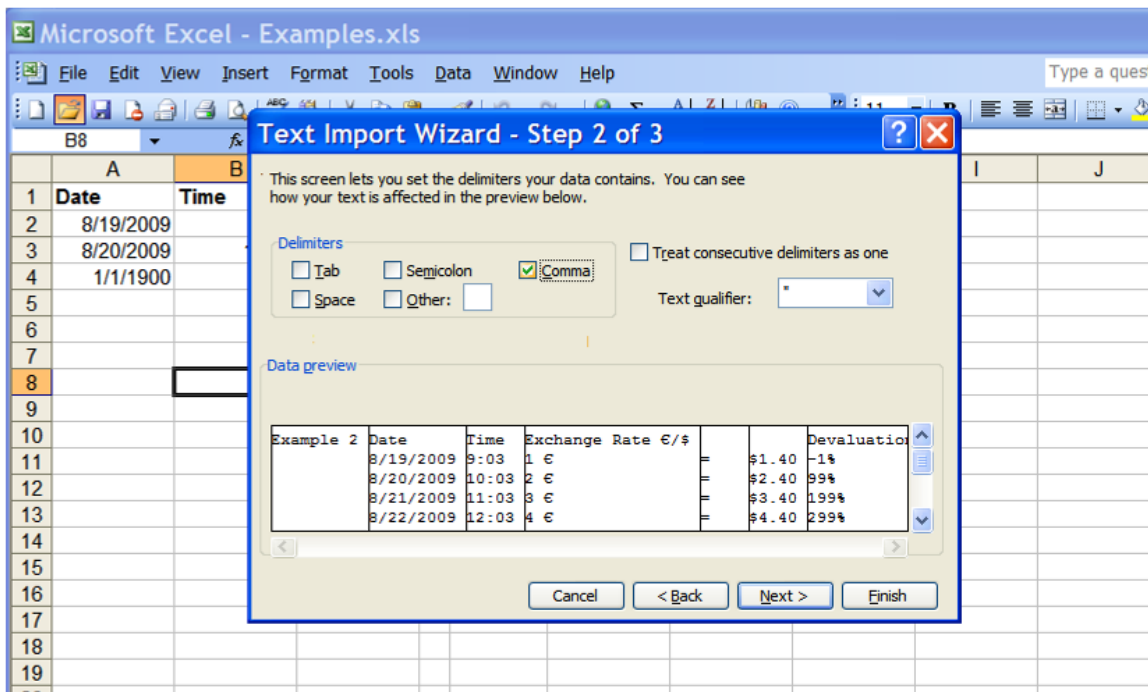
Screenshot 14: Importing data from text-file

Locate the file you would like to import. For this example, please use the comma delimited text file called example2_handbook.txt. The Text Import Wizard will launch automatically. Make sure "Delimited" is checked and click "Next".



Screenshot 15: Text import wizard

In the next menu (named “step 2 of 3”), you select the delimiter of you data. Most of the times you can see the delimiter by looking at the data preview in the “data preview” window. If you are not sure, you can also try the delimiters that are listed by checking the boxes next to them. When the right delimiter is found, you will see a well defined structure of columns and rows in the data preview window, as in our example, where the delimiter is “Comma”.

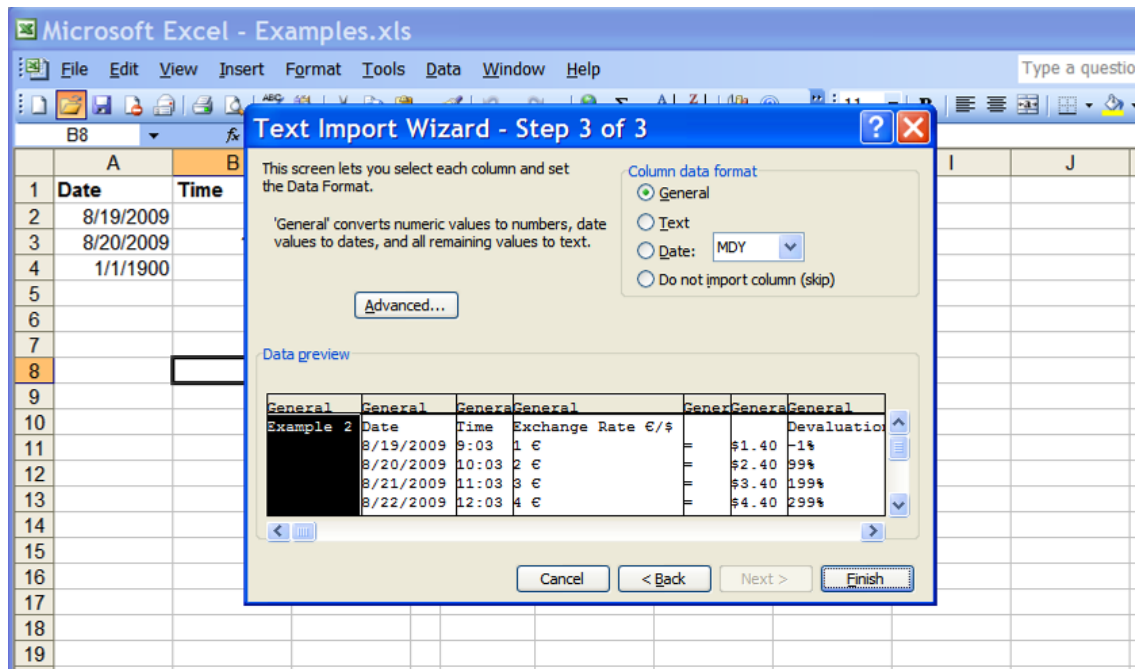


Screenshot 16: Choosing a delimiter in the text import wizard

After selecting the right delimiter, click “Next” and proceed with the last step. In this window, you can define the format of the cells that you would like to import. Simply click on the column in the preview window and select the appropriate format in the “Column data format” section. As an example, you might click on the Column that contains date formatted

Survey Data Management

cells and select “Date” in the “Column data format” section. Make sure that you select also the right date format in the drop down list right do the date option (MDY, where M=Month, D=Day and Y=Year).



Screenshot 17: Specifying the data format in the text import wizard

After clicking “Finish”, you should end up with the following spreadsheet:

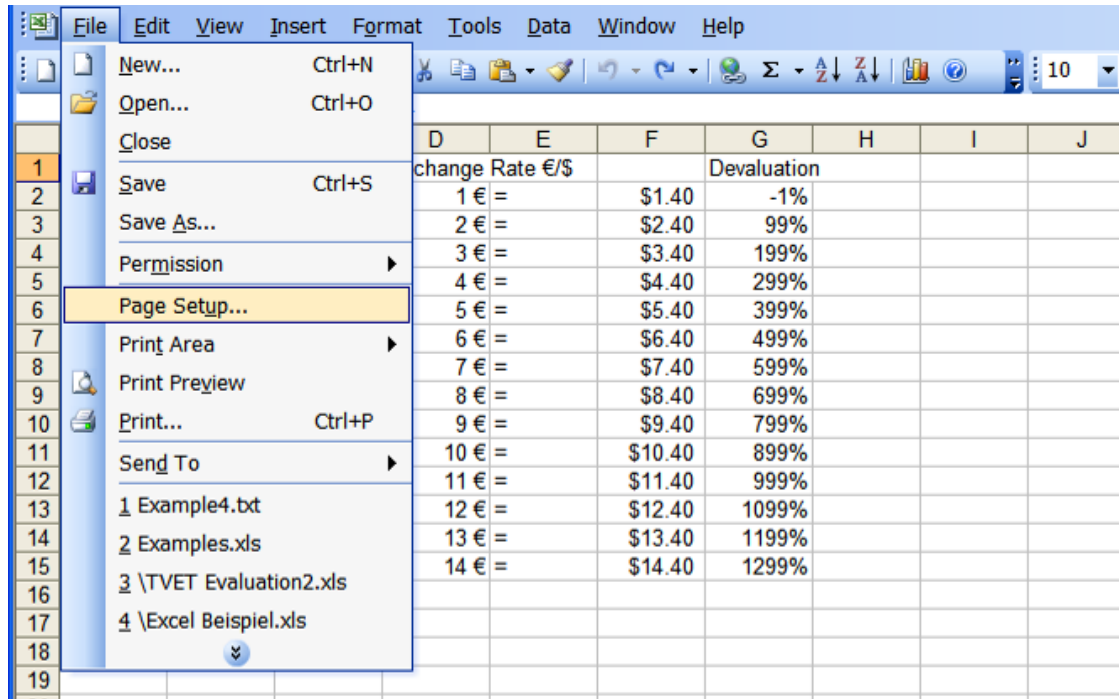
	A	B	C	D	E	F	G	H	I	J
1	Example 2	Date	Time	Exchange Rate €/€		Devaluation				
2		8/19/2009	9:03	1 € =		\$1.40	-1%			
3		8/20/2009	10:03	2 € =		\$2.40	99%			
4		8/21/2009	11:03	3 € =		\$3.40	199%			
5		8/22/2009	12:03	4 € =		\$4.40	299%			
6		8/23/2009	13:03	5 € =		\$5.40	399%			
7		8/24/2009	14:03	6 € =		\$6.40	499%			
8		8/25/2009	15:03	7 € =		\$7.40	599%			
9		8/26/2009	16:03	8 € =		\$8.40	699%			
10		8/27/2009	17:03	9 € =		\$9.40	799%			
11		8/28/2009	18:03	10 € =		\$10.40	899%			
12		8/29/2009	19:03	11 € =		\$11.40	999%			
13		8/30/2009	20:03	12 € =		\$12.40	1099%			
14		8/31/2009	21:03	13 € =		\$13.40	1199%			
15		9/1/2009	22:03	14 € =		\$14.40	1299%			
16										
17										
18										
19										

Screenshot 18: Imported data

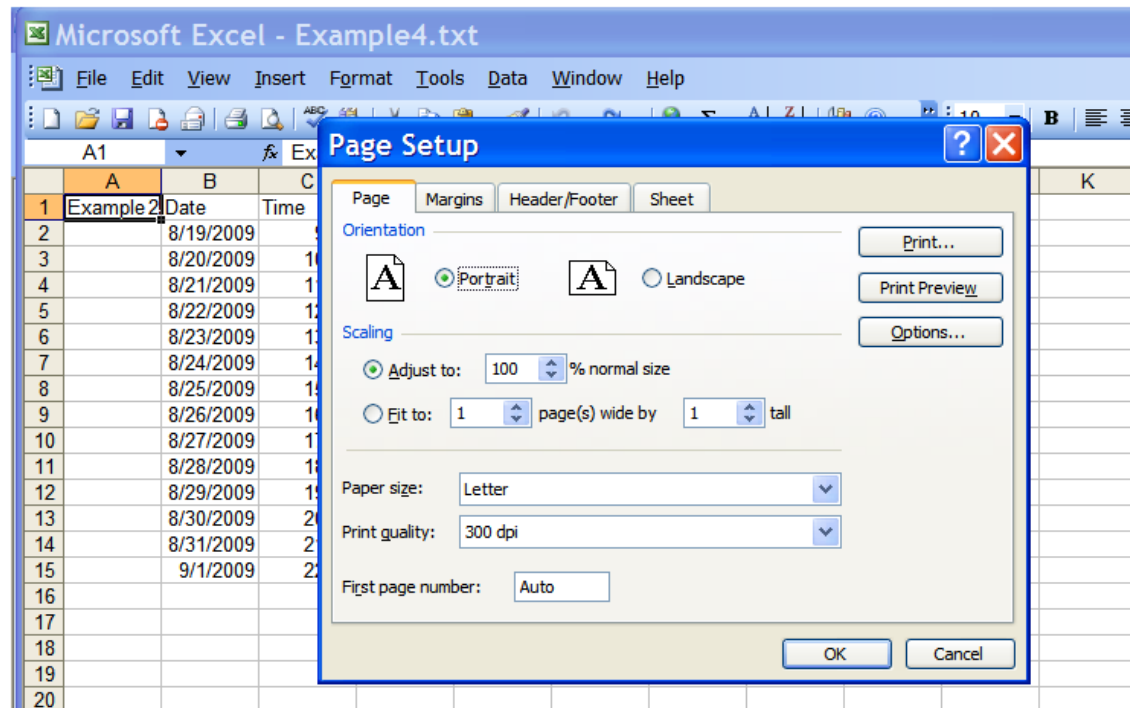
Given that you wanted to print this information, you first have to define the page setup that fits your data best. In case of few columns, it will be best to select a “Portrait” page setup. In case you have many columns to be printed, the “Landscape” format might fit your needs

Survey Data Management

best, because it allows more columns to be printed at once. To change the page setup, please go to the “File” menu and select “Page Setup...”.



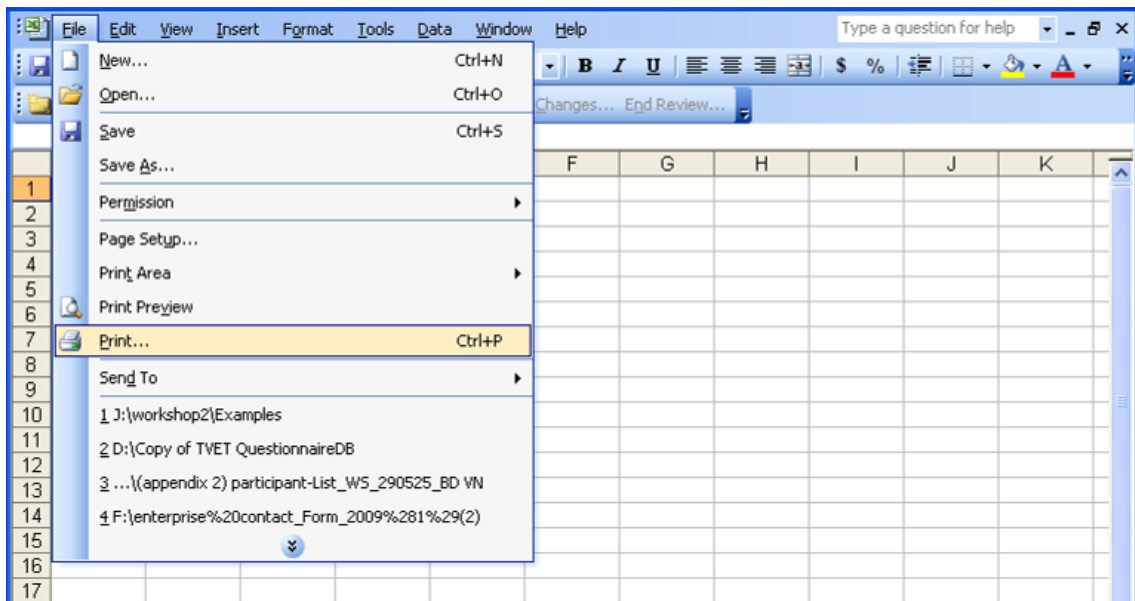
Screenshot 19: Selecting page setup for printing



Screenshot 20: Page setup window for printing

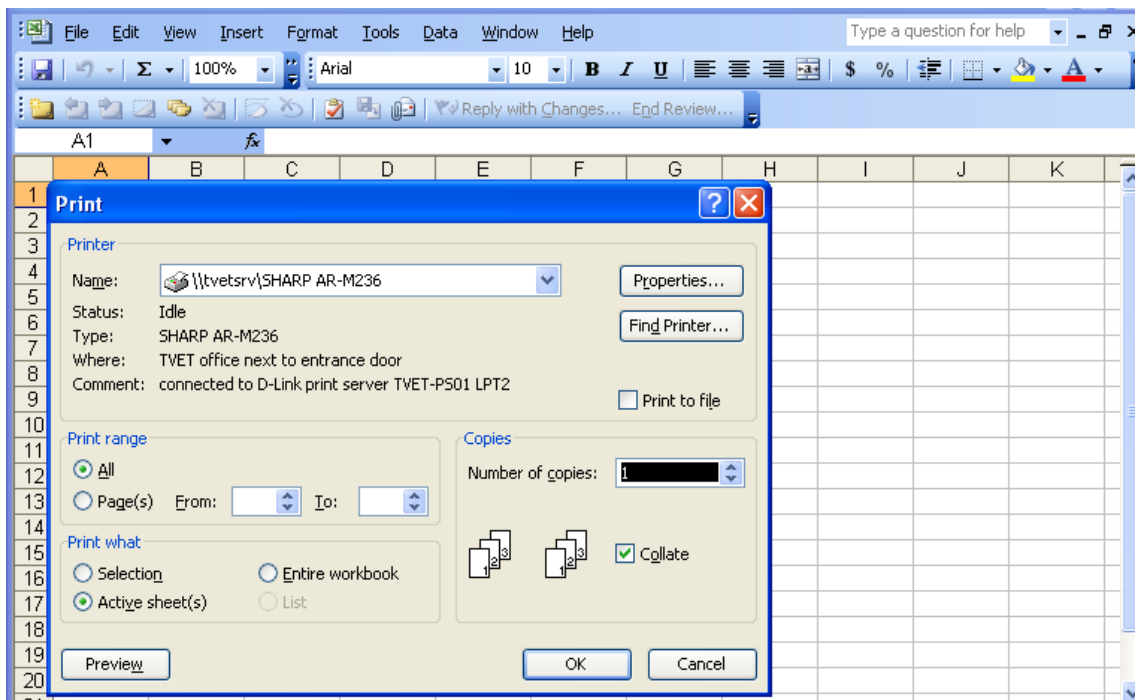
In the “Page Setup” window, you can select the Portrait and Landscape orientation, as well as the size of the pages that you want to print on. When you preview your page setup selection, you will find slim, dashed lines in your worksheet indicating the page breaks.

Survey Data Management



Screenshot 21: Printing a document.

After reviewing, you can go to the “File” menu again and select “Print” from the options to print your document. The “Print” window that appears, allows you to select the pages that you want to print. It also allows you to print only a selection of the open worksheet, meaning that only that part of the document will be printed, that you selected with your mouse before entering the print menu.



Screenshot 22: Printing dialogue

5 Sampling

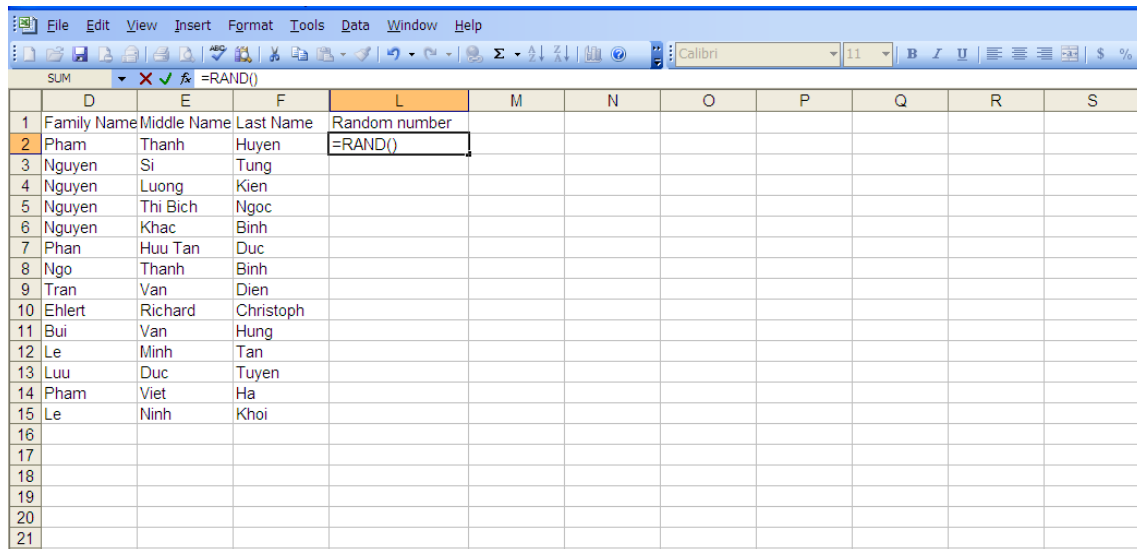
5.1 Introduction into sampling

Before we learn how to do sampling with Excel, you have to be familiar with the basic general steps of sampling. Sampling is concerned with the selection of individual observations. It is intended to yield some knowledge about the population of concern, especially for the purposes of statistical inference. In our case, this population of interest is the students in their graduation year. The properties that we measure to distinguish each observation are surveyed with the questionnaire. The sampling process can be split up in basically six steps:

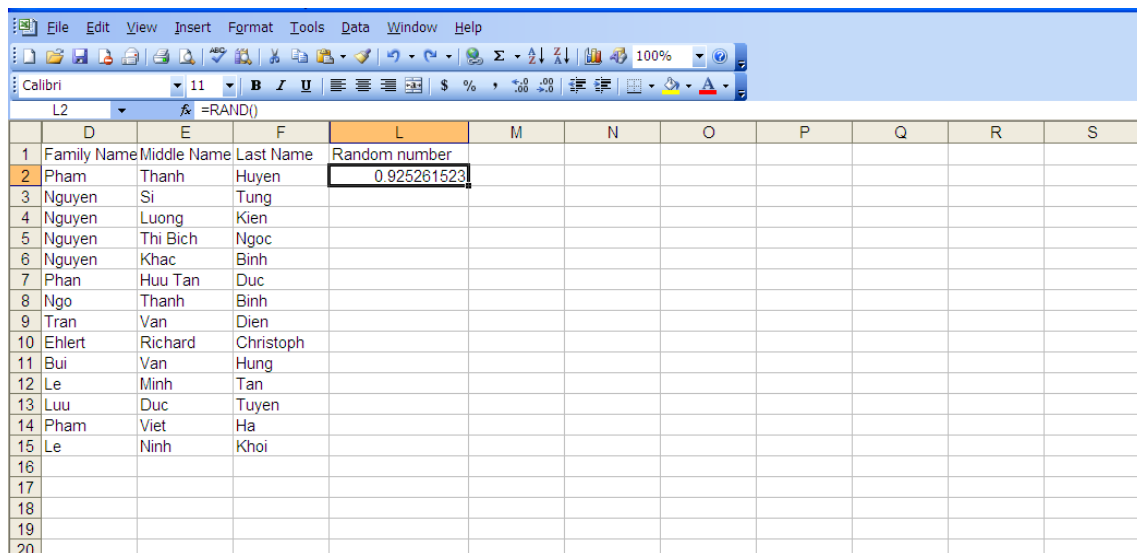
- 1) Defining the population that is important to answer the given question.
In the context of a tracer study, the population of interest is all students in the graduation year at your college
- 2) Specifying a sampling frame, which comprises each single member of the population
This is a list of all students in the graduation year at the college. It is needed in order to give every student the same chance of participation in the study.
- 3) Specifying a sampling method for selecting members of the population
We propose simple random sampling for randomly selecting students. With this method, each student has the same chance to get into the sample and to participate in the survey.
- 4) Determining the sample size
In order to allow inference from the sample about the whole population, information about a sufficient number of the students is needed. Schools with less than 400 students in the graduation year should include all of these students into the survey. Schools with more than 400 but less than 1000 students in the graduation year should include at least 400 students into the survey via random sampling. Schools with more than 1000 should include 40% of the students in graduation year
- 5) Sampling and data-collection
It is important in this step to following the defined sampling process, to note comments and to record non-responses
- 6) Review of the sampling process

5.2 Sampling with Excel

As mentioned in the first section, every command in Excel starts with an equal sign =. Followed by this sign, you type in the command that you want to execute. In case of our sampling example, this command is “RAND()”, which assigns a random number to each person. Each command is followed by an open and a closed bracket, which usually contain the parameters needed to execute the command. In case of the random number command, there are no parameters to define; therefore the space between the brackets is left empty. To execute the command, you press “enter”. Please replicate these first steps with the worksheet named “Example 3” in the Examples_Handbook.xls.



Screenshot 23: Random number function



Screenshot 24: Random number

Command Randomnumber:

TYPE: =RAND()

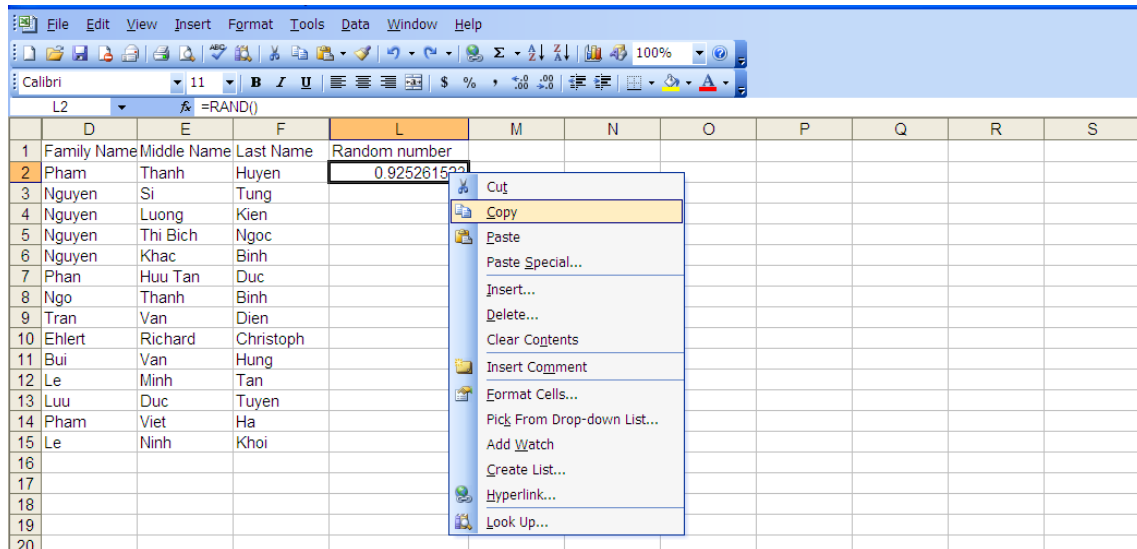
PRESS: ENTER

RESULT: Random number ranging from 0 to 1

In order to assign random numbers to all other cells too, you copy the content of the cell that contains the RAND() command and paste it to the other cells in the random number

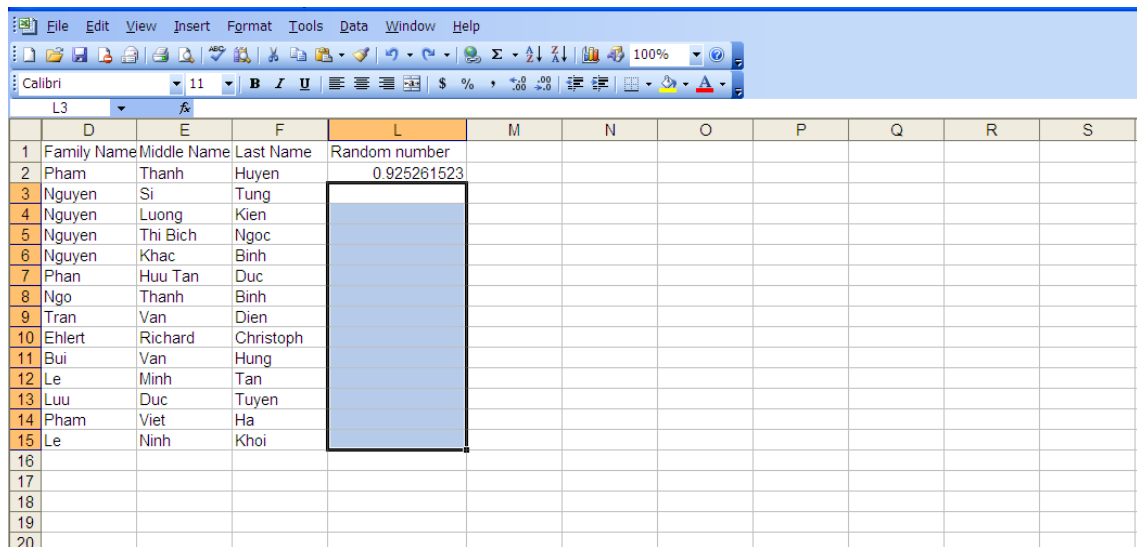
Survey Data Management

row. To do so, select the cell that you want to copy, press the right mouse button and select “Copy”



Screenshot 25: Random number – copy

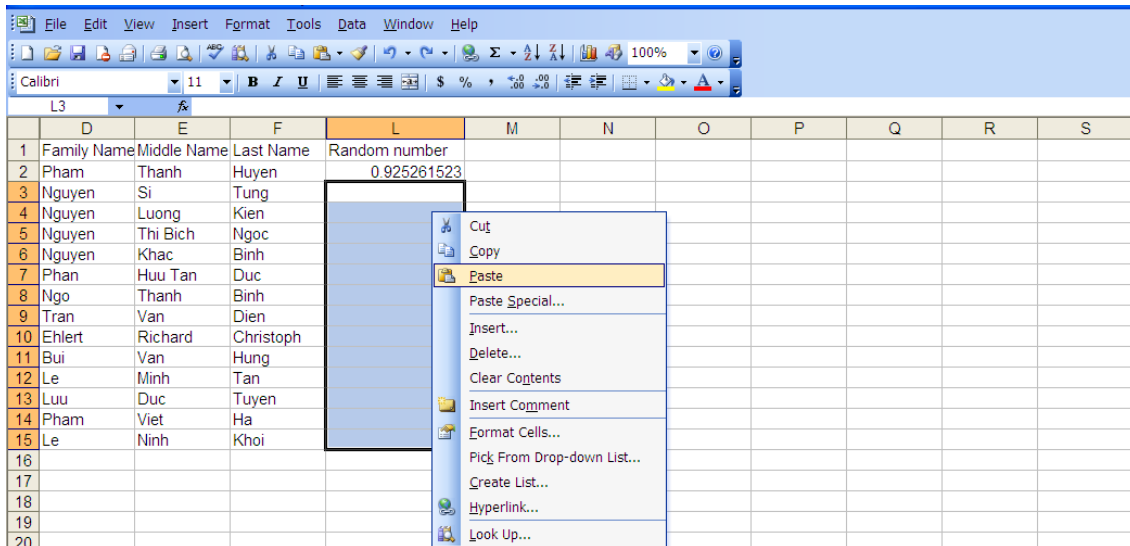
After you copied the content of the cell, you will have to paste it to the other cells. Please select the other cells that should contain the RAND() command of the copied cell, by moving the mouse to the first cell that you want to select and pressing the left mouse button. Keep the left mouse button pressed and move down the mouse to the last cell that you want to select, then release the left mouse button.



Screenshot 26: Random number – select

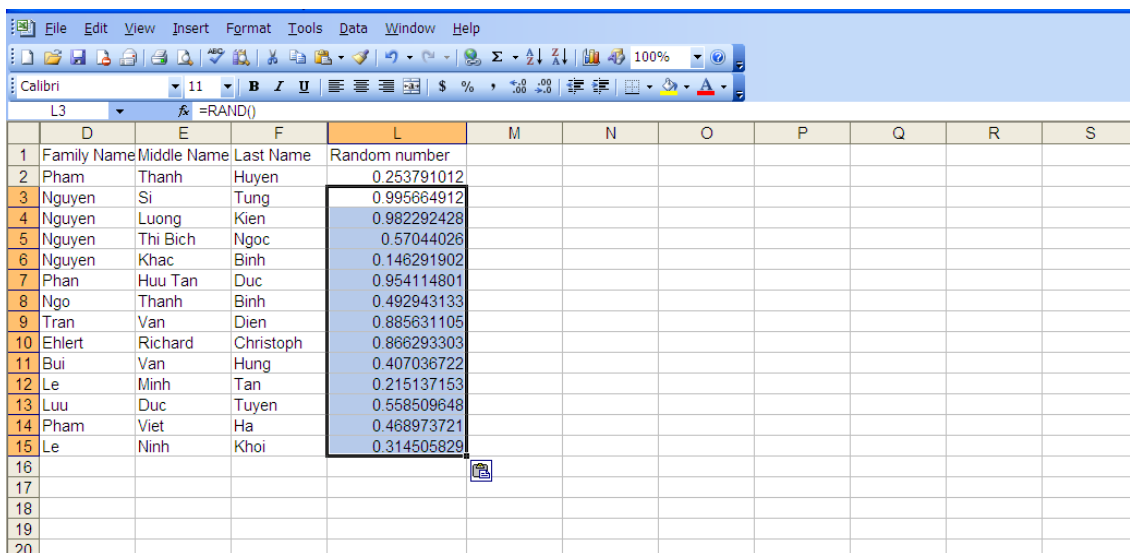
Now move the mouse cursor on the selected area, press the right mouse button and select “Paste” from the list.

Survey Data Management



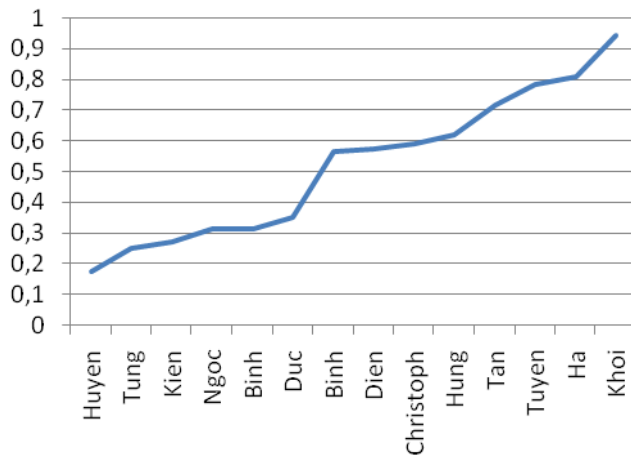
Screenshot 27: Random number - paste

After pasting the content, the other random numbers will appear. Please note, that the numbers will never be the same, since they are created randomly every time you use the command.



Screenshot 28: Random number – result

When you draw these random numbers into a graph, you will see that they are nearly equal distributed. The more numbers you create, that means the more names you have in your list, the more even the line will be.



Screenshot 29: Random number distribution

After you created the random number, you have to select a threshold or cut-off point, at which you want to select the participants. Remember, that the random number is equally distributed between 0 and 1. This means that all numbers between 0 and 1 have an equal likelihood to appear. If you selected 1 as a threshold, you would therefore select the whole sample. If you selected 0.5, you would select roughly 50% of the sample. Table 1 gives you an impression of this idea.

Eight people have a random number that is equal or higher than the threshold, we would therefore draw eight people from the population when using the threshold of 0.5.

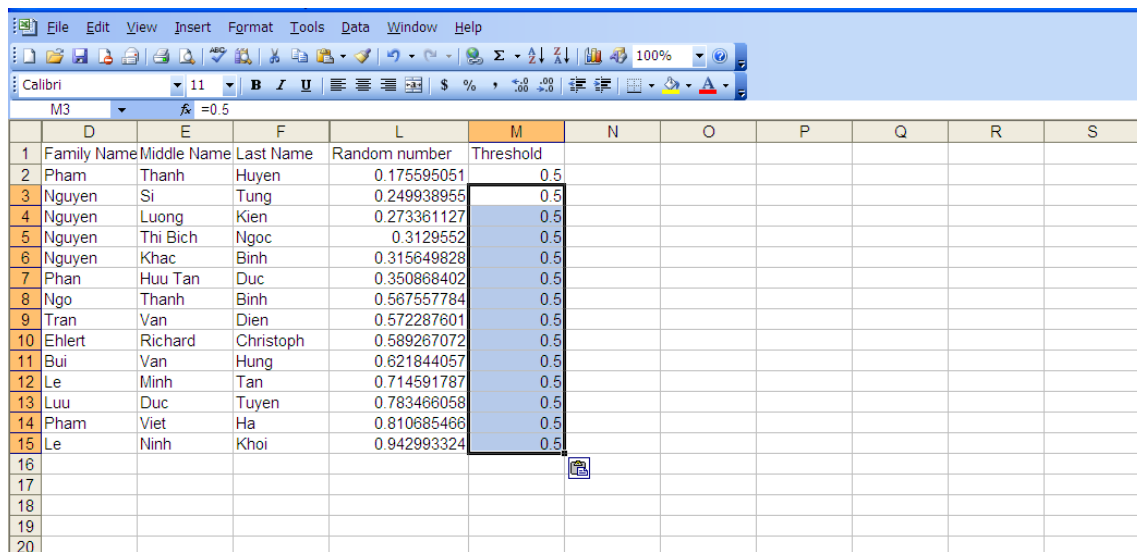
Table 1

Last Name	Random number	Cut-off
Huyen	0,17559505	0,5
Tung	0,24993895	0,5
Kien	0,27336113	0,5
Ngoc	0,3129552	0,5
Binh	0,31564983	0,5
Duc	0,3508684	0,5
Binh	0,56755778	0,5
Dien	0,5722876	0,5
Christoph	0,58926707	0,5
Hung	0,62184406	0,5
Tan	0,71459179	0,5
Tuyen	0,78346606	0,5
Ha	0,81068547	0,5
Khoi	0,94299332	0,5

To highlight all subjects that are within the threshold automatically, you will first have to create a column that contains the threshold. For this purpose, please select the first cell in the column next to the random number and type in the threshold: Type the equal sign, followed by the threshold 0.5 and press enter. To copy this threshold to the other cells, copy the cell with the threshold and paste it to the other cells as you did it with the RAND() command: Move the mouse to the threshold cell, press the right mouse button, select COPY, press the left mouse button. Then select the cells where you want to copy the content to

Survey Data Management

and move the mouse over the selected area. Press the right mouse button, select PASTE and press the left mouse button. You should end up with a spread sheet like the following.



	D	E	F	L	M	N	O	P	Q	R	S
1	Family Name	Middle Name	Last Name	Random number	Threshold						
2	Pham	Thanh	Huyen	0.175595051	0.5						
3	Nguyen	Si	Tung	0.249938955	0.5						
4	Nguyen	Luong	Kien	0.273361127	0.5						
5	Nguyen	Thi Bich	Ngoc	0.3129552	0.5						
6	Nguyen	Khac	Binh	0.315649828	0.5						
7	Phan	Huu Tan	Duc	0.350868402	0.5						
8	Ngo	Thanh	Binh	0.567557784	0.5						
9	Tran	Van	Dien	0.572287601	0.5						
10	Ehlert	Richard	Christoph	0.589267072	0.5						
11	Bui	Van	Hung	0.621844057	0.5						
12	Le	Minh	Tan	0.714591787	0.5						
13	Luu	Duc	Tuyen	0.783466058	0.5						
14	Pham	Viet	Ha	0.810685466	0.5						
15	Le	Ninh	Khoi	0.942993324	0.5						
16											
17											
18											
19											
20											

Screenshot 30: Threshold

For selecting the participants based on this threshold and the random number, you will have to use the IF function. The IF function evaluates, whether an expression fulfills a criterion that you define and executes conditional on this evaluation two possible actions. In our case, this criterion is that the random number should be larger or equal to the threshold. In case this is true, the IF command executes the content of “value_if_true”, if this is wrong, it executes “value_if_false”.

Command IF:

SYNTAX: =IF(EVALUATE; “value_if_true”; “value_if_false”)

EVALUATE, evaluates an expression. In case this expression is true, it executes “value_if_true”, otherwise it executes “value_if_false”. “value_if_true” and “value_if_false” can be both, functions, numbers or text.

TYPE: =IF(J2>=K2;”INCLUDE”;;”)

PRESS: ENTER

RESULT: Evaluates whether the random number is larger than 0.5. In case this is true, it will write INCLUDE into the cell.

Survey Data Management

	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	Family Name	Middle Name	Last Name	Random number	Threshold	Evaluation										
2	Pham	Thanh	Huyen	0.175595051	0.5	=IF(L2>=M2,										
3	Nguyen	Si	Tung	0.249938955	0.5											
4	Nguyen	Luong	Kien	0.273361127	0.5											
5	Nguyen	Thi Bich	Ngoc	0.3129552	0.5											
6	Nguyen	Khac	Binh	0.315649828	0.5											
7	Phan	Huu Tan	Duc	0.350868402	0.5											
8	Ngo	Thanh	Binh	0.567557784	0.5											
9	Tran	Van	Dien	0.572287601	0.5											
10	Ehlert	Richard	Christoph	0.589267072	0.5											
11	Bui	Van	Hung	0.621844057	0.5											
12	Le	Minh	Tan	0.714591787	0.5											
13	Luu	Duc	Tuyen	0.783466058	0.5											
14	Pham	Viet	Ha	0.810685466	0.5											
15	Le	Ninh	Khoi	0.942993324	0.5											

Screenshot 31: IF-function

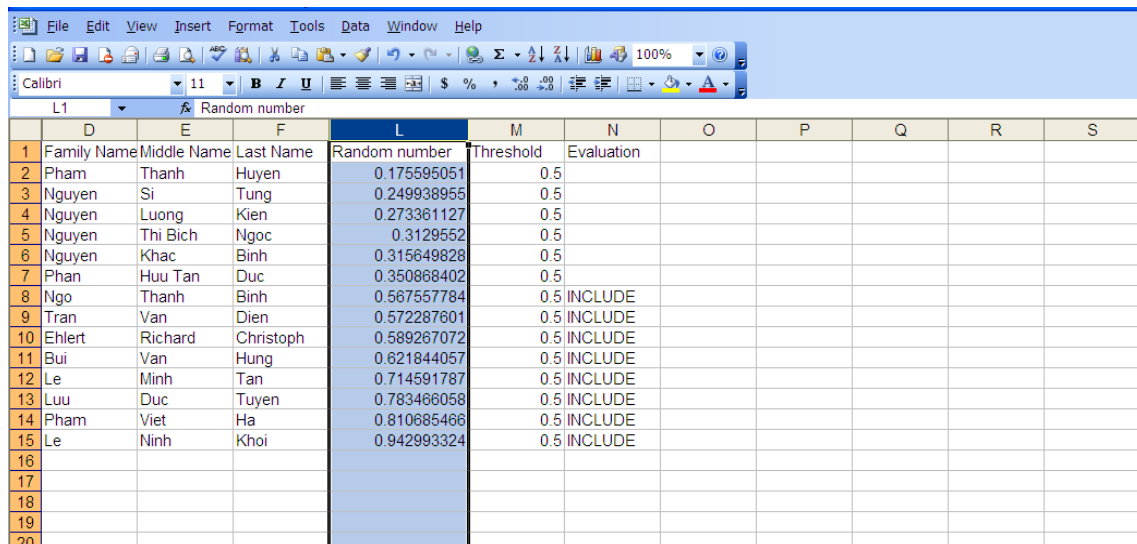
Please type "Evaluation" as a headline into the cell next to "Threshold". Next, type =IF(J2>=K2;"INCLUDE";"") into the first cell below this new headline and press ENTER (the quotation marks in the command are needed to indicate Excel that text is used instead of numbers or commands). Then copy the cells the same way you copied the threshold cells under the copying following these steps, you will see which persons to include into the survey.

	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	Family Name	Middle Name	Last Name	Random number	Threshold	Evaluation										
2	Pham	Thanh	Huyen	0.175595051	0.5											
3	Nguyen	Si	Tung	0.249938955	0.5											
4	Nguyen	Luong	Kien	0.273361127	0.5											
5	Nguyen	Thi Bich	Ngoc	0.3129552	0.5											
6	Nguyen	Khac	Binh	0.315649828	0.5											
7	Phan	Huu Tan	Duc	0.350868402	0.5											
8	Ngo	Thanh	Binh	0.567557784	0.5	INCLUDE										
9	Tran	Van	Dien	0.572287601	0.5	INCLUDE										
10	Ehlert	Richard	Christoph	0.589267072	0.5	INCLUDE										
11	Bui	Van	Hung	0.621844057	0.5	INCLUDE										
12	Le	Minh	Tan	0.714591787	0.5	INCLUDE										
13	Luu	Duc	Tuyen	0.783466058	0.5	INCLUDE										
14	Pham	Viet	Ha	0.810685466	0.5	INCLUDE										
15	Le	Ninh	Khoi	0.942993324	0.5	INCLUDE										

Screenshot 32: IF-function – result

After these steps, we have to make sure, that the random numbers do not change every time the cells are refreshed. To achieve this, you have to select the column with the random numbers by moving the mouse to the columns head letter and press the left mouse button.

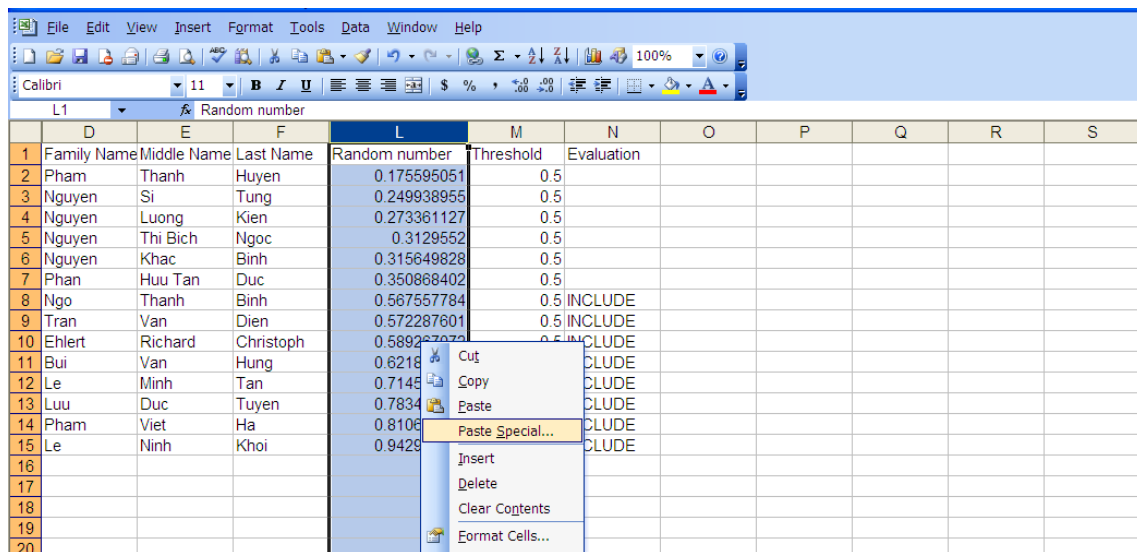
Survey Data Management



	D	E	F	L	M	N	O	P	Q	R	S
1	Family Name	Middle Name	Last Name	Random number	Threshold	Evaluation					
2	Pham	Thanh	Huyen	0.175595051	0.5						
3	Nguyen	Si	Tung	0.249938955	0.5						
4	Nguyen	Luong	Kien	0.273361127	0.5						
5	Nguyen	Thi Bich	Ngoc	0.3129552	0.5						
6	Nguyen	Khac	Binh	0.315649828	0.5						
7	Phan	Huu Tan	Duc	0.350868402	0.5						
8	Ngo	Thanh	Binh	0.567557784	0.5	INCLUDE					
9	Tran	Van	Dien	0.572287601	0.5	INCLUDE					
10	Ehlert	Richard	Christoph	0.589267072	0.5	INCLUDE					
11	Bui	Van	Hung	0.621844057	0.5	INCLUDE					
12	Le	Minh	Tan	0.714591787	0.5	INCLUDE					
13	Luu	Duc	Tuyen	0.783466058	0.5	INCLUDE					
14	Pham	Viet	Ha	0.810685466	0.5	INCLUDE					
15	Le	Ninh	Khoi	0.942993324	0.5	INCLUDE					
16											
17											
18											
19											
20											

Screenshot 33: Paste special – select

Now move the mouse over the selected cells and copy the cells. After copying, make sure that the cells are still selected. Now paste the cells at the same position, moving the mouse over the selected cells, doing a click with the right mouse button and selecting “paste special”.

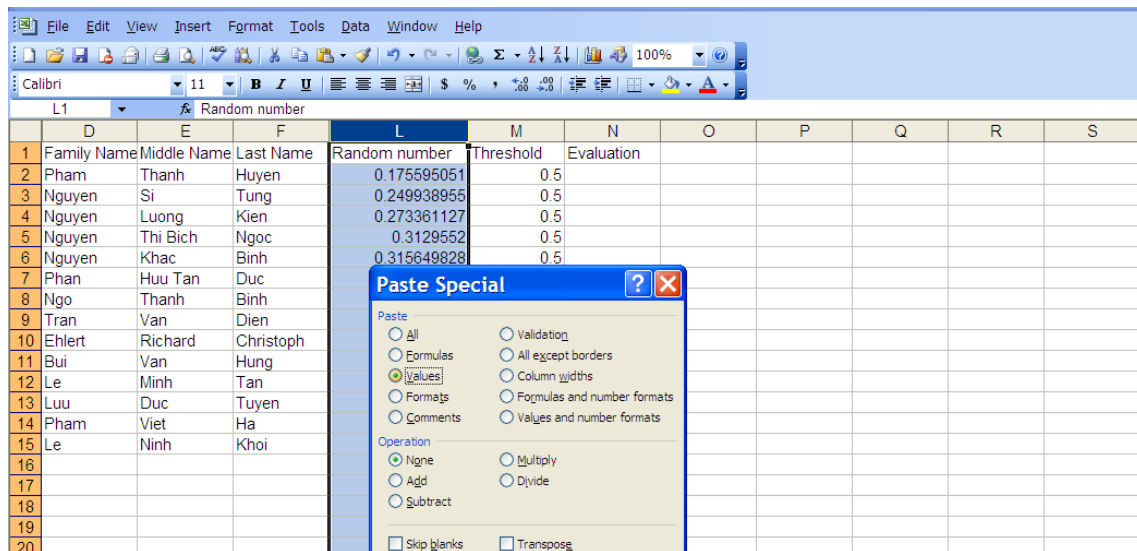


	D	E	F	L	M	N	O	P	Q	R	S
1	Family Name	Middle Name	Last Name	Random number	Threshold	Evaluation					
2	Pham	Thanh	Huyen	0.175595051	0.5						
3	Nguyen	Si	Tung	0.249938955	0.5						
4	Nguyen	Luong	Kien	0.273361127	0.5						
5	Nguyen	Thi Bich	Ngoc	0.3129552	0.5						
6	Nguyen	Khac	Binh	0.315649828	0.5						
7	Phan	Huu Tan	Duc	0.350868402	0.5						
8	Ngo	Thanh	Binh	0.567557784	0.5	INCLUDE					
9	Tran	Van	Dien	0.572287601	0.5	INCLUDE					
10	Ehlert	Richard	Christoph	0.589267072	0.5	INCLUDE					
11	Bui	Van	Hung	0.621844057	0.5	INCLUDE					
12	Le	Minh	Tan	0.714591787	0.5	INCLUDE					
13	Luu	Duc	Tuyen	0.783466058	0.5	INCLUDE					
14	Pham	Viet	Ha	0.810685466	0.5	INCLUDE					
15	Le	Ninh	Khoi	0.942993324	0.5	INCLUDE					
16											
17											
18											
19											
20											

Screenshot 34: Paste special

A list of options will appear where you select “Values” and click “OK”. Doing this will remove the function from the cells, leaving only their values. Therefore, the contents of the cells will not change anymore.

Survey Data Management



Screenshot 35: Paste special dialogue

Action Paste special, values only:

SELECT: Cell that you want to fill with the copied content

PRESS: Right mouse button

SELECT: Paste Special...

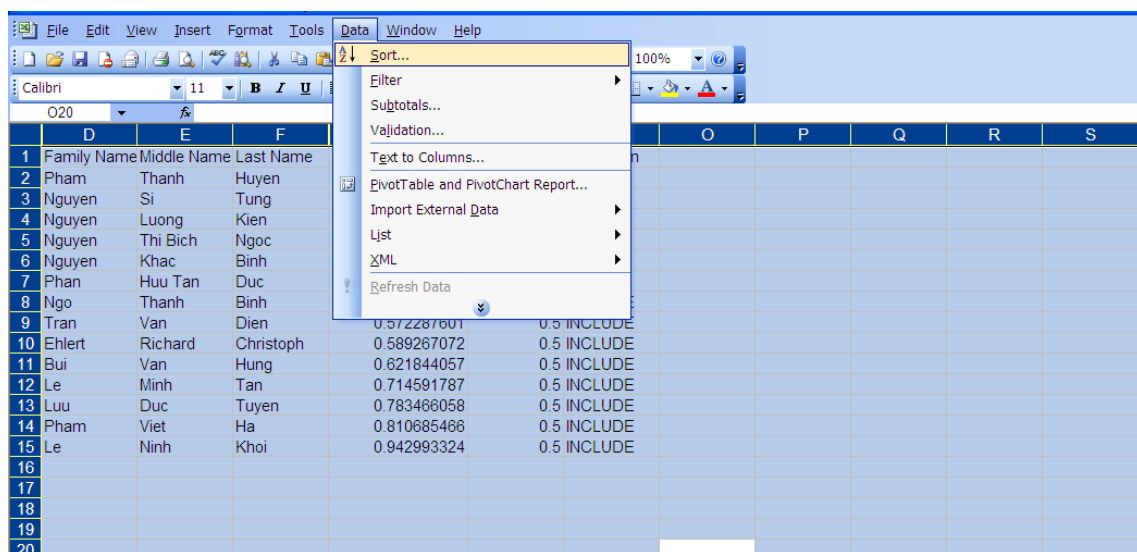
PRESS: Left mouse button

SELECT: Values

SELECT: OK

PRESS: Left mouse button

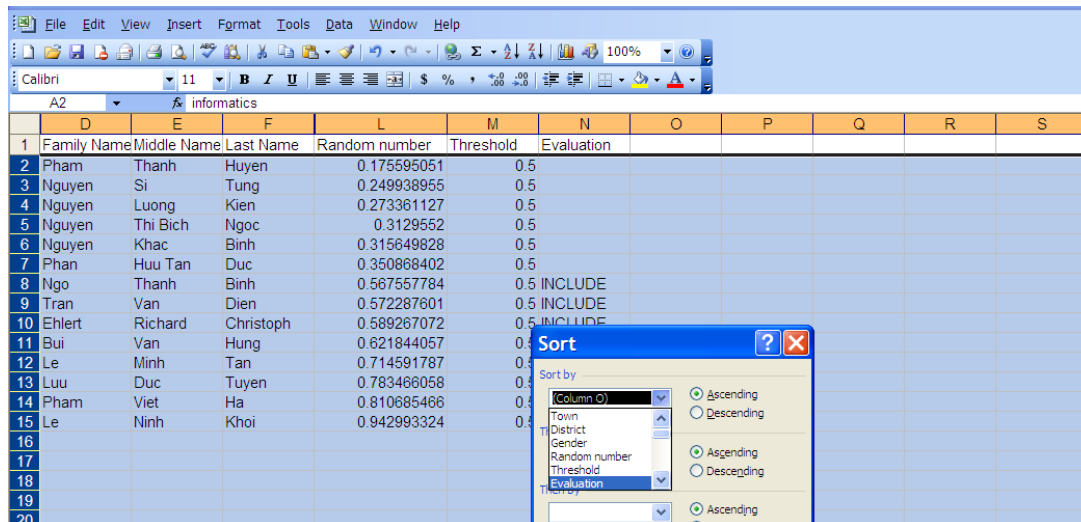
Now you can sort the cells according to the evaluation column. For this purpose, simultaneously press the keys "Ctrl" and "A" on your keyboard, this will select all cells in the spreadsheet. Now go to the Data Menu und select "Sort"



Screenshot 36: Sort

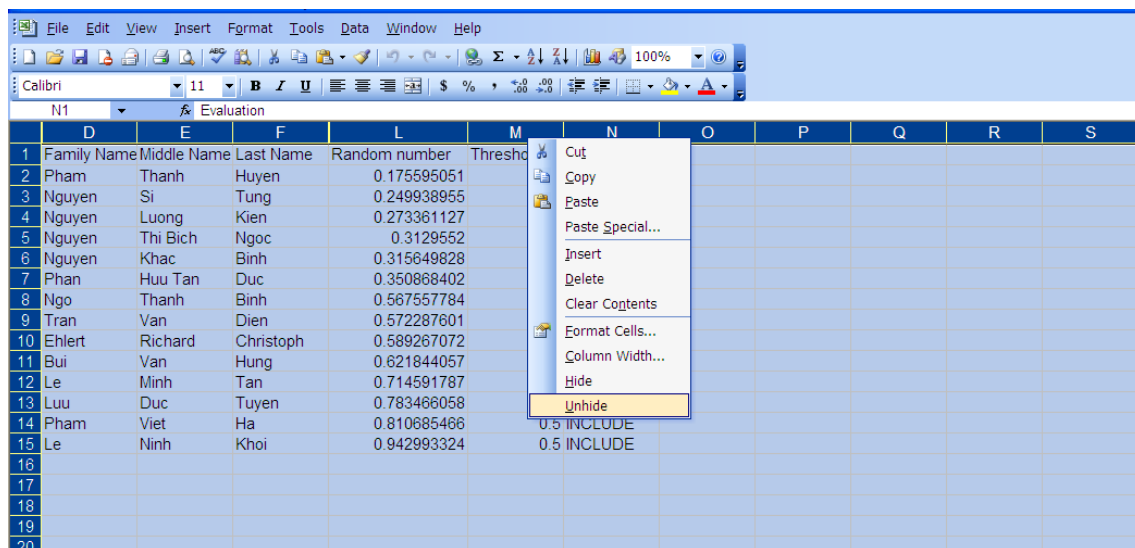
Make shure that you select „header row“, than select the column according to which the spreadsheet should be sorted. You can do this by selecting a column header in the drop down list of "sort by". After chosing the evaluation column, please press OK.

Survey Data Management



Screenshot 37: Sort dialog

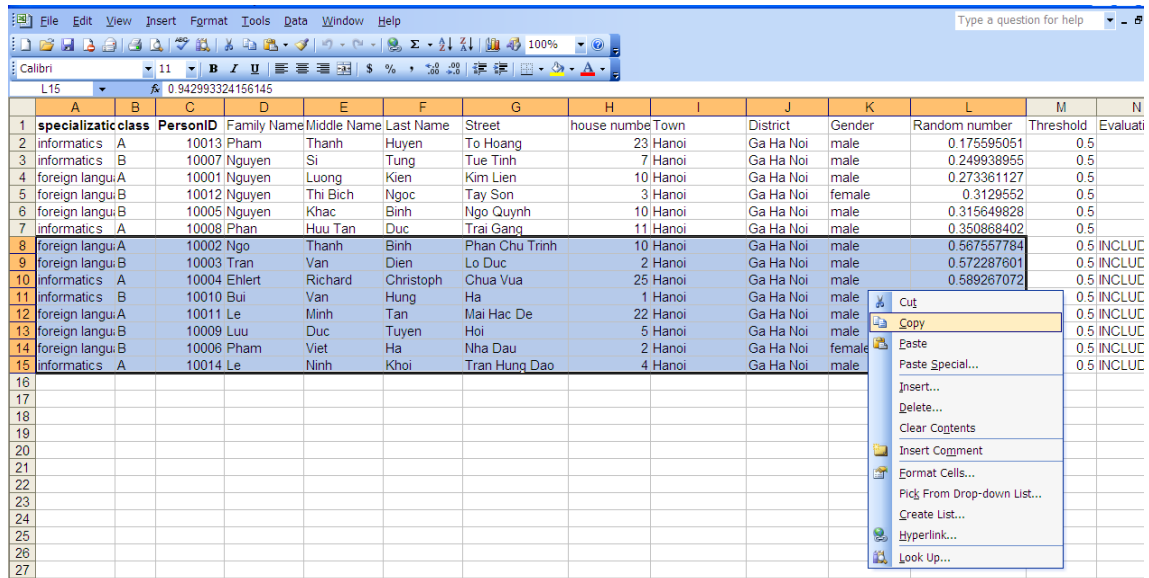
Now we have to copy all persons that have the “include” statement in the evaluation column and paste it into the corresponding columns of the Evaluation Excel sheet. Before doing this step, we have to unhide the columns that were previously hidden in order to make the example easier to understand. Please select the whole worksheet by pressing “Ctrl” and “A” simultaneously. Then move the mouse to the header of the spreadsheet that contains the column letters, press the right mouse button and select “Unhide” from the list.



Screenshot 38: Unhide

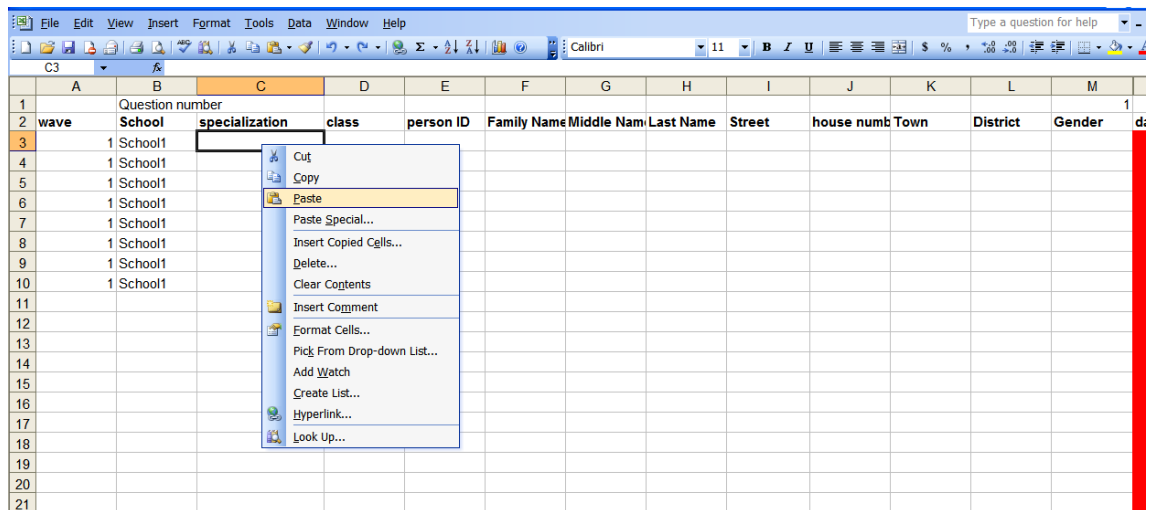
Now select all entries in the rows that contain the “INCLUDE” entry in the evaluation column and copy the persons that were chosen by the sampling procedure into the Evaluation file. In doing so, we first have to open the “Evaluation.xls”. Then you select all subjects that have to be included into the survey (those that have the INCLUDE statement) from the “Example 3” worksheet and copy this information.

Survey Data Management



Screenshot 39: Copy

After copying the information, please switch to the “Evaluation.xls” window that we just opened and paste the cells that we copied.



Screenshot 40: Paste

This is done by selecting the first empty cell in the specialization column and pasting the copied content there.

Survey Data Management

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
Question number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Wave	School	Specialization	Class	Person ID	Family Name	Middle Name	Last Name	Street	House Number	Town	District	Gender									
1	School1	foreign language	A	10002	Ngo	Thanh	Binh	Phan Chu Tri	10	Hanoi	Ga Ha Noi	male									
2	School1	foreign language	B	10003	Tran	Van	Dien	Lo Duc	2	Hanoi	Ga Ha Noi	male									
3	School1	informatics	A	10004	Ehlert	Richard	Christoph	Chua Vua	25	Hanoi	Ga Ha Noi	male									
4	School1	informatics	B	10010	Bui	Van	Hung	Ha	1	Hanoi	Ga Ha Noi	male									
5	School1	foreign language	A	10011	Le	Minh	Tan	Mai Hac De	22	Hanoi	Ga Ha Noi	male									
6	School1	foreign language	B	10009	Luu	Duc	Tuyen	Hoi	5	Hanoi	Ga Ha Noi	male									
7	School1	foreign language	B	10006	Pham	Viet	Ha	Nha Dau	2	Hanoi	Ga Ha Noi	female									
8	School1	informatics	A	10014	Le	Ninh	Khoi	Tran Hung Dz	4	Hanoi	Ga Ha Noi	male									

Screenshot 41: Result

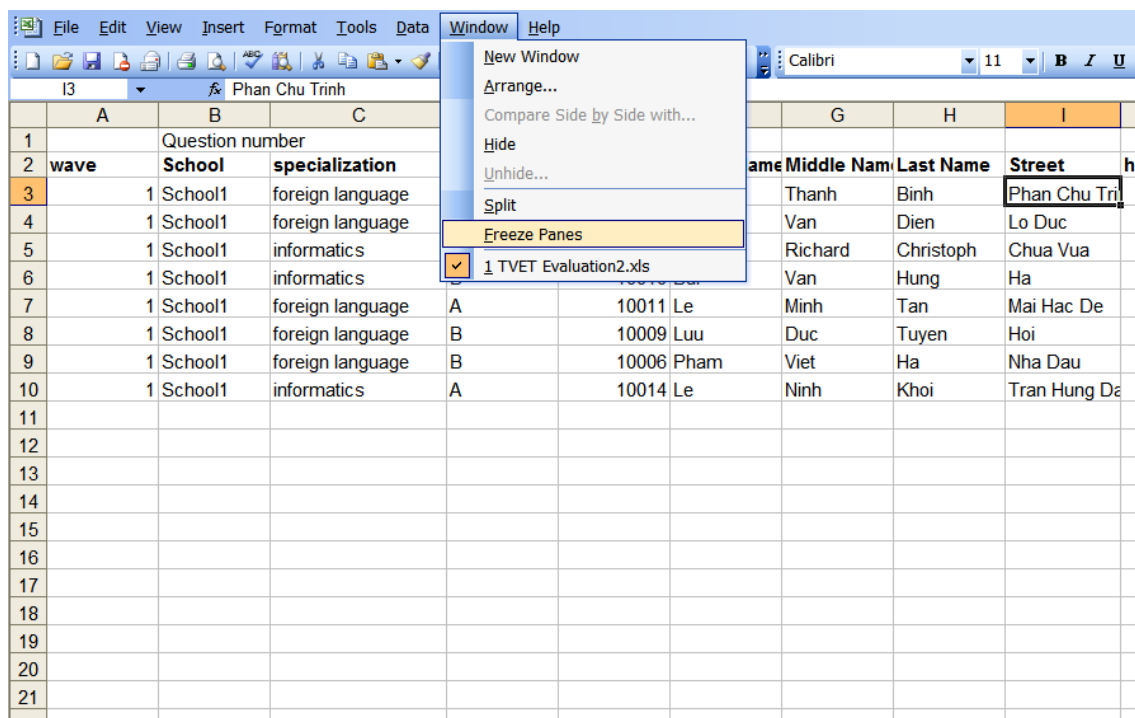
You should end up with a table similar to the one in screenshot 41. We have now prepared everything for the process of (baseline-) data-entry in the next chapter.

6 Coding questionnaires with Excel

Questionnaires coding is transformation of our questionnaire data into another format that computer software (in this case is Excel 2003) could understand

6.1 Structure of the data-entry file

In the following, we will go through the process of using the evaluation Excel file for questionnaire management purposes.. The file that you work with should already be filled with some information about your students, such as their names, birthdays, field of specialization and address. If you did not complete the steps in the previous chapter, you may want to use the Examples_backup.xls file to follow the steps in this chapter. The Excel file covers the baseline survey. Before entering the data, we have to make sure that the name of the person whose questionnaire data we fill in, is always present. To do so, please select the cell right to the first name in the first row below the headline. Now go to the “View” menu and select “Freeze Panes”, which will freeze the cells above and left from the selected cell.



Screenshot 42: Freeze panes

Survey Data Management

	A	B	C	D	E	F	G	H	I
1		Question number							
2	wave	School	specialization	class	person ID	Family Name	Middle Nam	Last Name	Street
3	1	School1	foreign language	A	10002	Ngo	Thanh	Binh	Phan Chu Tr
4	1	School1	foreign language	B	10003	Tran	Van	Dien	Lo Duc
5	1	School1	informatics	A	10004	Ehler	Richard	Christoph	Chua Vua
6	1	School1	informatics	B	10010	Bui	Van	Hung	Ha
7	1	School1	foreign language	A	10011	Le	Minh	Tan	Mai Hac De
8	1	School1	foreign language	B	10009	Luu	Duc	Tuyen	Hoi
9	1	School1	foreign language	B	10006	Pham	Viet	Ha	Nha Dau
10	1	School1	informatics	A	10014	Le	Ninh	Khoi	Tran Hung Da
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									

Screenshot 43: Freeze panes – result

Because there is still a lot of information that is not needed in the left part of the screen, please hide the columns A to D by selecting them, clicking with the right mouse button onto cells header and pressing “Hide” in the drop-down menu.

	A	B	C	D	E	F	G	H	I
1		Question number							
2	wave	School	specialization	class	person ID	Family Name	Middle Nam	Last Name	Street
3	1	School1	foreign language	A	10002	Ngo	Thanh	Binh	Phan Chu Tr
4	1	School1	foreign language	B	10003	Tran	Van	Dien	Lo Duc
5	1	School1	informatics	A	10004	Ehler	Richard	Christoph	Chua Vua
6	1	School1	informatics	B	10010	Bui	Van	Hung	Ha
7	1	School1	foreign language	A	10011	Le	Minh	Tan	Mai Hac De
8	1	School1	foreign language	B	10009	Luu	Duc	Tuyen	Hoi
9	1	School1	foreign language	B	10006	Pham	Viet	Ha	Nha Dau
10	1	School1	informatics	A	10014	Le	Ninh	Khoi	Tran Hung Da
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									

Screenshot 44: Hide

	E	F	G	H	I	J	K	L	M	O	P	
1										1	2	3
2	person ID	Family Name	Middle Nam	Last Name	Street	house numb	Town	District	Gender	date of birth	schooling degree	v
3	10002	Ngo	Thanh	Binh	Phan Chu Tri	10	Hanoi	Ga Ha Noi	male	17/5/1982		
4	10003	Tran	Van	Dien	Lo Duc	2	Hanoi	Ga Ha Noi	male			
5	10004	Ehlert	Richard	Christoph	Chua Vua	25	Hanoi	Ga Ha Noi	male			
6	10010	Bui	Van	Hung	Ha	1	Hanoi	Ga Ha Noi	male			
7	10011	Le	Minh	Tan	Mai Hac De	22	Hanoi	Ga Ha Noi	male			
8	10009	Luu	Duc	Tuyen	Hoi	5	Hanoi	Ga Ha Noi	male			
9	10006	Pham	Viet	Ha	Nha Dau	2	Hanoi	Ga Ha Noi	female			
10	10014	Le	Ninh	Khoi	Tran Hung Da	4	Hanoi	Ga Ha Noi	male			
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												

Screenshot 45: Hide – result

You can now conveniently enter data into the Excel sheet. In case of the multiple choice questions, you can easily select the answer using the provided drop-down lists.

6.2 Handling missing data

During the data-entry phase of your survey, you will often be confronted with missing data. Data can be missing for several reasons, in some cases these reasons are known – as it is most likely in telephone interviews – sometimes these reasons are unknown – as it often happens in postal mail surveys. Information about missing data is important, because it might help to understand shortcomings of the questionnaire in use. It also helps later to make guesses about the answers that would have been given and about the characteristics of those who did not respond to a specific question. Therefore, leaving non-answered questions open in the data-entry file is the least best option of handling missing data. Doing so, makes it indistinguishable whether there was an error in coding, whether the respondent didn't answer or whether the question simply didn't apply. A more appropriate way of handling these missing data would be to code the reasons explicitly. This can be done by using numerical values that are not occupied by any possible valid answer. Table 2 gives an example for coding missing data with numerical or non-numerical values. Whichever system you use, make sure that you document this system.

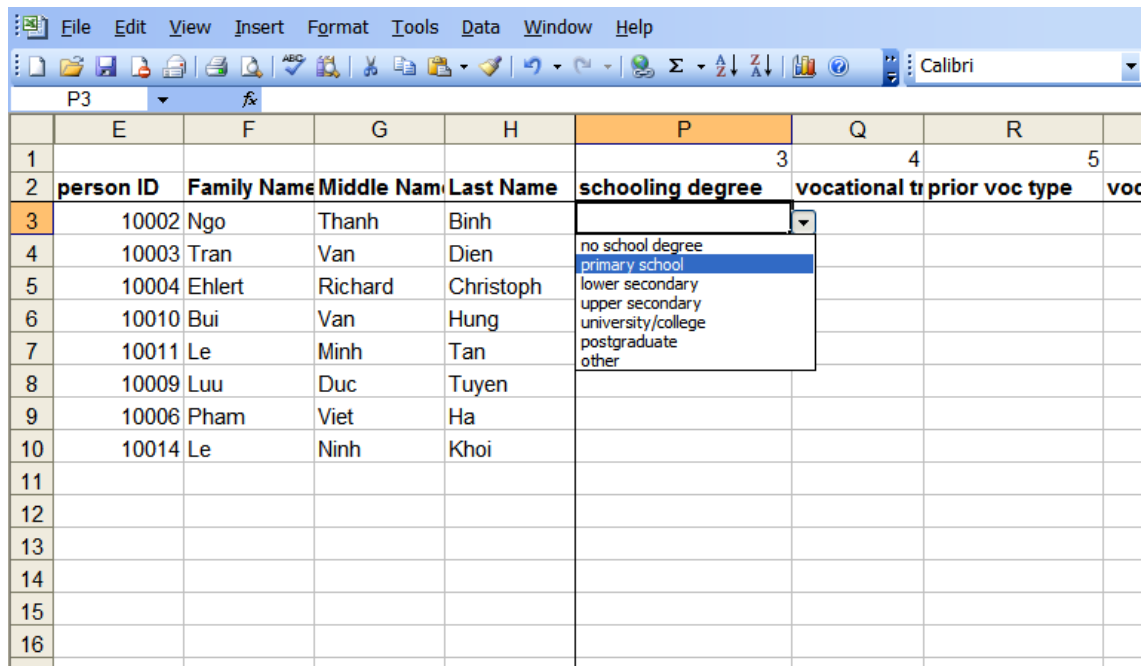
Table 2

Missing value	Numerical missing	Non-numerical missing
Question did not apply	-1	a
Answer refused	-2	b
Question not understood	-3	c
Other reason	-4	d

The next chapter gives some examples of modifying the multiple choice option sets of the data-entry form. You may use these examples to add missing value codes into the option-sets in the data-entry form.

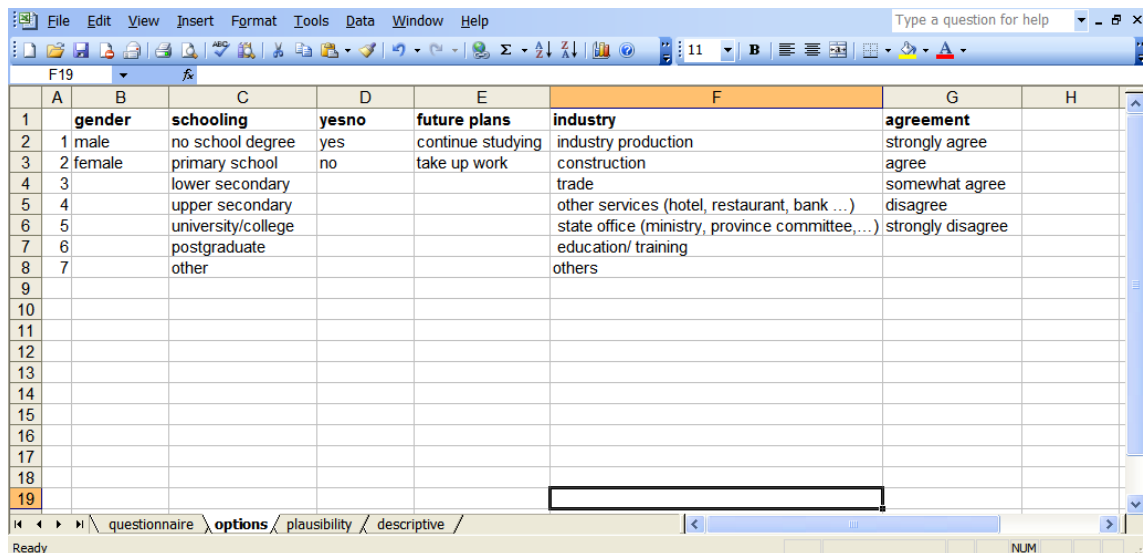
6.3 Defining new option-sets

The data-entry Excel file works with drop down lists that can be easily modified.



Screenshot 46: Option menus

To alter the content of these lists, you have to click onto the worksheet tab that has the label “options”, where you can change the content of the list by simply changing the contents of the cells.



Screenshot 47: Options tab

To add a new drop down menu, you have to add a new column to this worksheet. This column has to start with the name of the option set followed by the options that you want to have in your drop down menu. As an example, please type “newoption” into the cell right to the header of the “agreement” option set. Below, you may add the three options “option1”, “option2” and “option3”. Before creating the drop down menu, you have to create a so-called “name” for this option set. Please select the cells that contain the headline and options of the newly created option set.

Survey Data Management

	A	B	C	D	E	F	G	H
1		gender	schooling	yesno	future plans	industry	agreement	newoption
2	1	male	no school degree	yes	continue studying	industry production	strongly agree	option1
3	2	female	primary school	no	take up work	construction	agree	option2
4	3		lower secondary			trade	somewhat agree	option3
5	4		upper secondary			other services (hotel, restaurant, bank ...)	disagree	
6	5		university/college			state office (ministry, province committee,...)	strongly disagree	
7	6		postgraduate			education/ training		
8	7		other			others		
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								

Screenshot 48: Inserting new options

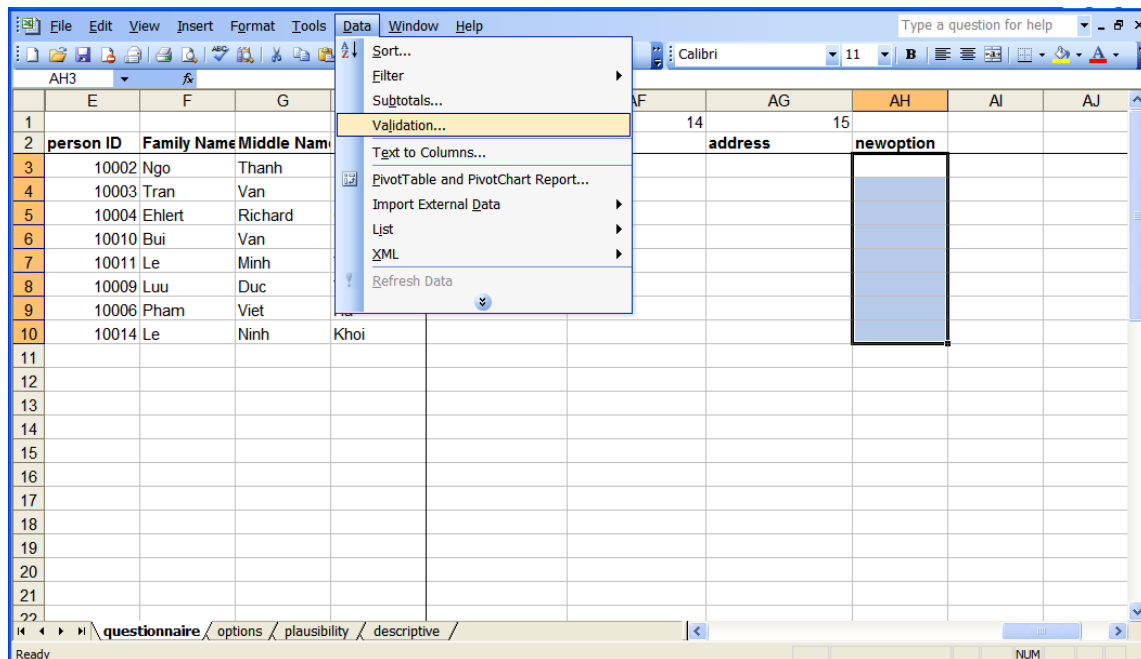
Then go to the “Insert” menu in the option bar of Excel, select “Name” and then click on “Define...”. In the following window press “OK”. This will create an option-set which has the name “newoption” in our example.

	E	G	H	I	J	K
1	future plans	agreement	newoption			
2	continue studying	strongly agree	option1			
3	take up work	agree	option2			
4		somewhat agree	option3			
5		disagree				
6		strongly disagree				
7						
8	others					
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						

Screenshot 49: Defining names

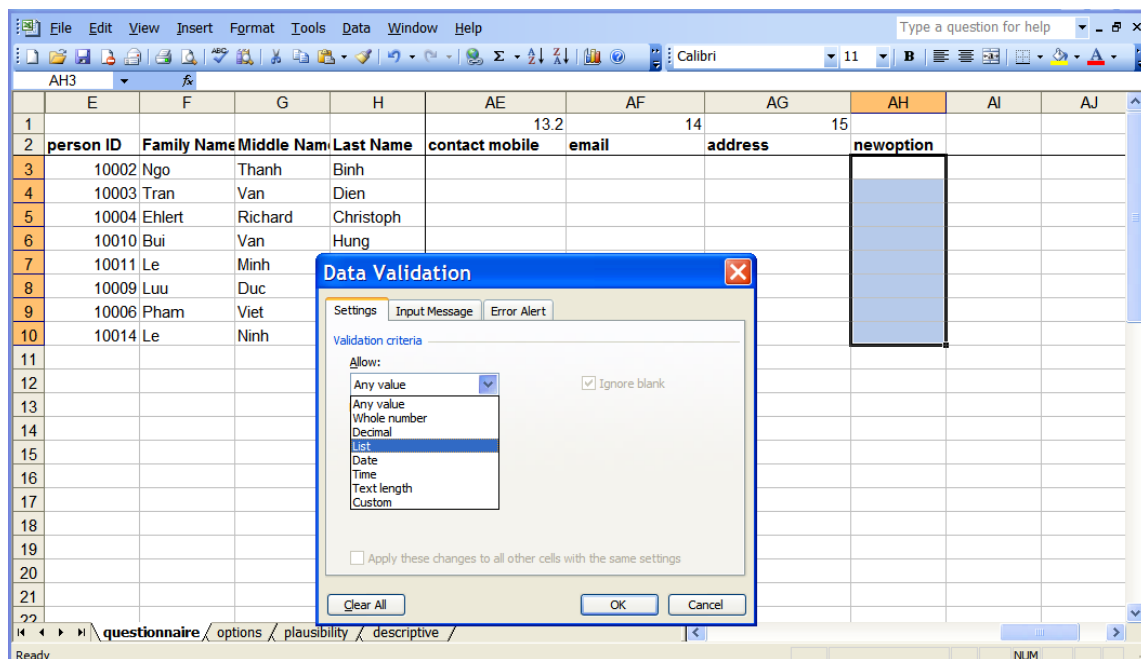
To use this option-set in the data-entry worksheet, you will have to select the cells that shall contain the drop down lists later. Then go to the “Data” menu in the option bar of Excel and select the point “Validation...”

Survey Data Management



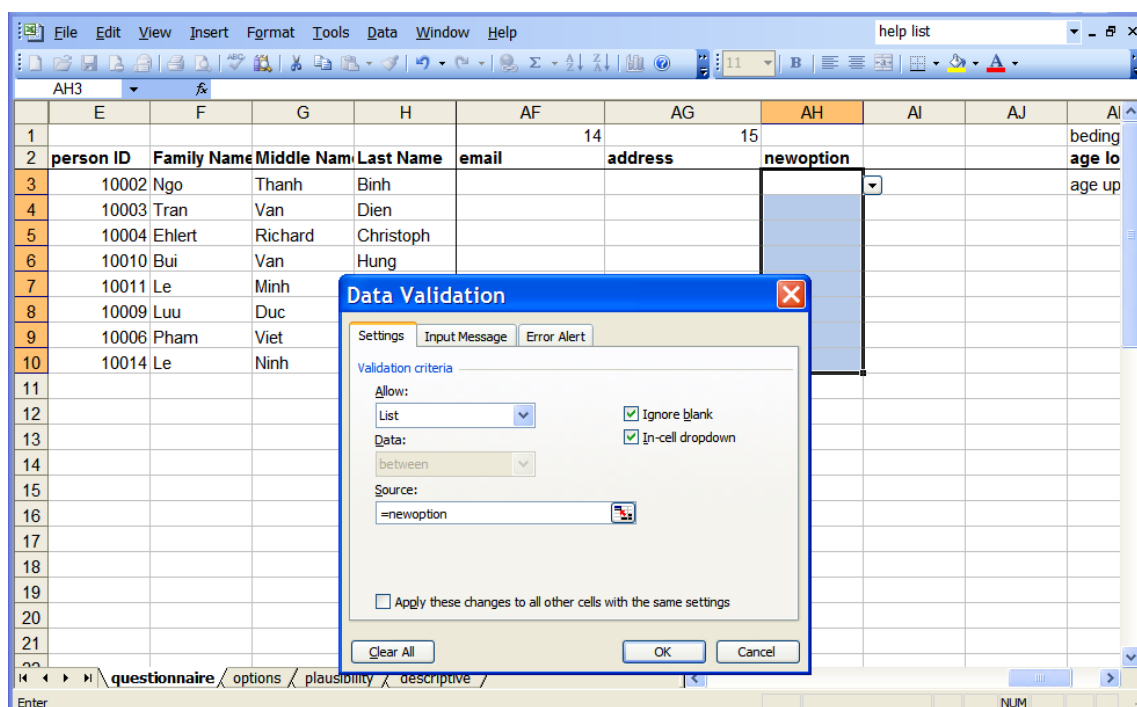
Screenshot 50: Data Validation

Now select “check data” from the data menu and “List” in the “Validation criteria” field. To finish, please type the name of the option set starting with an equal sign into the “Source” field.



Screenshot 51: Data validation dialogue

Survey Data Management



Screenshot 52: Data validation dialogue – select

In case of our example, you have to type “=newoption”. After finishing by clicking “Enter”, you will find the new drop-down lists in the selected cells.

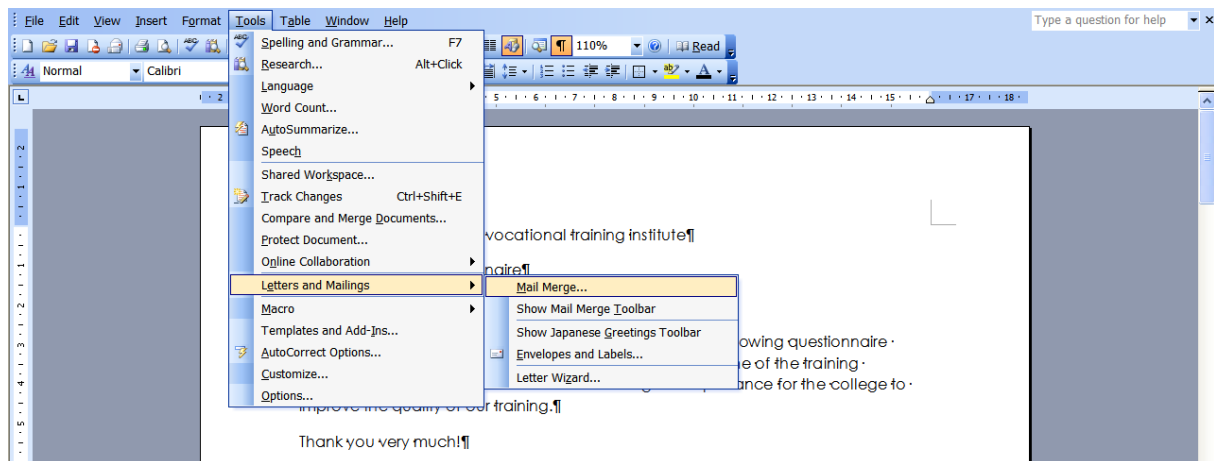
7 Automated postal mail survey

This section will help you to produce multiple questionnaire letters in the same standard format to send to graduates via courier. Mail merge function enables you to do that in a fast and accurate manner.

7.1 Selecting a data-base

Word can access data from Excel via the “Mail Merge” function. This function allows defining fields that refer to information in columns of Excel. Examples are personalized letters, where only name and address changes, while the rest of the text remains unaltered.

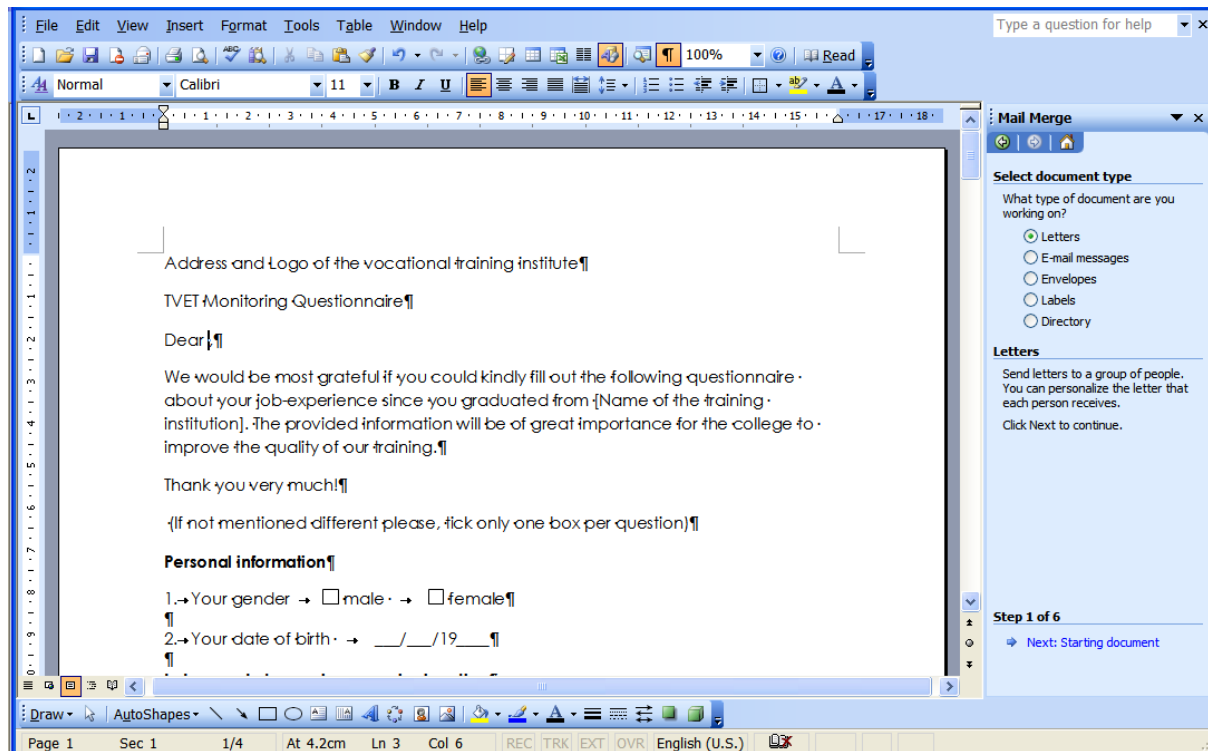
To use this function, please open a Word document that you might want to personalize and start the “Mail Merge” function. In this example, we use an excerpt of the English version of the tracer questionnaire, which is also found in the file “Handbook_Letter.doc” distributed with this handbook. After having the Word document opened, please start the Mail Merge function by clicking on “Tools” in the option bar of Word, selecting “Letters and Mailings” and clicking on “Mail Merge...”. This will open an assistant that is located at the right side of the Word window.



Screenshot 53: Mail merge

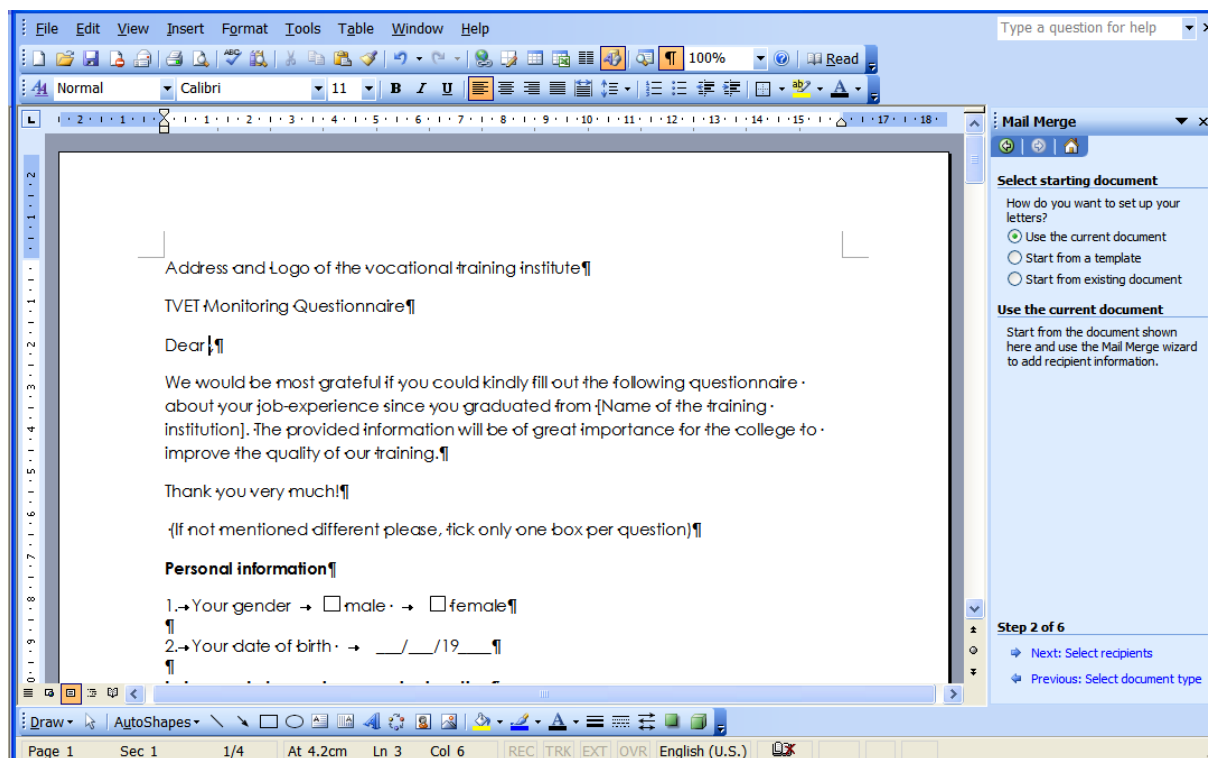
In this assistant, you can select the kind of documents you want to personalize. As this example deals with creating automated letters for survey purpose, you will have to select the first option, “Letters”, in the menu.

Survey Data Management



Screenshot 54: Mail merge – step 1

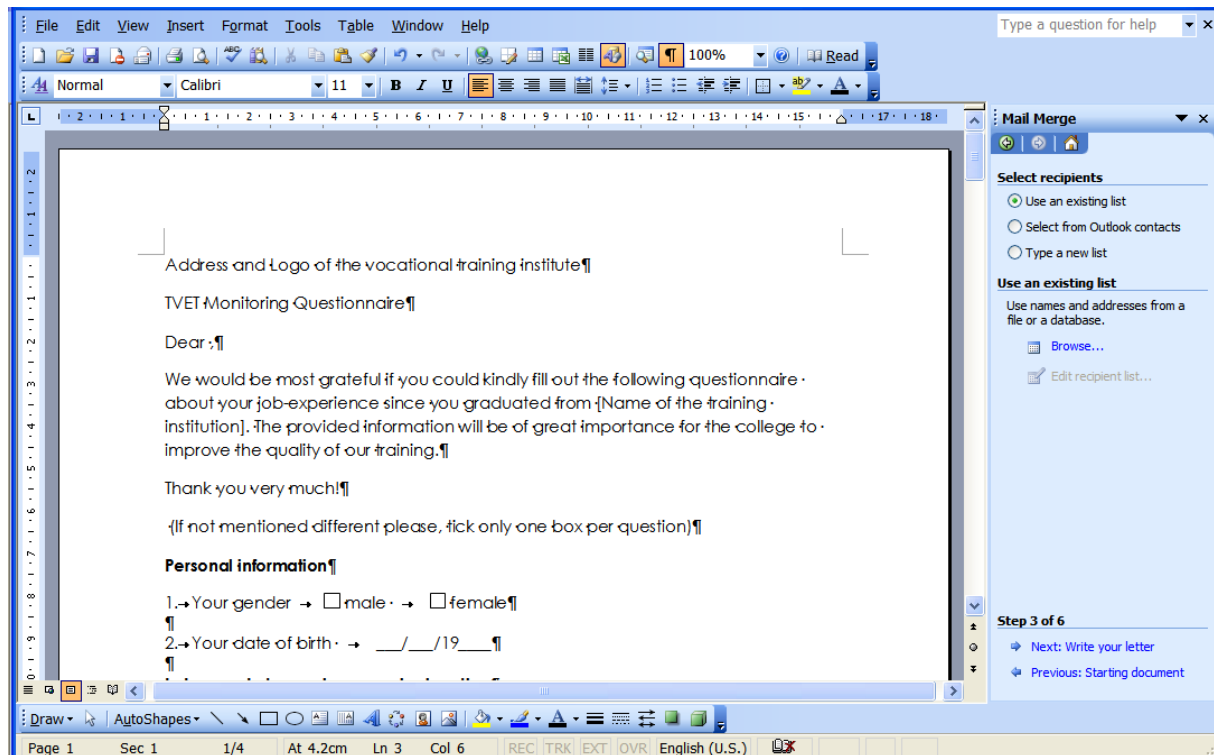
After doing so, please click “Next: Starting document” at the bottom of the assistant window. In the following step, you are asked to select a document which you want to personalize, as we already opened this document, we stick to the pre-selected option “Use current document”. Following, please click “Next: Select recipients” to jump to the next step.



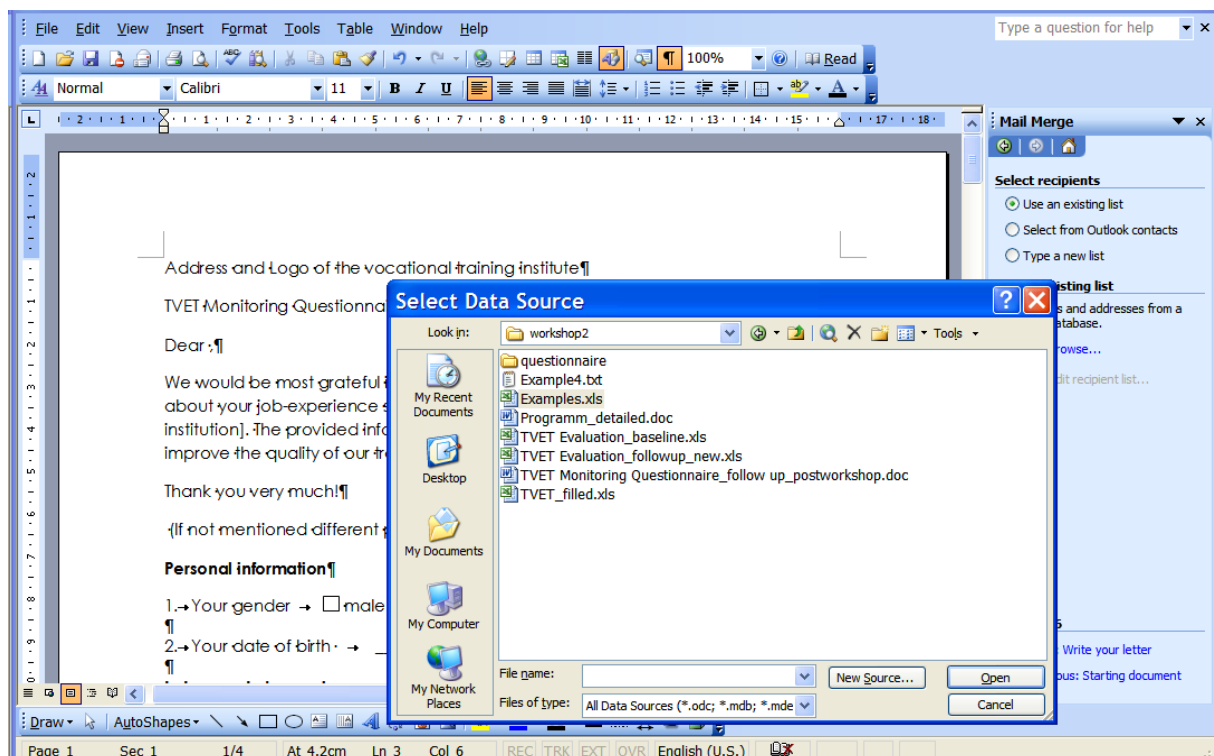
Screenshot 55: Mail merge – step 2

Survey Data Management

In this step, we can open the Excel document which contains the information about the students, such as their names and addresses. You can open this document by clicking on “Browse...” in the Mail merge assistant window. In this example we open the Examples_Handbook.xls that you received with this handbook.



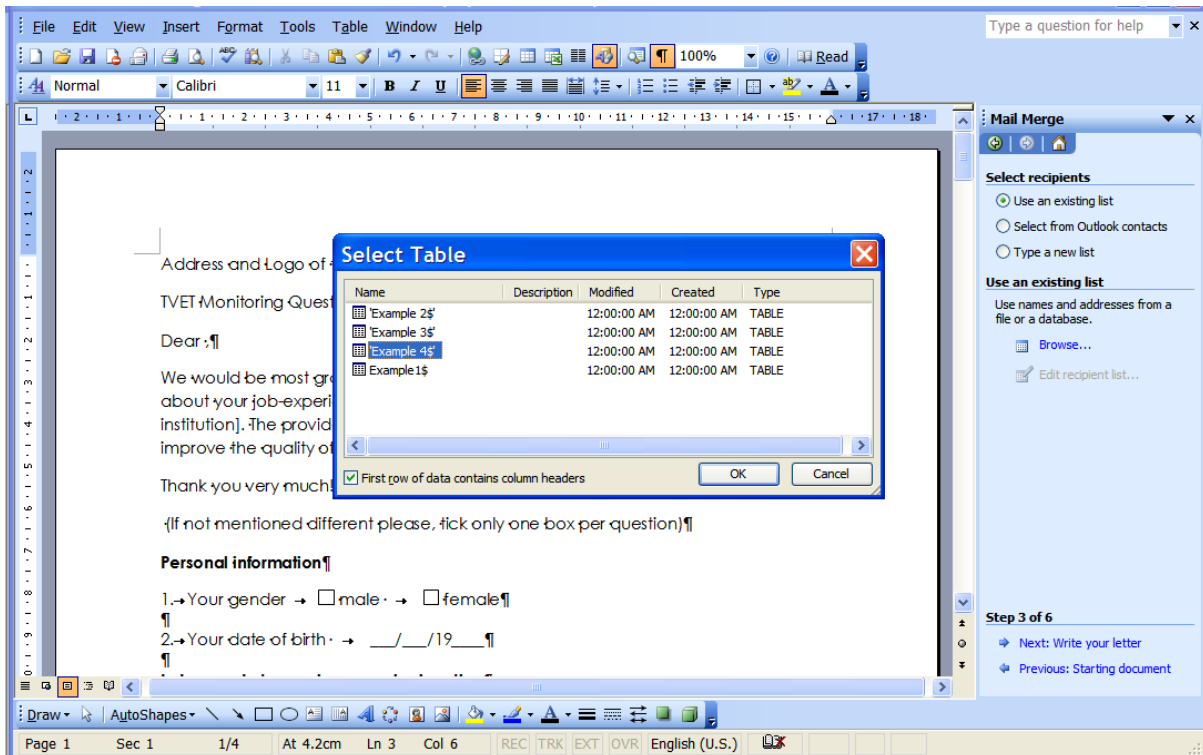
Screenshot 56: Mail merge – step 3



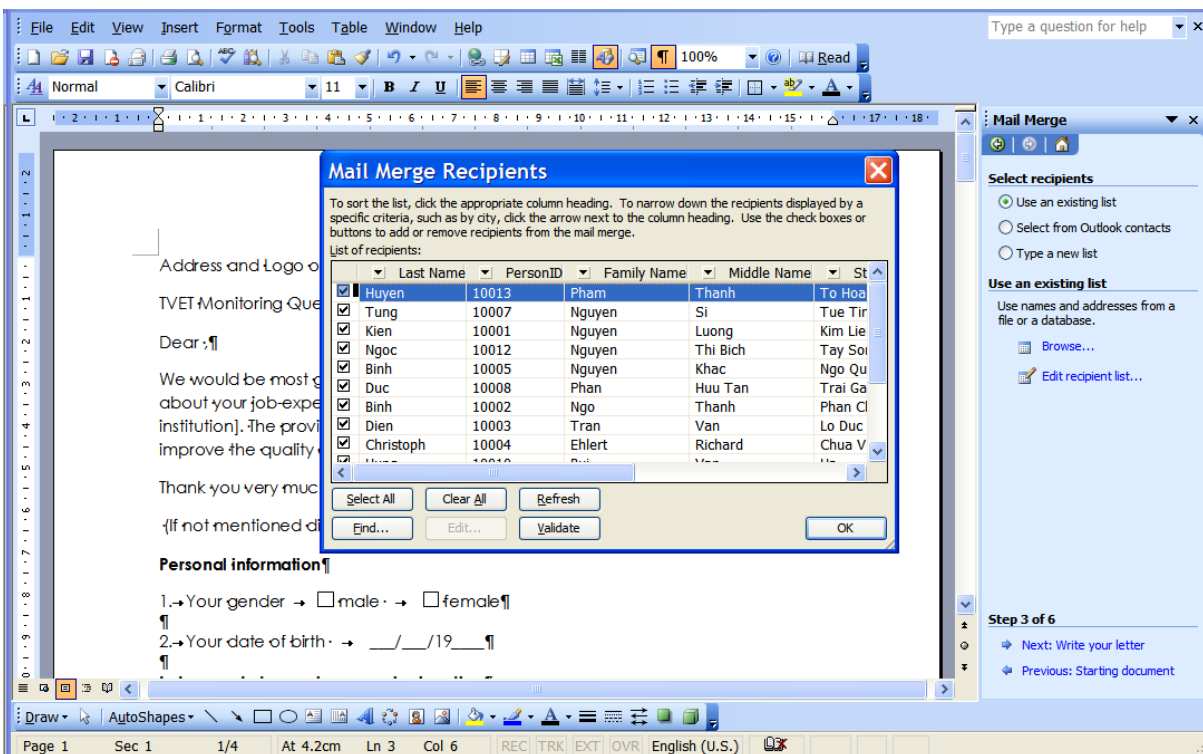
Screenshot 57: Mail merge – Selecting data source

Survey Data Management

When opening the Examples_Handbook.xls document, the Mail Merge assistant will ask you, which of the working sheets contained in the Excel file you would like to use. As the names and addresses are contained in the “Example 4” table, we select this table and click on “OK”.



Screenshot 58: Mail merge – Selecting table

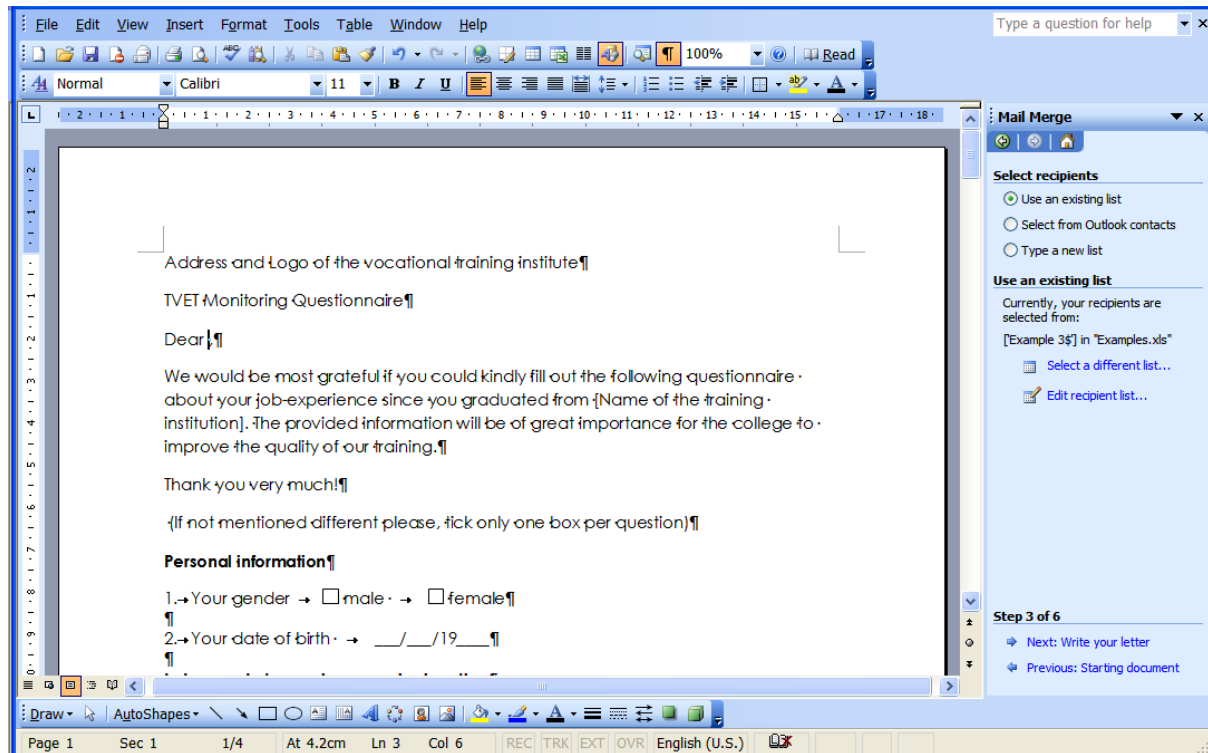


Screenshot 59: Mail merge – selecting recipients

To provide you with information about the data contained in the spreadsheet we just selected, the Mail Merge assistant automatically opens an preview window with the data

Survey Data Management

contained in the spreadsheet. In this window you can de-select the persons to which you do not want to send the personalized letters by un-checking the corresponding boxes on the left. Please proceed with “OK”, then click “Next: Write your letter”.

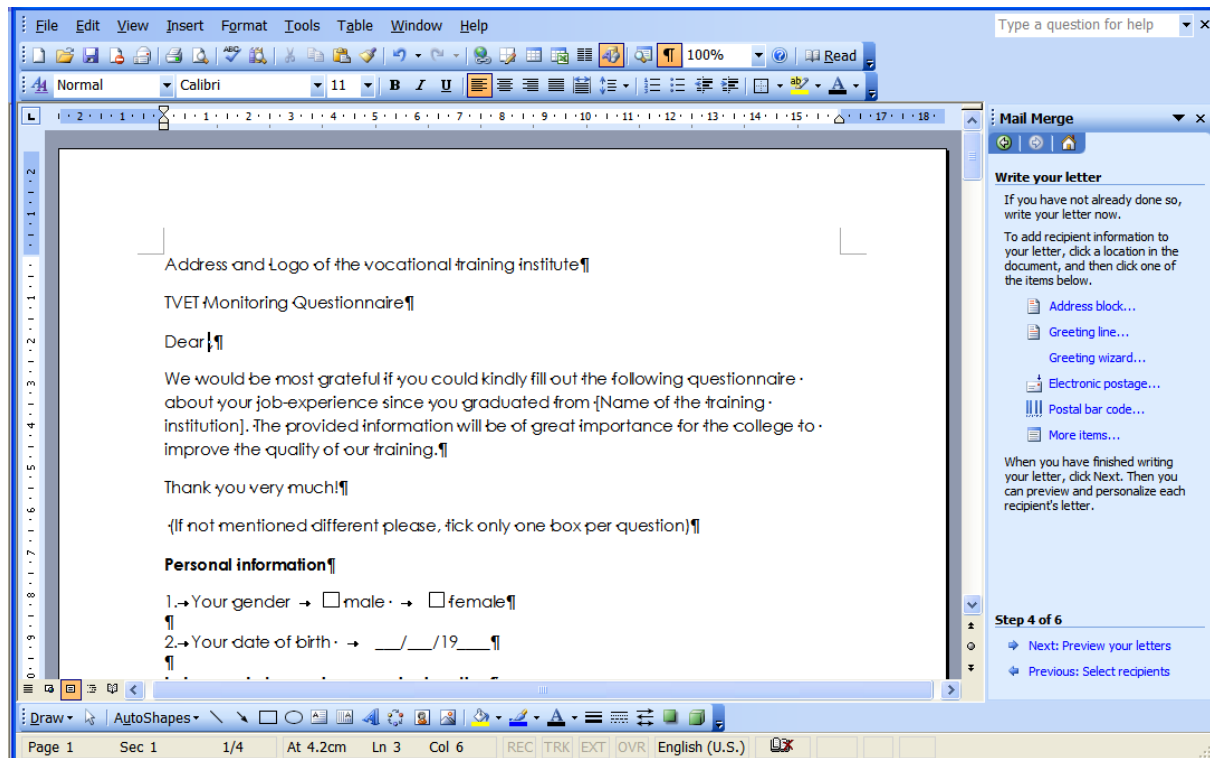


Screenshot 60: Mail merge – step 3

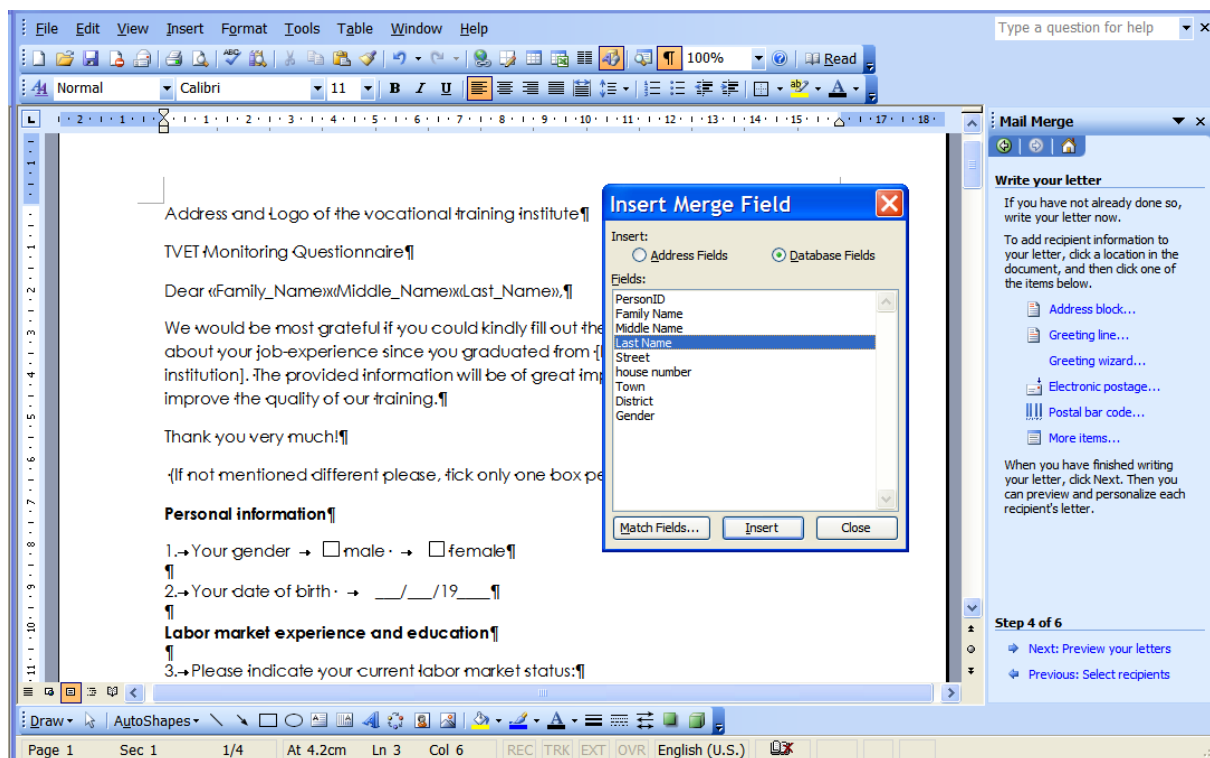
7.2 Using data-fields

To place fields that refer to the information in the Excel sheet, the first step is to place the cursor at the position you want to include certain information. In this example, we want to include a field that contains the name of the persons listed in the Excel document into the address. Therefore we place the cursor after “Dear” and click on “More items” in the Mail Merge assistant window. In the box that opens, you can select the items that you want to insert at this position. Please select “Family Name”, click “Insert”, then insert “Middle Name” and “Last Name” the same way by selecting them and clicking “Insert”.

Survey Data Management

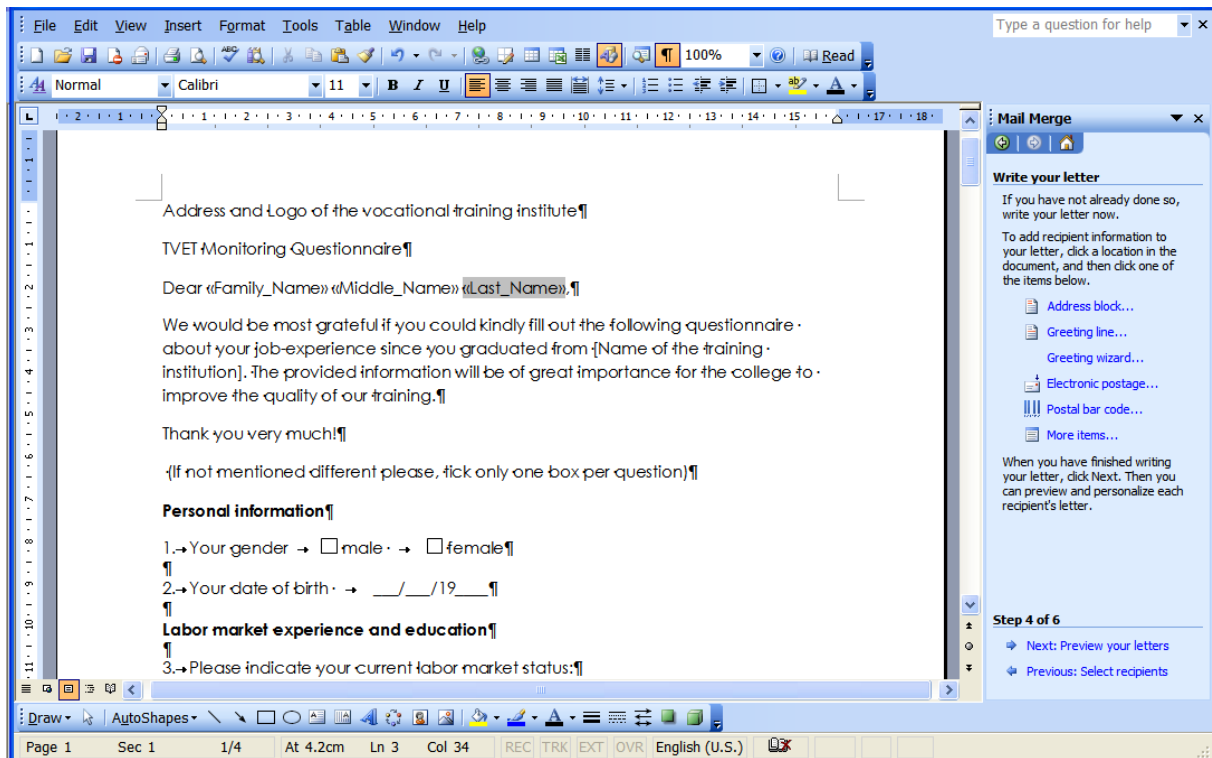


Screenshot 61: Mail merge – step 4



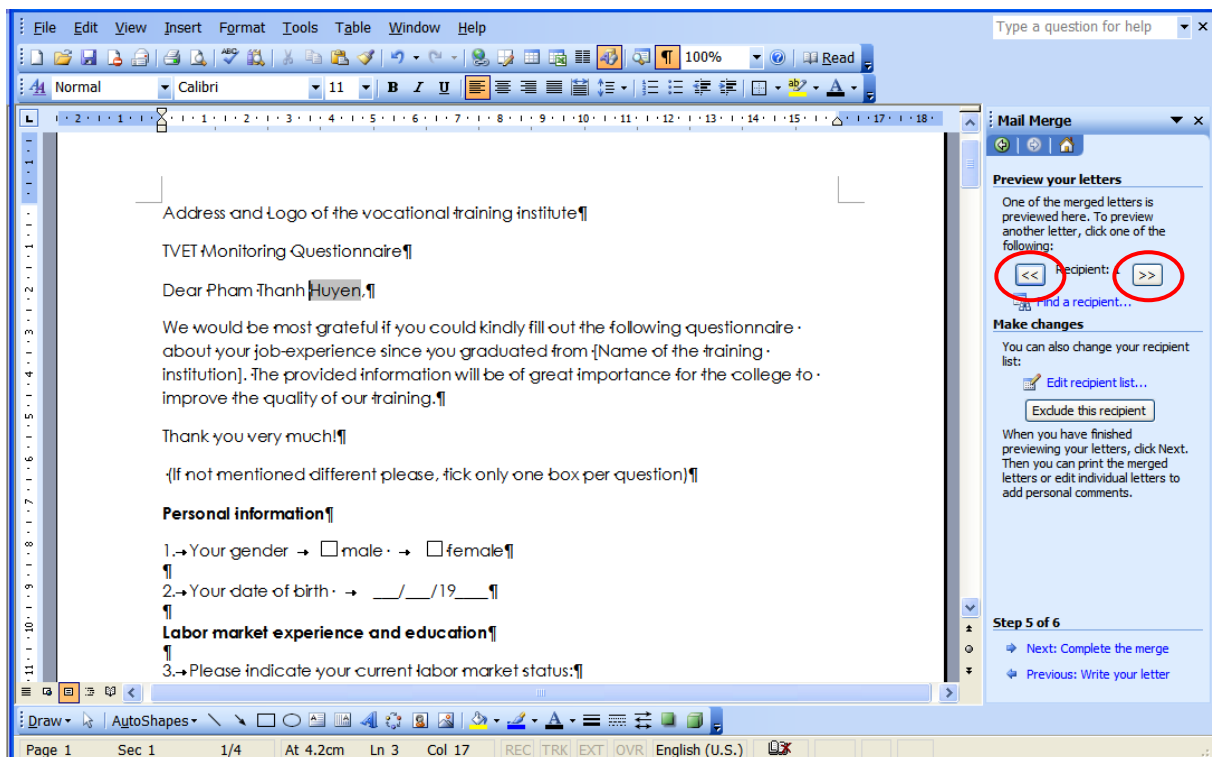
Screenshot 62: Mail merge – insert mail merge field dialogue

After closing the “Insert Merge Field” window, you will see these three fields. By clicking on them, they will turn grey, indicating that they are Mail Merge fields. You can select these fields and insert and copy them to any position in the text – the same way as ordinary text.



Screenshot 63: Mail merge – insert mail merge field

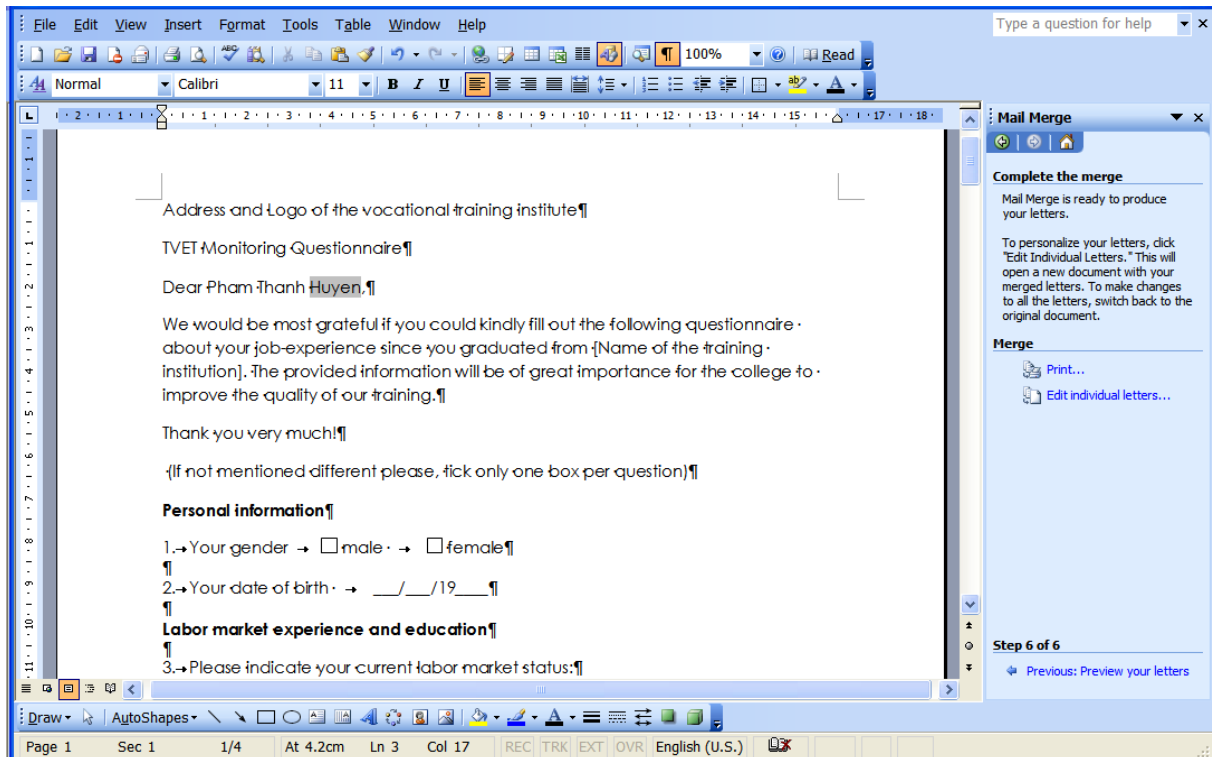
In the next step, please click “Next: Preview your letters”, which allows you to preview the individualized letters that will be created. You can go through these mails by simply clicking onto the double arrows in the “Preview your letters” section (highlighted with red circles in screenshot 64).



Screenshot 64: Mail merge – preview

Survey Data Management

In the last step, you can complete the mail merge by clicking on “Next: Complete the merge” and then on “Print”, which will print the letters for all persons that are in the Excel file you selected.



Screenshot 65: Mail merge – print

8 Merging data from different data-sets (baseline and follow-up)

8.1 Importance of a consistent person identifier

For being able to merge different data-sets, it is necessary to have a unique person identifier in each of the data-sets that are going to be merged. This identifier allows Excel to find information connected to this identifier in a second data-set. As an example, let's have a look at table 1 and table 2 in the screenshot 66 below (see Example 5 in the Examples_Handbook.xls). In table 1, you find a list of names and a person identifier, in table 2 you find the birthday and the person identifier. Given that we wish to combine this information, we have to be sure that the variable "person_id" is consistent across both tables, that is, identification number 1 belongs to the same person in table 2 as in table 1.

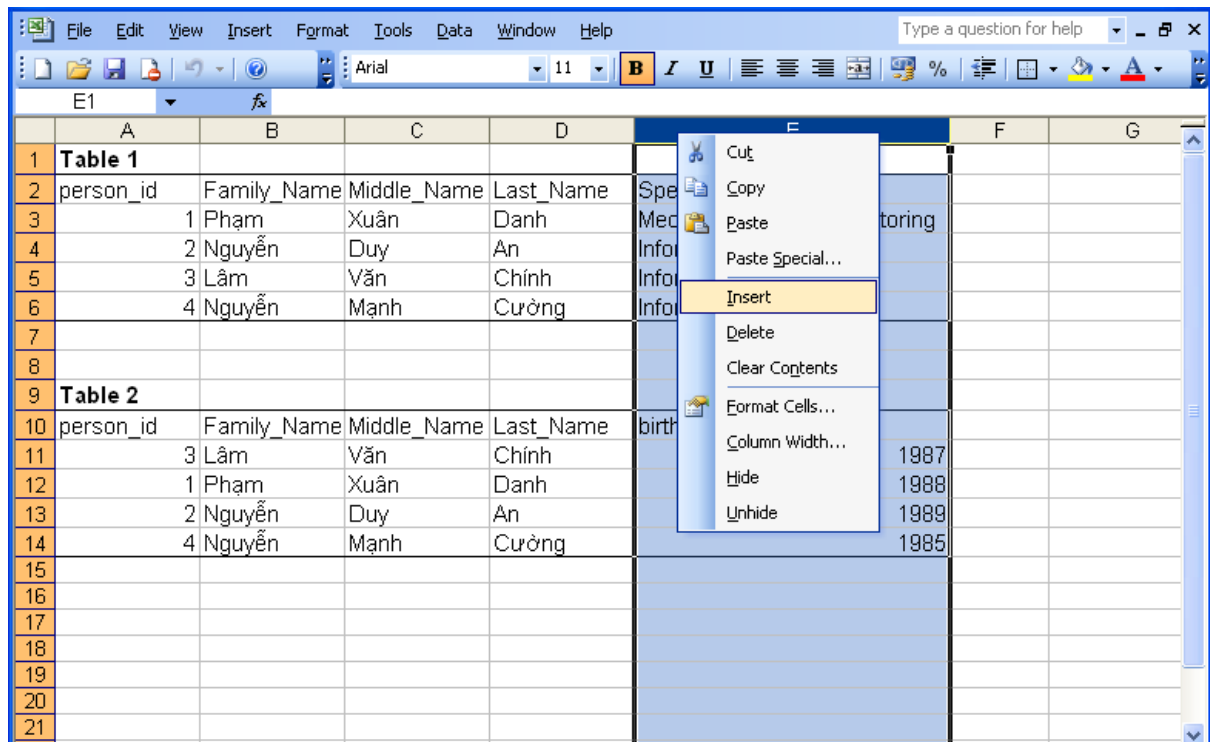
Table 1				
person_id	Family_Name	Middle_Name	Last_Name	Specialisation
1	Phạm	Xuân	Danh	Mechanics and Manufacturing
2	Nguyễn	Duy	An	Informatic Technology
3	Lâm	Văn	Chính	Informatic Technology
4	Nguyễn	Mạnh	Cường	Informatic Technology

Table 2				
person_id	Family_Name	Middle_Name	Last_Name	birthday
3	Lâm	Văn	Chính	1987
1	Phạm	Xuân	Danh	1988
2	Nguyễn	Duy	An	1989
4	Nguyễn	Mạnh	Cường	1985

Screenshot 66: Tables to merge

8.2 Creating a new identifier

In cases where such an identifier does not exist, we can create it with information common in both data-sets. In our case, the name is common in both data-sets. In order to create a new identifier from the name, the information of all three columns containing family, middle and last name has to be combined into one cell. For this purpose, a new column has to be created first.



Screenshot 67: Inserting new column

Please add the new column right of the column that holds the last name (to add a new column, follow the instructions in the screenshot above or in chapter 4.2). This new column is then used to generate the new identifier by writing the full name of each person into one cell. The formula for writing the contents of all three cells into one cell uses the “&” to connect the cells. To start, type “=” into cell E3 to start the formula, and then select the first cell that contains information to be written into the new cell. You can do this by selecting the cell with the mouse or by typing the position of the cell (B3) directly into the formula. As we want to have information of three cells in one, we now type “&”, which allows us to add the second cell (C3). This way, you can select all three cells within one formula. After doing so, you should end up with “=B3&C3&D3”. Please press “Enter” so see the result of the formula. The result should look like the one shown in the screenshot 69 – the whole name is now written in one cell. We can now copy this formula into the other cells.

Survey Data Management

SUM X ✓ ✖ =B3&C3&D3

	A	B	C	D	E	F	G
1	Table 1						
2	person_id	Family_Name	Middle_Name	Last_Name		Specialisation	
3	1	Phạm	Xuân	Danh	=B3&C3&D3	Mechanics and Manufacturing	
4	2	Nguyễn	Duy	An		Informatic Technology	
5	3	Lâm	Văn	Chính		Informatic Technology	
6	4	Nguyễn	Mạnh	Cường		Informatic Technology	
7							
8							
9	Table 2						
10	person_id	Family_Name	Middle_Name	Last_Name		birthday	
11	3	Lâm	Văn	Chính		1987	
12	1	Phạm	Xuân	Danh		1988	
13	2	Nguyễn	Duy	An		1989	
14	4	Nguyễn	Mạnh	Cường		1985	
15							
16							
17							
18							
19							
20							
21							

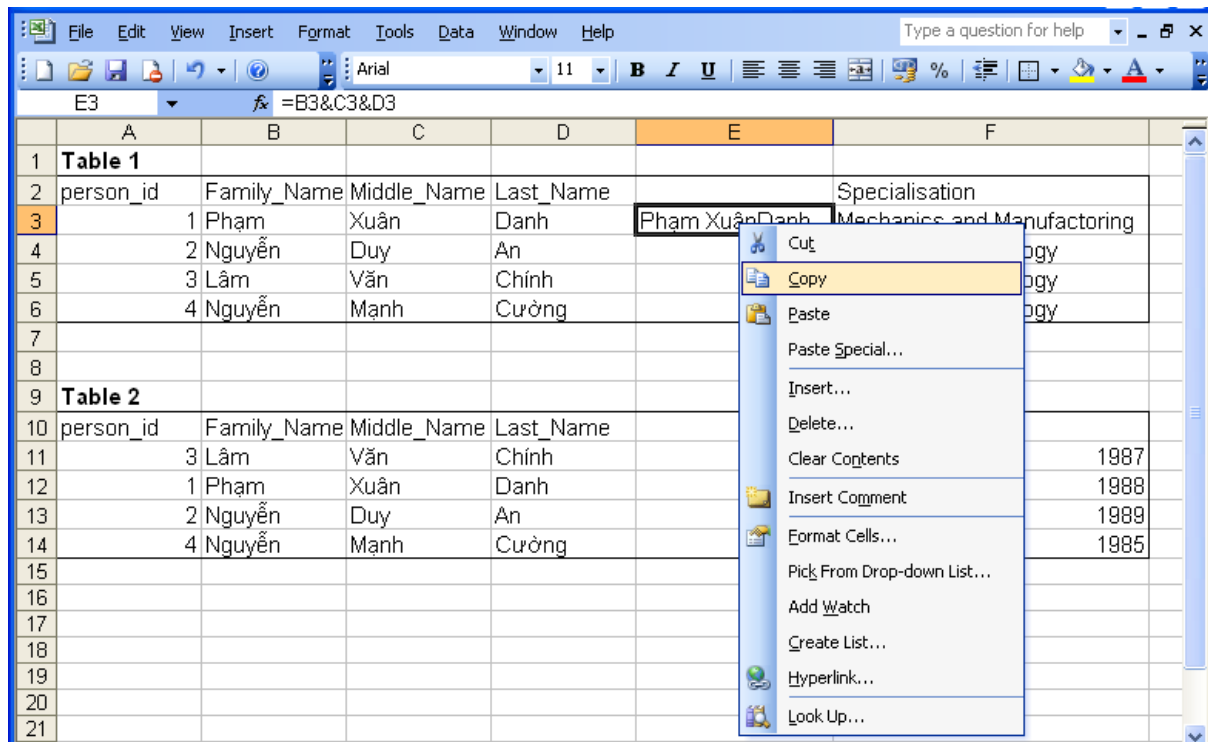
Screenshot 68: Formula to connect cells

E3 X ✓ ✖ =B3&C3&D3

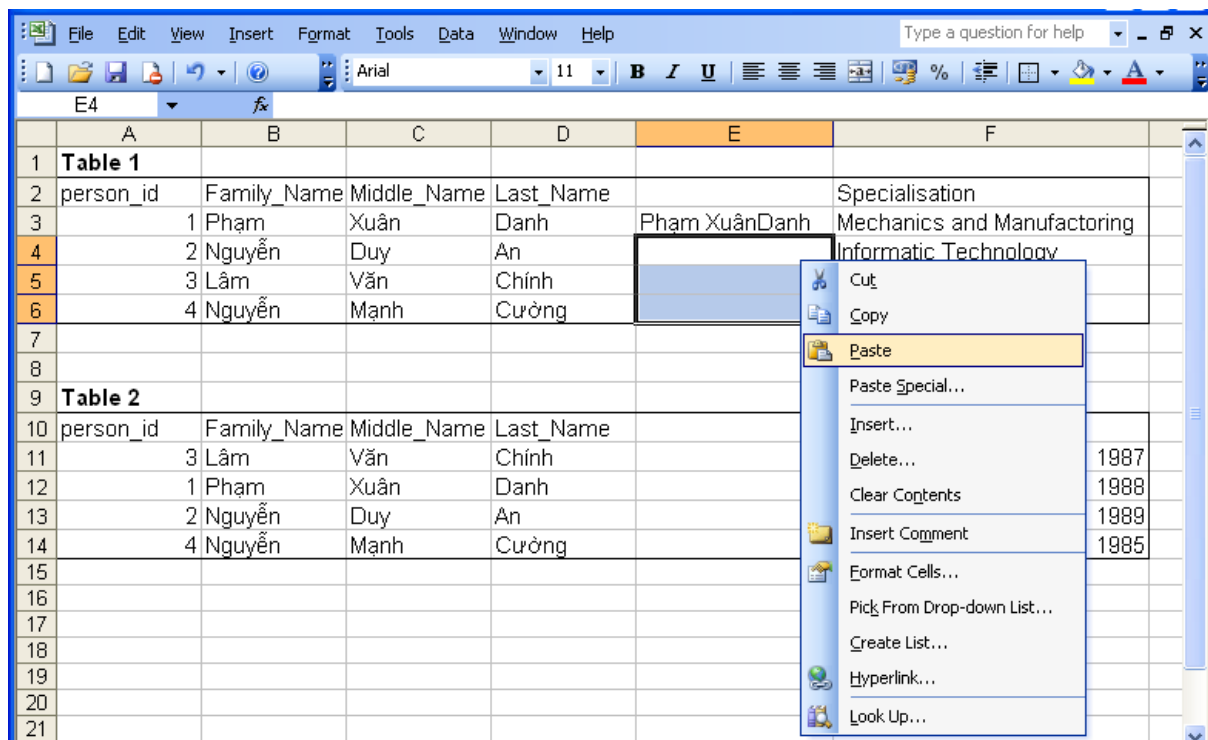
	A	B	C	D	E	F	G
1	Table 1						
2	person_id	Family_Name	Middle_Name	Last_Name		Specialisation	
3	1	Phạm	Xuân	Danh	Phạm XuânDanh	Mechanics and Manufacturing	
4	2	Nguyễn	Duy	An		Informatic Technology	
5	3	Lâm	Văn	Chính		Informatic Technology	
6	4	Nguyễn	Mạnh	Cường		Informatic Technology	
7							
8							
9	Table 2						
10	person_id	Family_Name	Middle_Name	Last_Name		birthday	
11	3	Lâm	Văn	Chính		1987	
12	1	Phạm	Xuân	Danh		1988	
13	2	Nguyễn	Duy	An		1989	
14	4	Nguyễn	Mạnh	Cường		1985	
15							
16							
17							
18							
19							
20							
21							

Screenshot 69: Formula to connect cells

Survey Data Management



Screenshot 70: Formula to connect cells – copy



Screenshot 71: Formula to connect cells – paste

Repeating these steps with the second table will lead to the results depicted in screenshot 72. As you can see, there are some blanks in the identifier of some individuals. These blanks have to be removed before we can use this new person identifier. This helps to ensure a consistent identifier in cases where the name was written with blanks in one table and without in another table.

Survey Data Management

Table 1					
person_id	Family_Name	Middle_Name	Last_Name	Specialisation	
1	Phạm	Xuân	Danh	Phạm XuânDanh	Mechanics and Manufacturing
2	Nguyễn	Duy	An	Nguyễn DuyAn	Informatic Technology
3	Lâm	Văn	Chính	LâmVănChính	Informatic Technology
4	Nguyễn	Manh	Cường	Nguyễn ManhCườ	Informatic Technology

Table 2					
person_id	Family_Name	Middle_Name	Last_Name	birthday	
3	Lâm	Văn	Chính	LâmVănChính	1987
1	Phạm	Xuân	Danh	Phạm XuânDanh	1988
2	Nguyễn	Duy	An	Nguyễn DuyAn	1989
4	Nguyễn	Manh	Cường	Nguyễn ManhCườ	1985

Screenshot 72: Formula to connect cells – end result

In order to be able to remove the blanks, the underlying formula, which allowed writing information from three cells into one, has to be removed. This can be done by copying the cells with the formula (next screenshot) and pasting them again with “paste special”.

Table 1					
person_id	Family_Name	Middle_Name	Last_Name	Specialisation	
1	Phạm	Xuân	Danh	Phạm XuânDanh	Mechanics and Manufacturing
2	Nguyễn	Duy	An	Nguyễn DuyAn	Informatic Technology
3	Lâm	Văn	Chính	LâmVănChính	Informatic Technology
4	Nguyễn	Manh	Cường	Nguyễn ManhCườ	Informatic Technology

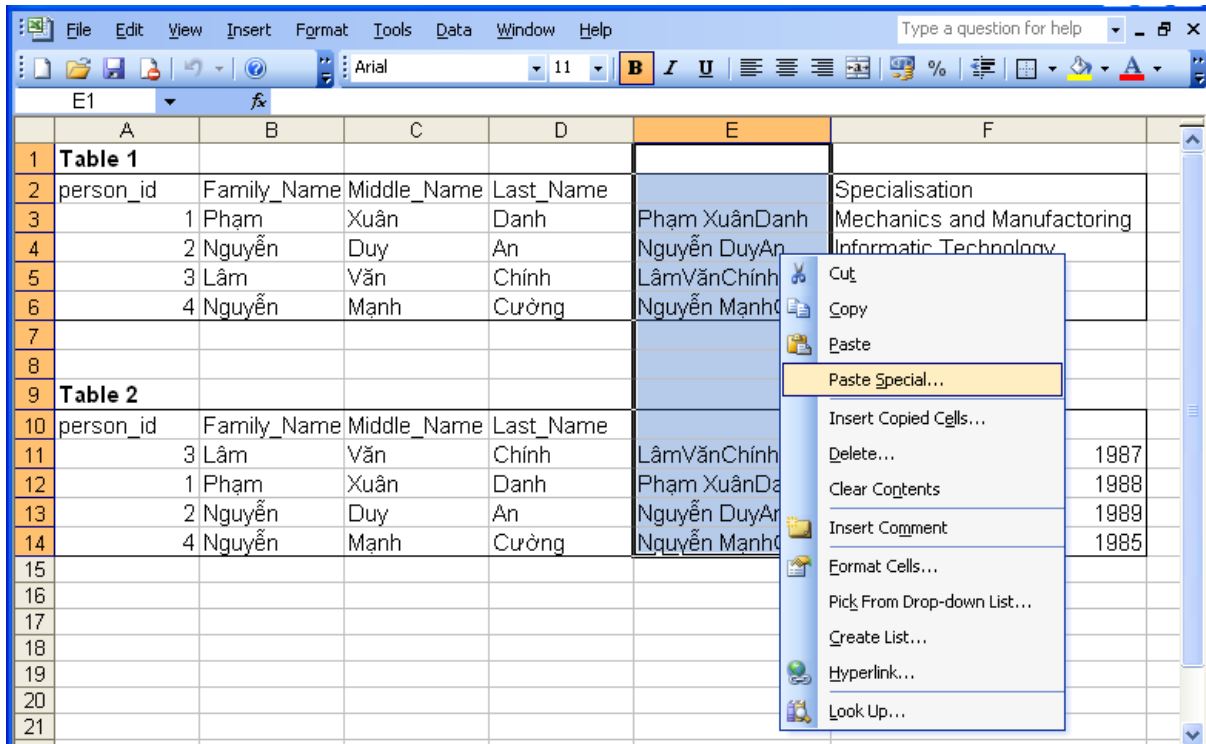
Table 2					
person_id	Family_Name	Middle_Name	Last_Name	birthday	
3	Lâm	Văn	Chính	LâmVănChính	1987
1	Phạm	Xuân	Danh	Phạm XuânDanh	1988
2	Nguyễn	Duy	An	Nguyễn DuyAn	1989
4	Nguyễn	Manh	Cường	Nguyễn ManhCườ	1985

Screenshot 73: Formula to connect cells – removing formula

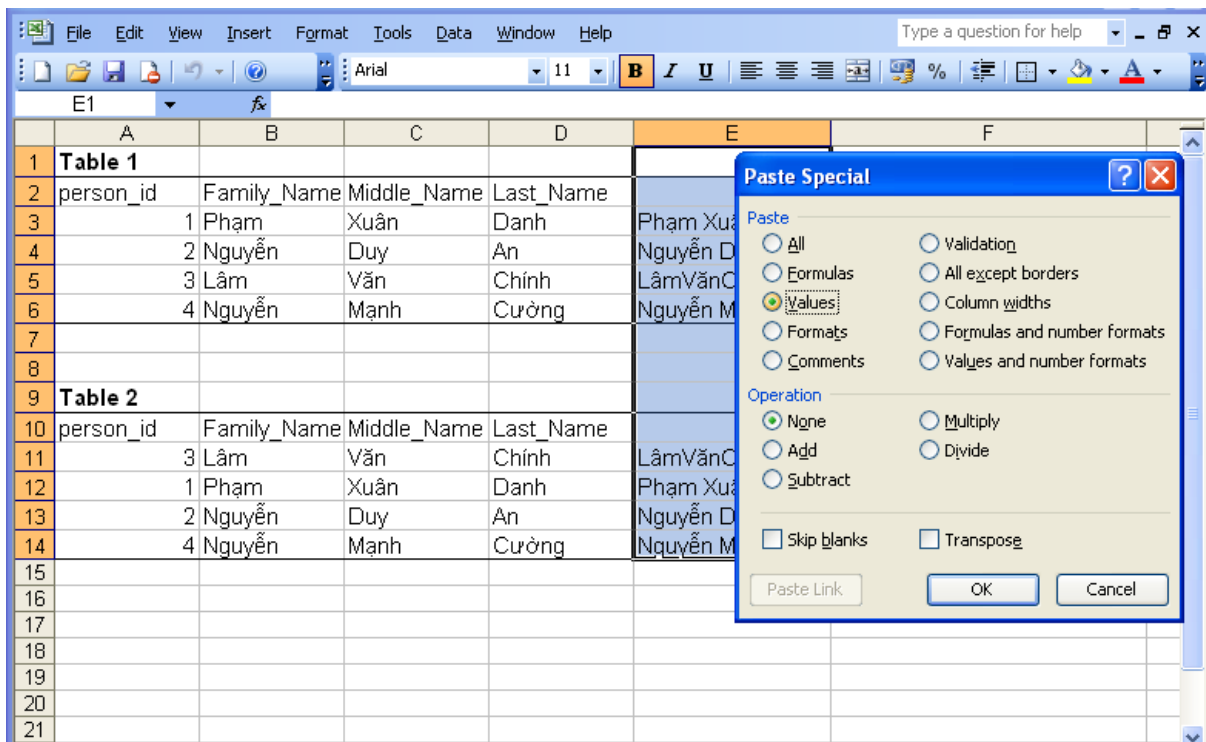
Right click on the selected cells and press “copy”. Then do another right click on the cells and select “Paste Special...”. A dialogue will appear where you can select which information you

Survey Data Management

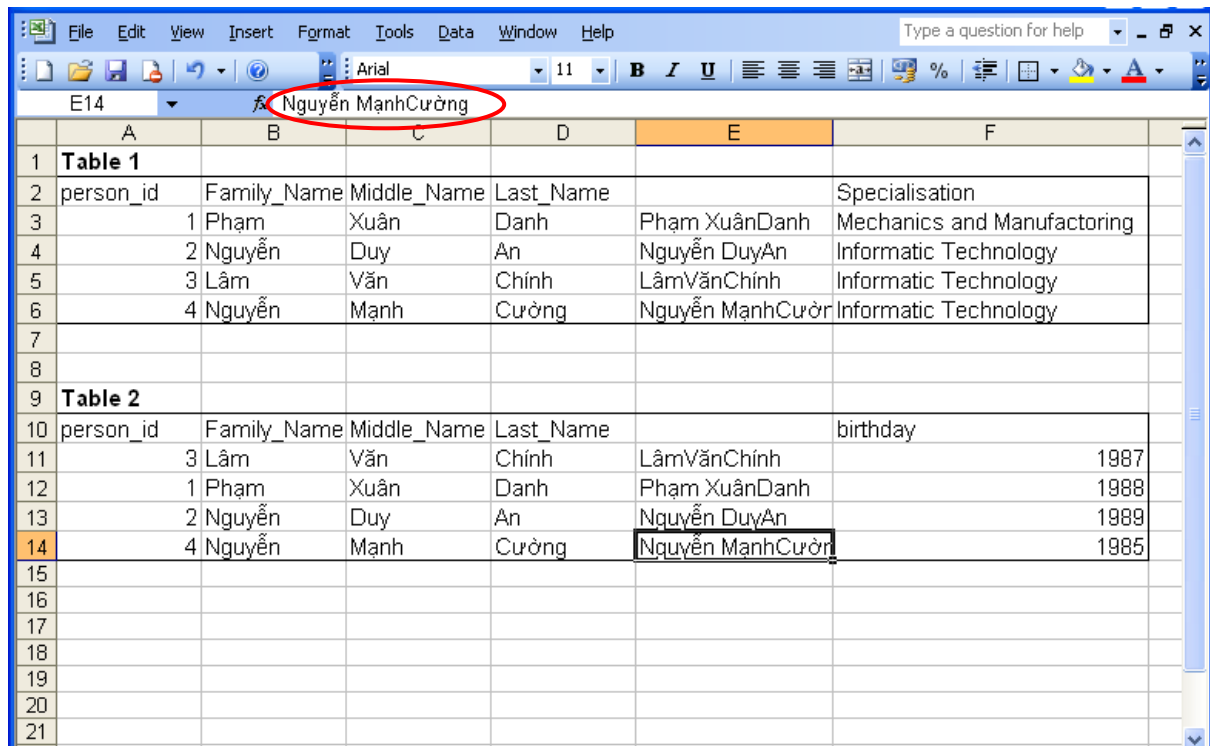
would like to paste. As we want to remove the formulas while keeping the content of the cells, we select “Values” in the dialogue and finish by clicking on the “OK” button.



Screenshot 74: Formula to connect cells – removing formula

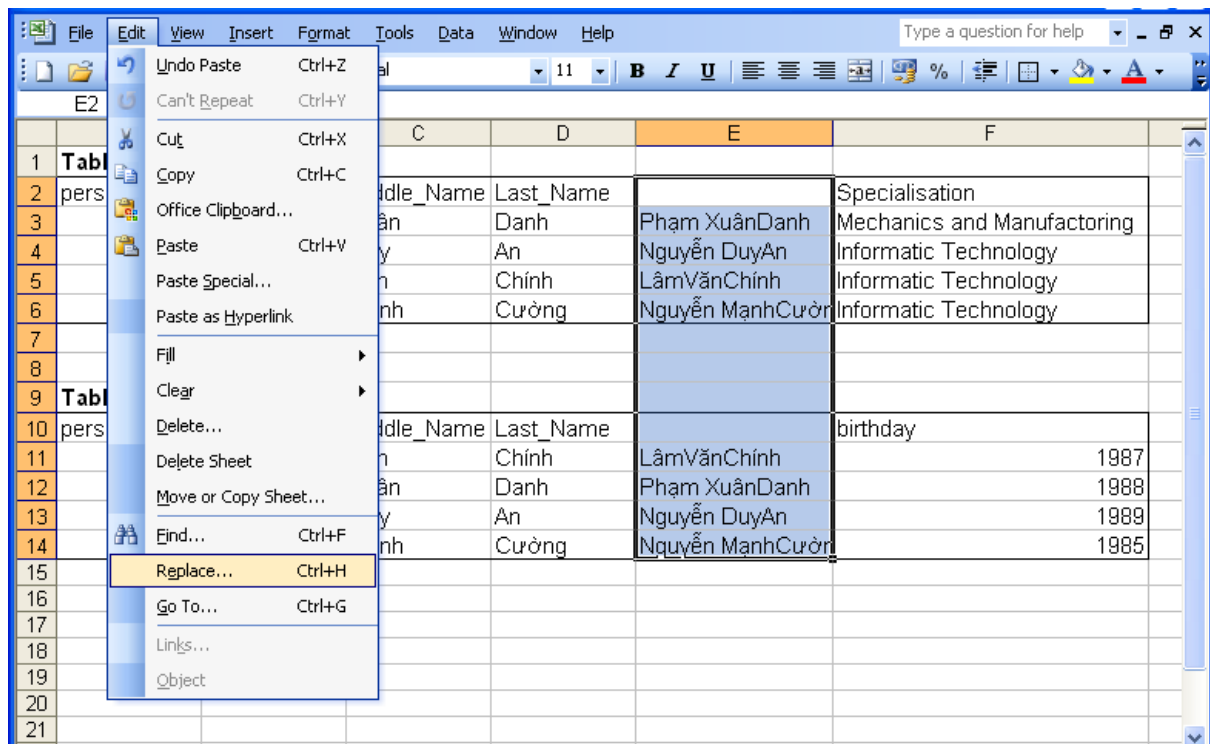


Screenshot 75: Formula to connect cells – paste-special dialogue

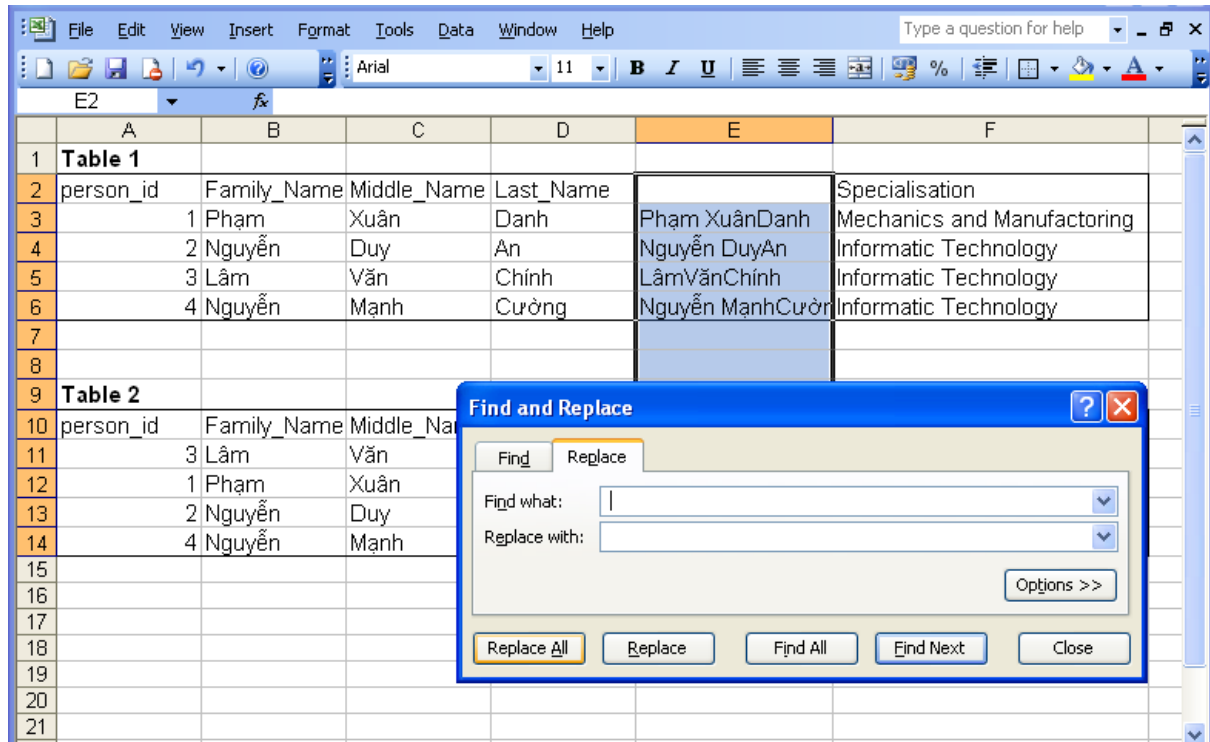


Screenshot 76: Formula to connect cells – removed formula

As depicted in the screenshot above, the formula is removed from the cells. You can now remove the blanks by using the replace tool of Excel. Please select the cells of the new identifier first. Then go to the “Edit” tab, and click on “Replace...” (a faster and easier way to do this is to press the keys “Ctrl” and “h” on your keyboard at the same time), which will open the find and replace dialogue window.

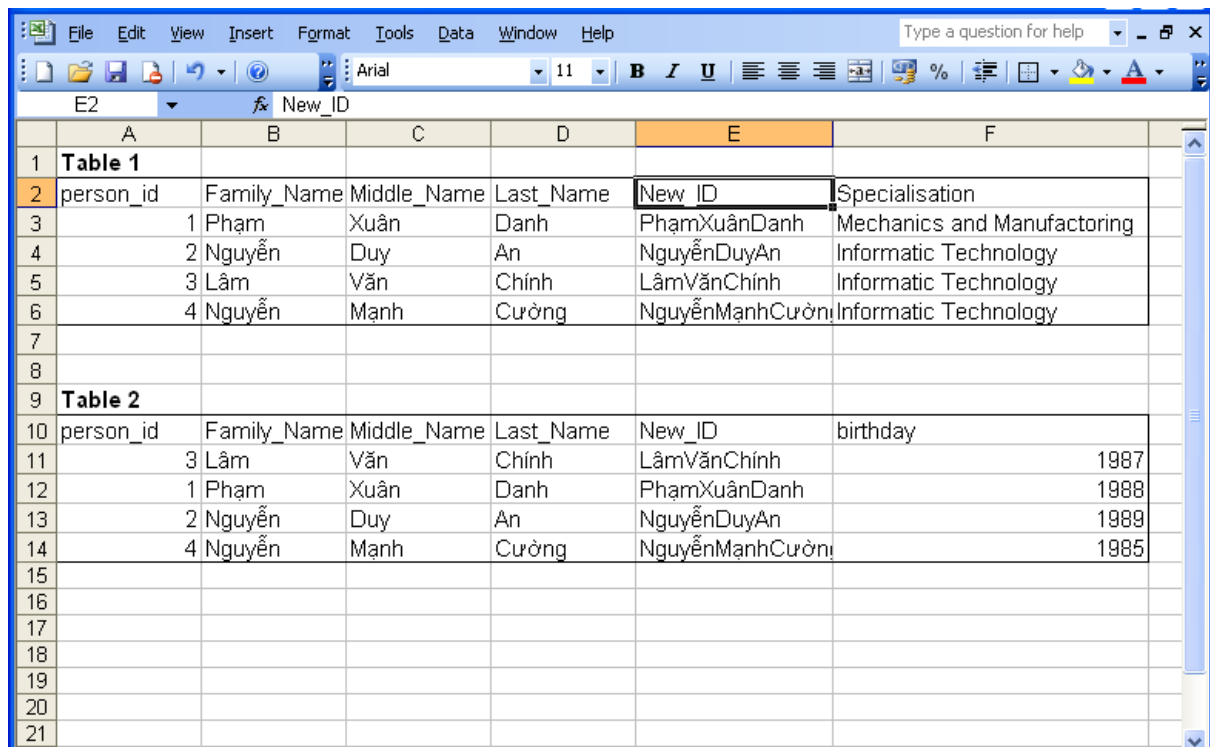


Screenshot 77: Remove blanks - Replace



Screenshot 78: Remove blanks - Replace

For removing the blanks, Excel has to search for blanks and replace them with “nothing”. Therefore, we only have to type one blank into the “Find what:” line. Note that nothing is typed in the “Replace with:” line. Now you click on “Replace all” and close the dialogue.

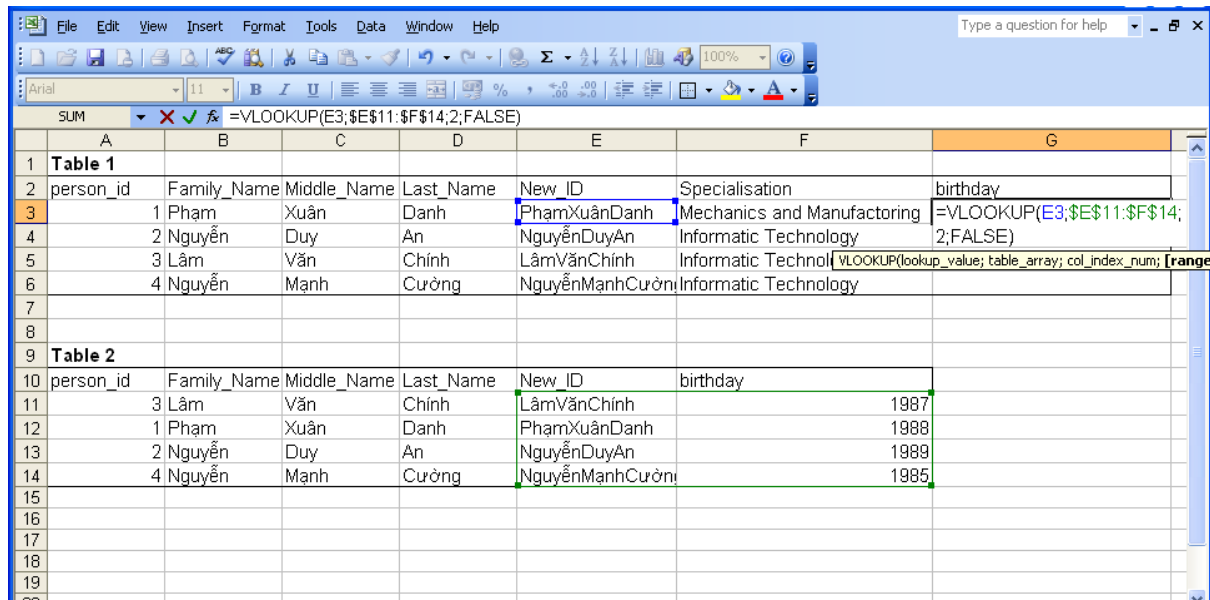


Screenshot 79: New identifier

As you see, all blanks have been removed. The data from the two data-sets can now be merged on this new identifier using the VLOOKUP function, which is explained in detail in the next chapter.

8.3 Merging data

Excel provides a function that looks up entries from one list and returns entries from another list. This can be compared with a telephone book. You give Excel a name (the new identifier in table 1 of the example) and Excel will look up the number (the birthday in table 2 of the example) and return it to you. In our example, we want to merge the birthday information from table 2 into table 1.



Screenshot 80: Function VLOOKUP

Command VLOOKUP:

SYNTAX: =VLOOKUP(lookup_value;table_array;col_index_num;[range_lookup])
 “lookup_value” is the value that Excel should lookup in a second table. This second table is defined by “table_array”. “col_index_num” tells Excel which of the columns of the second table to return in case it finds the “lookup_value” in the second table. Finally, “range_lookup” defines whether Excel should return a value only in case of an exact match with the lookup value (type “FALSE”) or whether it should return anything that is similar to the lookup value (type “TRUE”).

TYPE: =VLOOKUP(E3;\$E\$11:\$F\$14;2;FALSE)

PRESS: ENTER

RESULT: Looks up the new identifier of table 1 (cell E3) in table 2 (E11:F14) and returns information on birthday (column 2) only if it finds an exact match (FALSE).

The function used for merging is the VLOOKUP function. To start, please start in cell G3 and type: “=VLOOKUP(E3;E11:F14;2;FALSE)”. The first part (E3) tells Excel which value to look up in the second table. This second table is defined in the next part of the function after the semicolon by (\$E\$11:\$F\$14). You can actually select this second table with your mouse. Make sure that the first column of the area that you select holds the new identifier. After selecting the area with the mouse, you have to insert dollar symbols (\$) before and after the characters of the table reference. This will keep the reference to the table fixed, even when you move the formula down to the next cells. Alternatively, you can press “F4” on your keyboard after selecting the table with the mouse; this will also add the dollar signs. The

Survey Data Management

third part of the function tells Excel which information from which column to return. In this example you type “2”, because the birthday is found in the second column of the selected area. Finally, we tell Excel to return only exact matches of the identifier in both tables by typing “FALSE”. If we typed “TRUE”, Excel would return also approximate matches of the identifier. After you finish by pressing “Enter”, Excel will return the birthday from the second table. Note that we could also have used the person_id instead of the New_ID to match information in this example.

The screenshot shows an Excel spreadsheet with two tables. Table 1 is in rows 2-6 and Table 2 is in rows 10-14. Both tables have columns for person_id, Family_Name, Middle_Name, Last_Name, New_ID, Specialisation, and birthday. Cell G4 contains the formula =VLOOKUP(2,Table2,2,FALSE) and displays the value 1987.

Table 1						
person_id	Family_Name	Middle_Name	Last_Name	New_ID	Specialisation	birthday
1	Phạm	Xuân	Danh	PhạmXuânDanh	Mechanics and Manufacturing	1988
2	Nguyễn	Duy	An	NguyễnDuyAn	Informatic Technology	
3	Lâm	Văn	Chính	LâmVănChính	Informatic Technology	
4	Nguyễn	Manh	Cường	NguyễnManhCường	Informatic Technology	

Table 2						
person_id	Family_Name	Middle_Name	Last_Name	New_ID	Specialisation	birthday
3	Lâm	Văn	Chính	LâmVănChính		1987
1	Phạm	Xuân	Danh	PhạmXuânDanh		1988
2	Nguyễn	Duy	An	NguyễnDuyAn		1989
4	Nguyễn	Manh	Cường	NguyễnManhCường		1985

Screenshot 81: Function VLOOKUP

We can now copy the formula as in the previous example.

The screenshot shows the same Excel spreadsheet as Screenshot 81. The formula =VLOOKUP(2,Table2,2,FALSE) has been copied to cell G9, and it now displays the value 1989.

Table 1						
person_id	Family_Name	Middle_Name	Last_Name	New_ID	Specialisation	birthday
1	Phạm	Xuân	Danh	PhạmXuânDanh	Mechanics and Manufacturing	1988
2	Nguyễn	Duy	An	NguyễnDuyAn	Informatic Technology	1989
3	Lâm	Văn	Chính	LâmVănChính	Informatic Technology	1987
4	Nguyễn	Manh	Cường	NguyễnManhCường	Informatic Technology	1985

Table 2						
person_id	Family_Name	Middle_Name	Last_Name	New_ID	Specialisation	birthday
3	Lâm	Văn	Chính	LâmVănChính		1987
1	Phạm	Xuân	Danh	PhạmXuânDanh		1988
2	Nguyễn	Duy	An	NguyễnDuyAn		1989
4	Nguyễn	Manh	Cường	NguyễnManhCường		1985

Screenshot 82: Function VLOOKUP

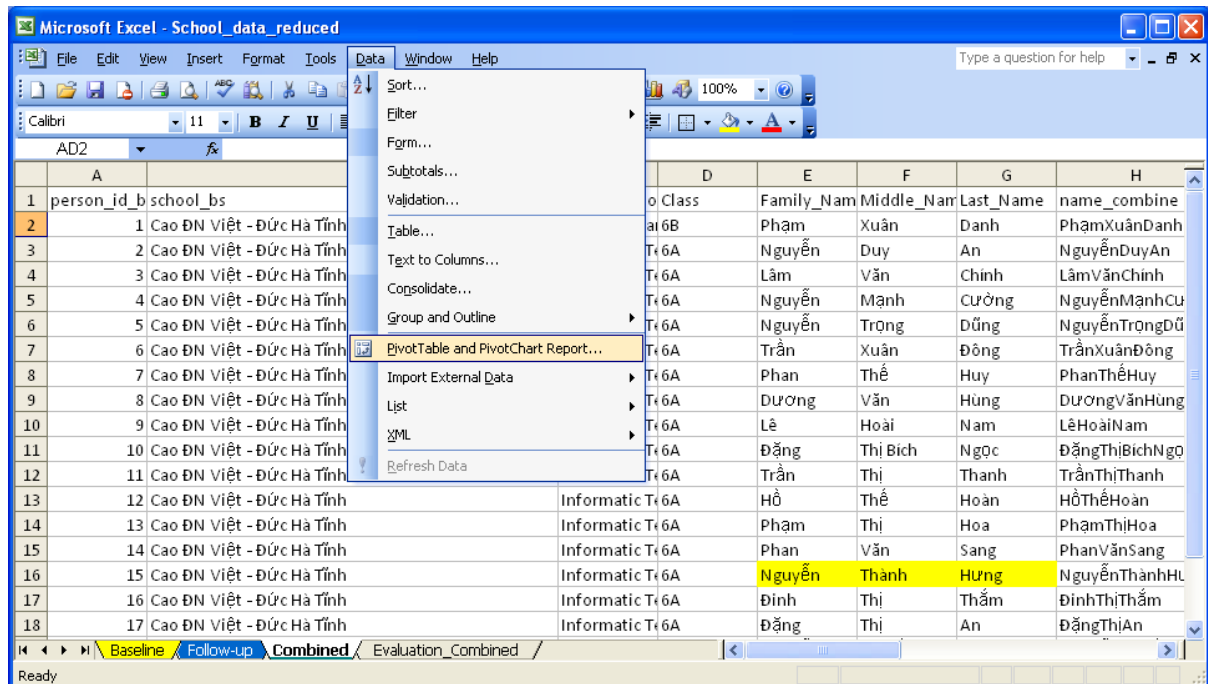
Using this function, you can also merge information from different spreadsheets. An example will be given in chapter 9.2.

9 Creating overview statistics

9.1 Pivot tables to summarize questionnaire information

For the purpose of learning about overview statistics, a larger data-set is used. Please open the Excel file School_data_reduced.xls that was delivered with this handbook and select the “Combined” worksheet. The information in this worksheet holds baseline and follow-up tracer information which were merged using the VLOOKUP function.

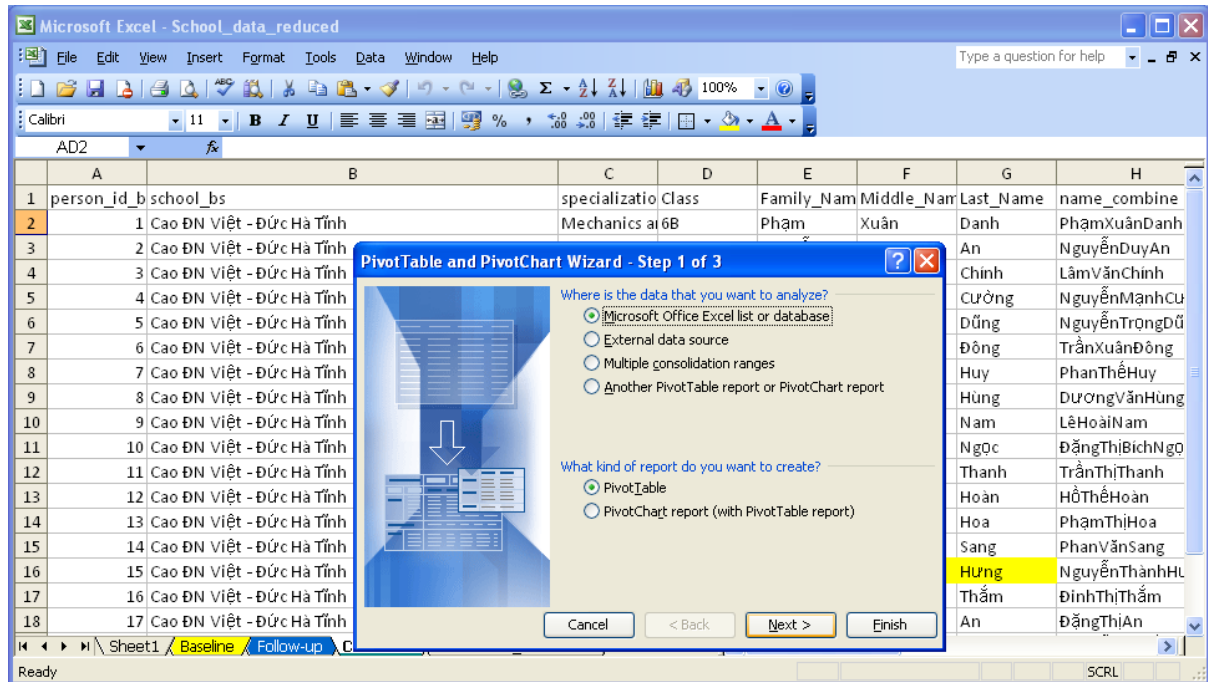
Among the different tools that Excel provides for data-analysis, the pivot-table tool is the most powerful, inasmuch as it allows analyzing data in a very time effective manner. In order to start the pivot table tool, please click on “Data” in the menu and select “PivotTable and PivotChart Report...”



Screenshot 83: Creating a PivotTable

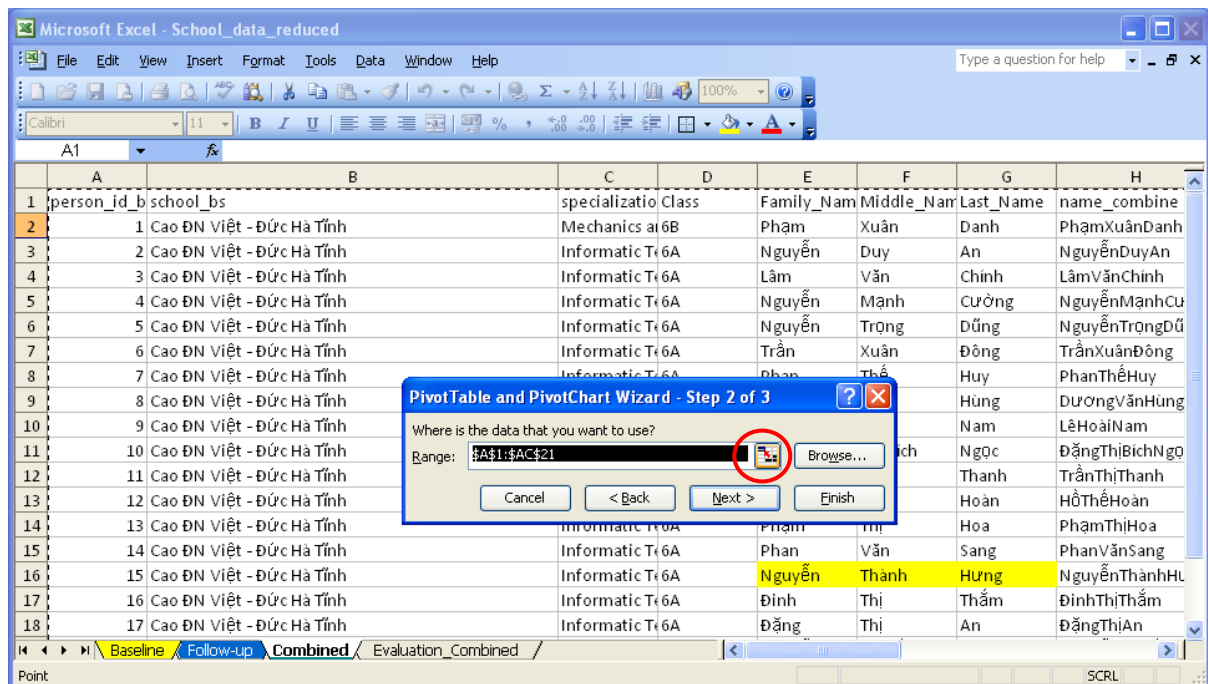
A dialogue window will appear where you can select the data source. As we are using data in the spreadsheet we are working with, no changes are needed in this dialogue and you can click “Next”.

Survey Data Management



Screenshot 84: Creating a PivotTable

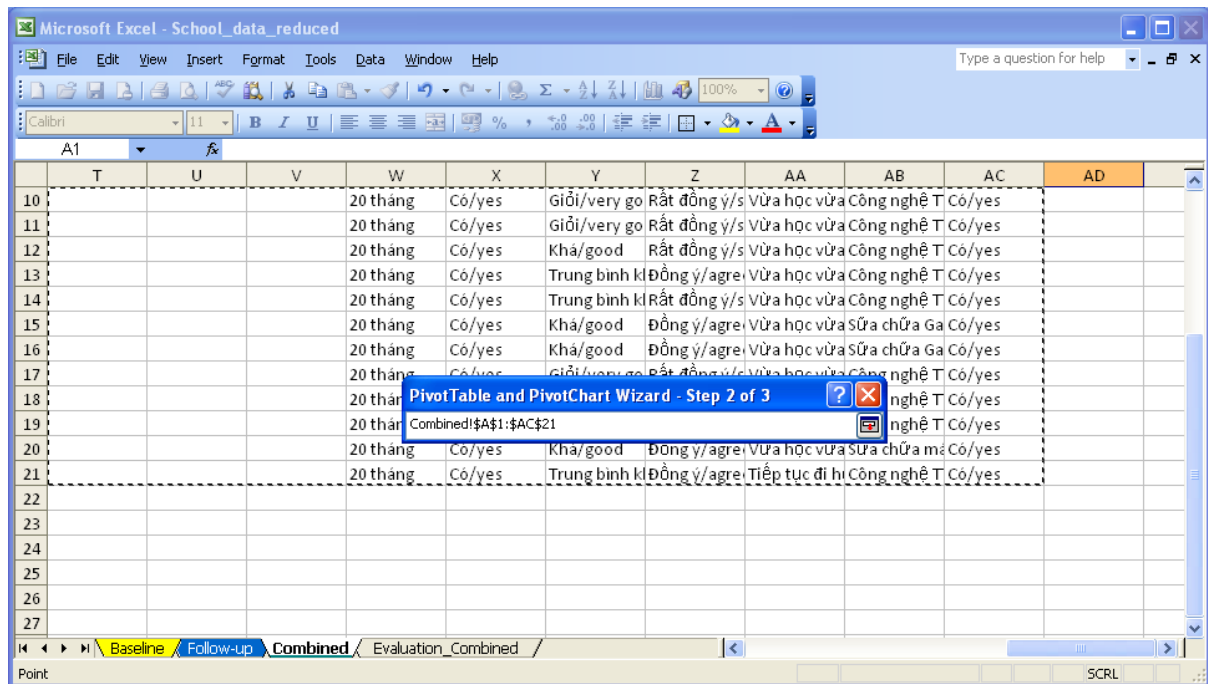
Excel will now ask you to select the data that you want to work with in the PivotTable. When you are satisfied with the pre-selection of data that Excel made, click “Next”, otherwise click onto the button that is left of the “Browse” button (highlighted with a red circle in screenshot 85).



Screenshot 85: Creating a PivotTable – selecting the relevant data

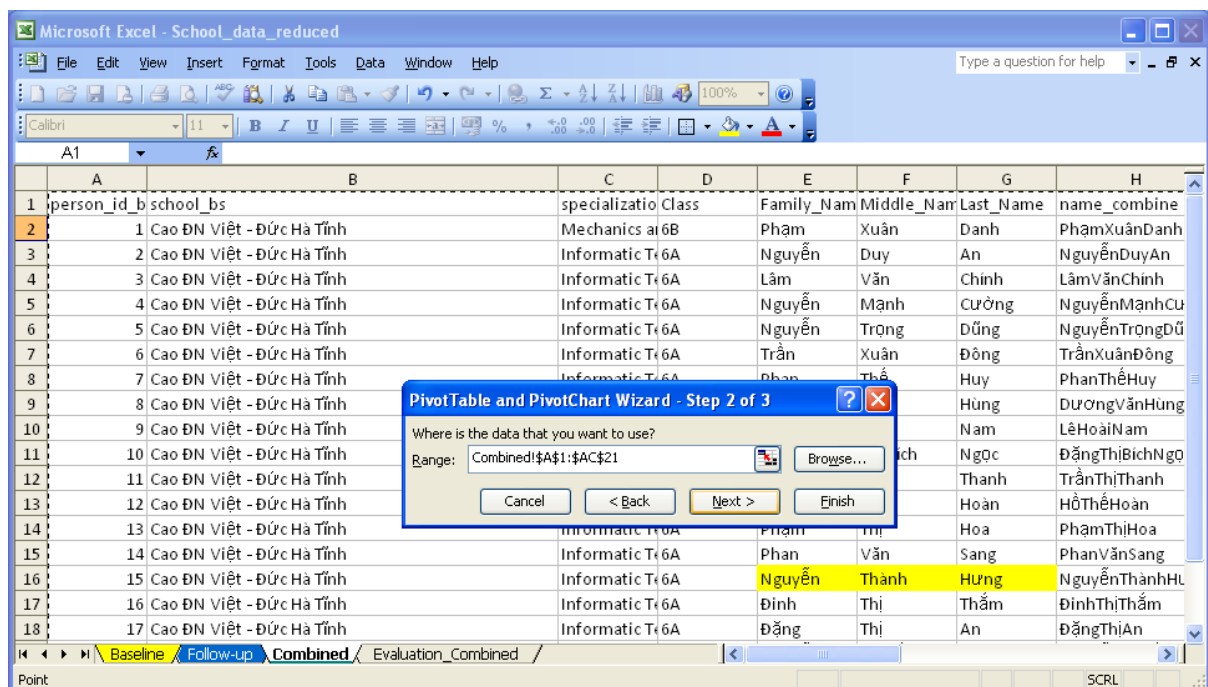
In case you opted for selecting the data yourself, you can now use your mouse to select the area that you want to include into the PivotTable. Make sure that you also select the first row with the variable names when selecting the data.

Survey Data Management



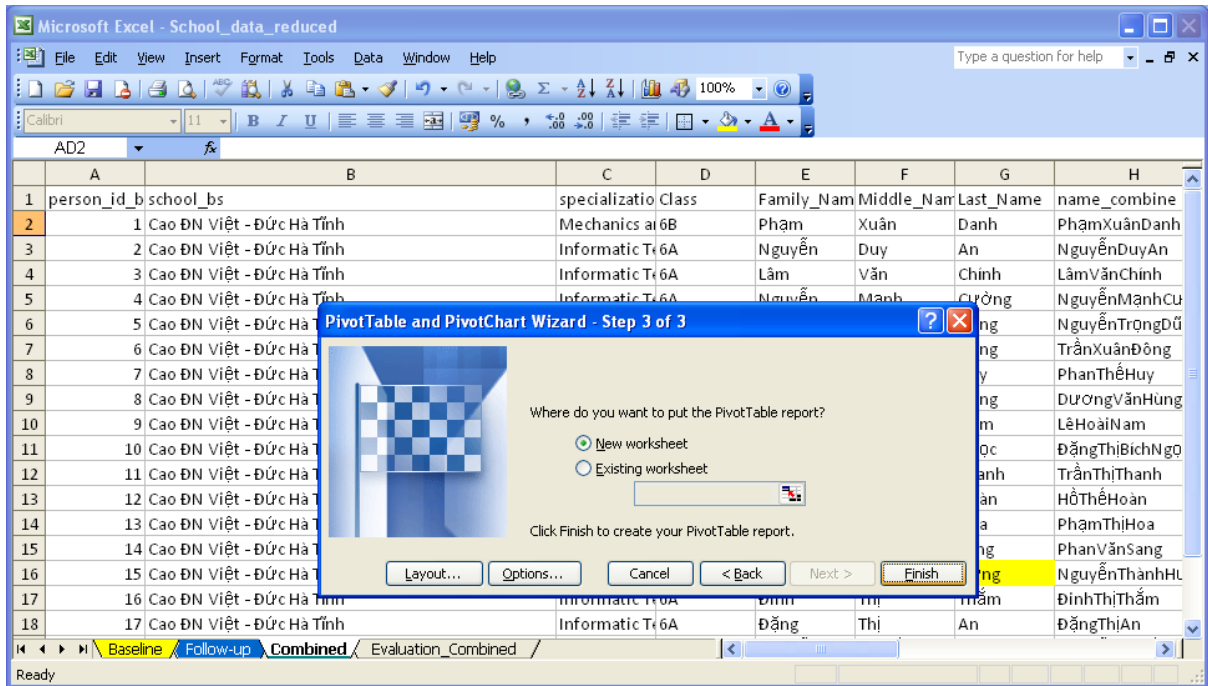
Screenshot 86: Creating a PivotTable – selecting the relevant data

After you selected the data, press enter and then proceed by clicking “Next”.



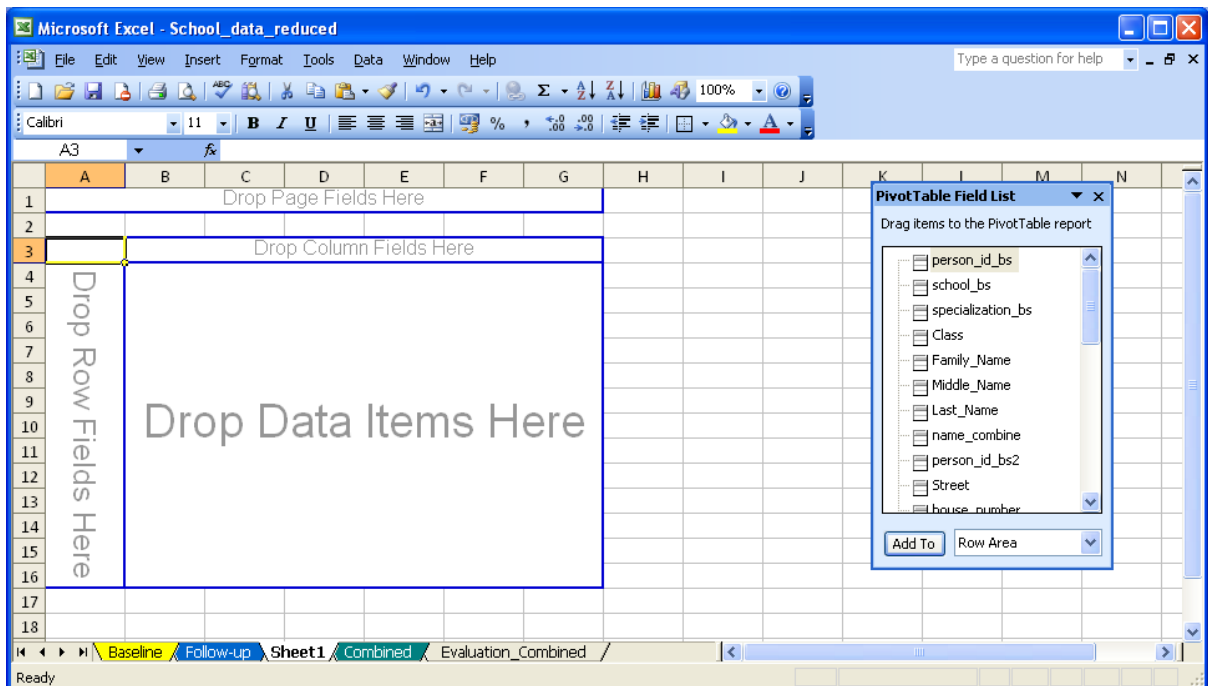
Screenshot 87: Creating a PivotTable – selecting the relevant data

Finally, Excel will ask you, whether you would like to place the PivotTable in the existing, or in a new worksheet. Please click on “New worksheet” and press “Finish”.



Screenshot 88: Creating a PivotTable

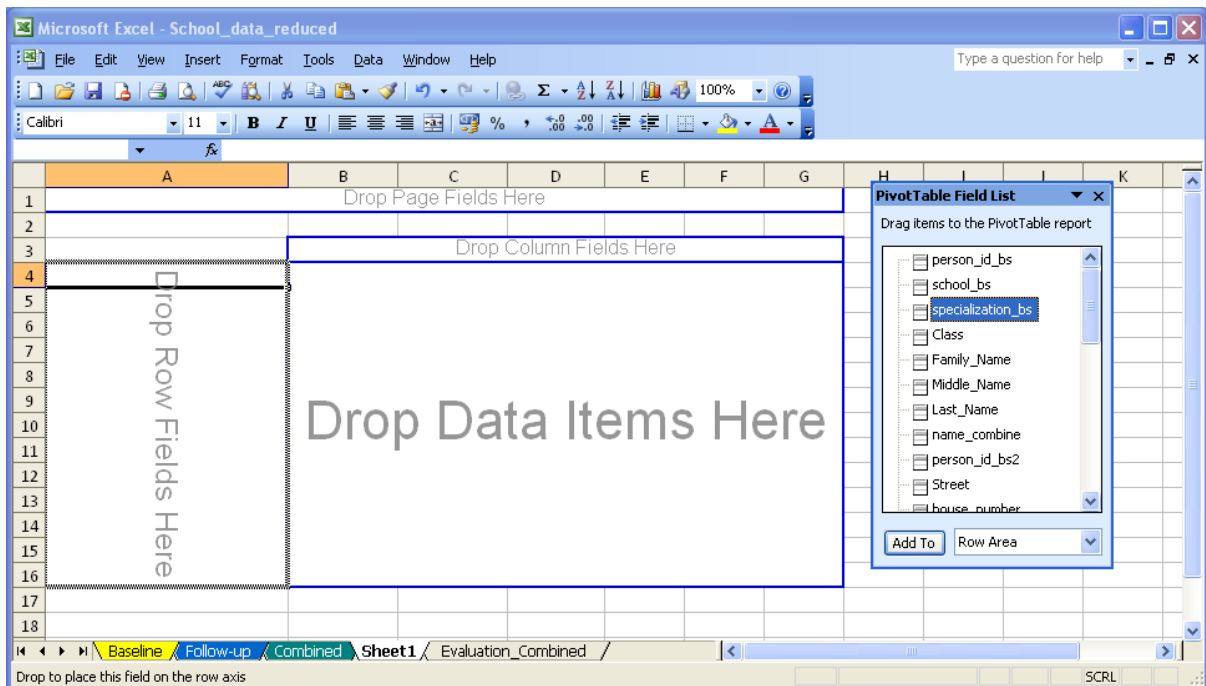
You should now end up with an empty table similar to the one in the screenshot below. On the right-hand side of the table you find all the variables in the data-set.



Screenshot 89: Creating a PivotTable

You can drag items from the variable list onto the empty fields of the table. In this example, we would like to know the number of males and females studying in the different fields of specialization. Please drag the variable specialization_bs from the variable menu into the row field of the table (the “_bs” attached to the variable name indicates that this is a variable from the baseline).

Survey Data Management



Screenshot 90: Creating a PivotTable

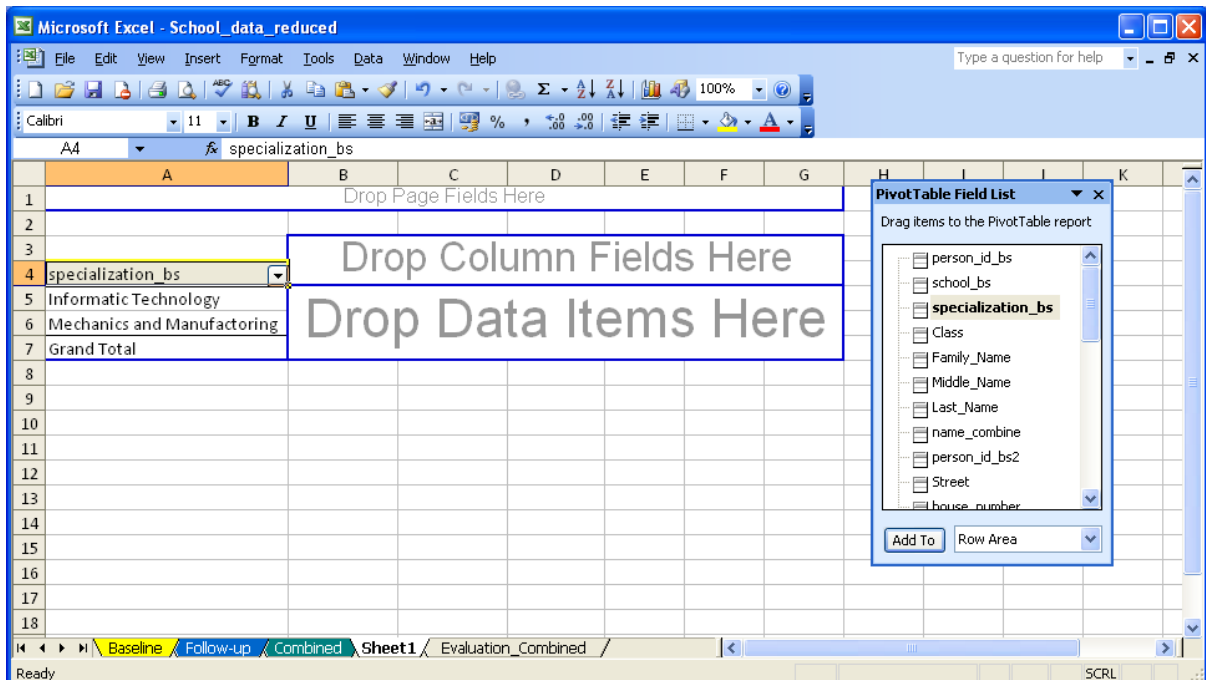
Action drag and drop:

SELECT: Place the mouse on the variable that you want to move to the table.

PRESS: Press the left mouse button and hold

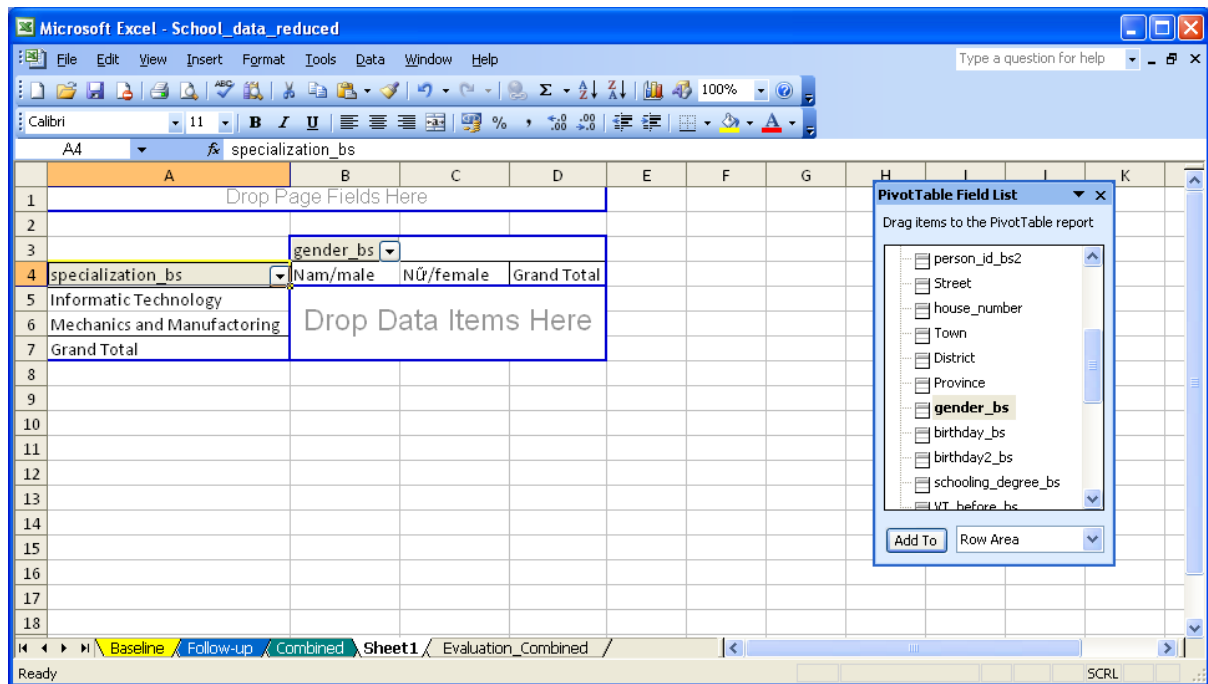
DRAG: Move the mouse to the place where you want to place the variable

DROP: Release the left mouse button



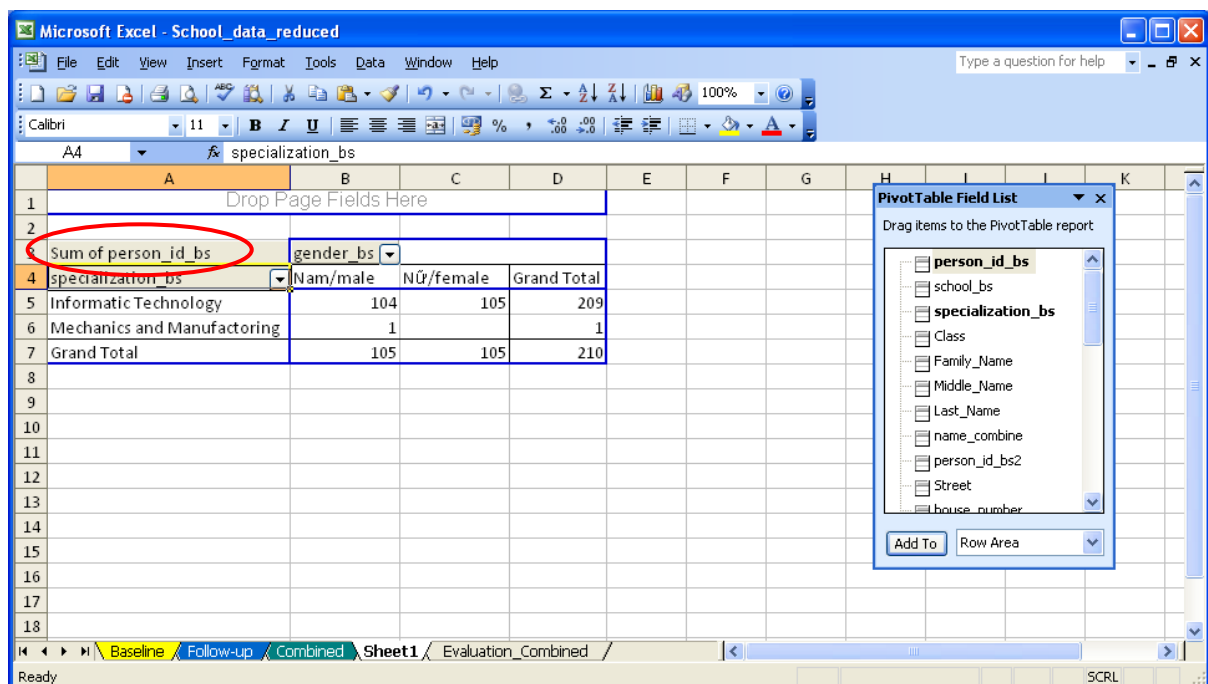
Screenshot 91: Creating a PivotTable – selecting variables

Survey Data Management

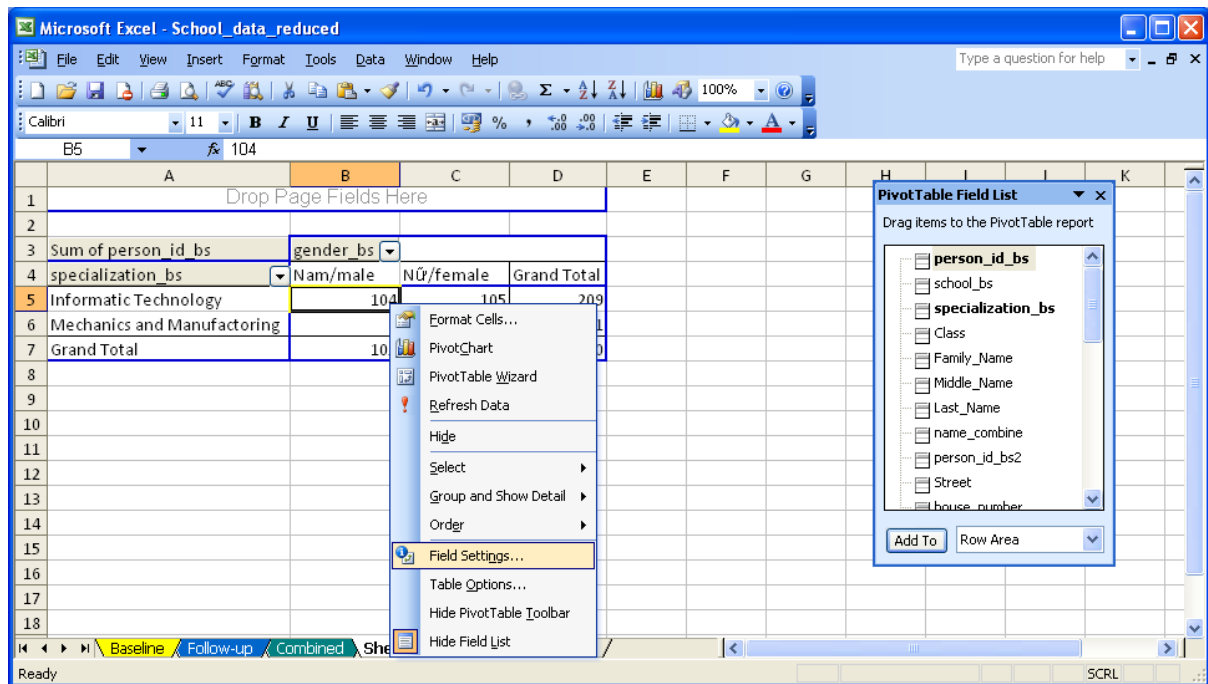


Screenshot 92: Creating a PivotTable – selecting variables

Then drop `gender_bs` onto the column field of the table. As you see, the table is still empty except for the column and row fields. The reason is, that Excel needs to know, which information to show within the table. This can be a frequency or the mean of a variable. To show the frequency, we select a variable that has observations for all individuals in the data-set. `person_id_bs` is such a variable. Please drop it into the table's empty data-field. As you can see in the upper left of the table, Excel displays now the sum of `person_id_bs`. In fact, we need the frequency. In order to select the frequency or any other statistics, please click right onto the data-field and select "Field Settings..." from the drop-down menu.

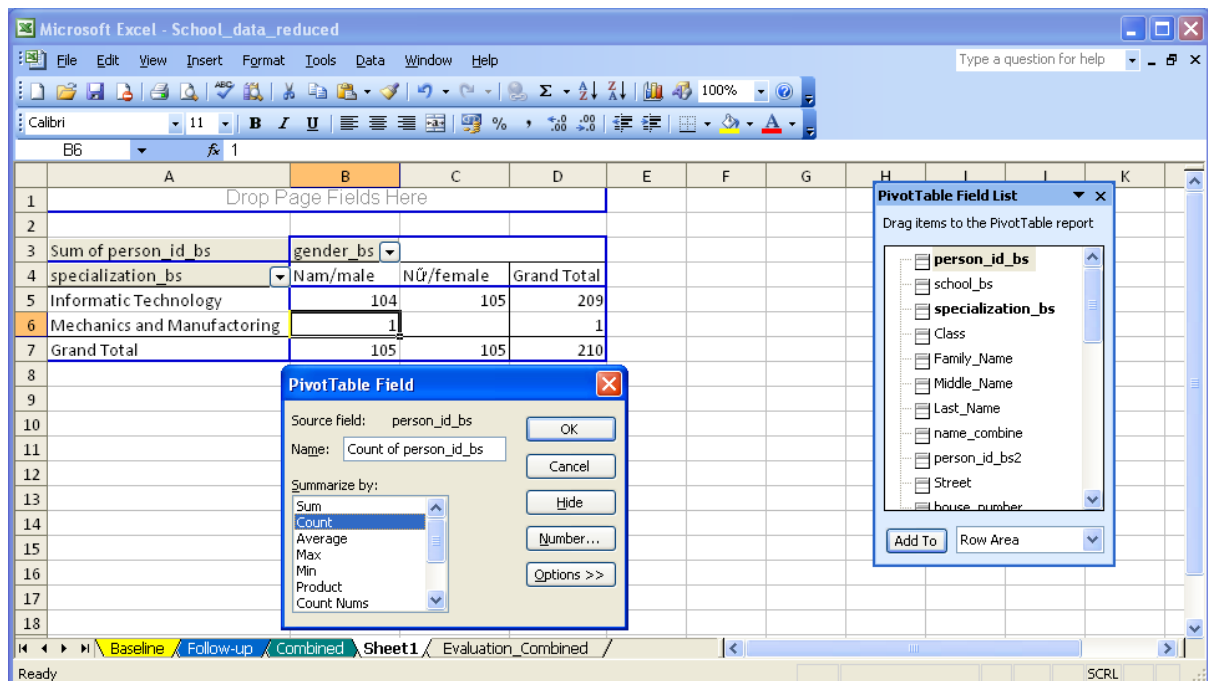


Screenshot 93: Creating a PivotTable – selecting statistics



Screenshot 94: Creating a PivotTable – selecting statistics

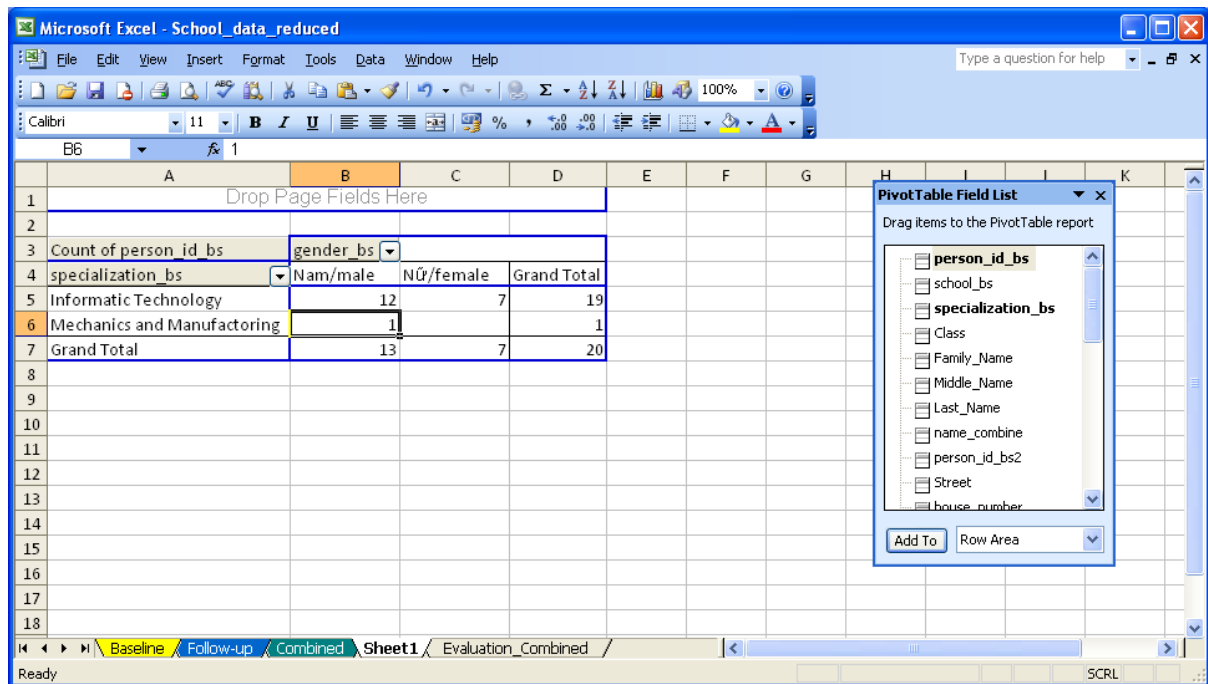
A dialogue box will appear where you can select “Count” among other statistics. When you press “OK”, you will end up with a frequency table of specialization and gender.



Screenshot 95: Creating a PivotTable – selecting statistics

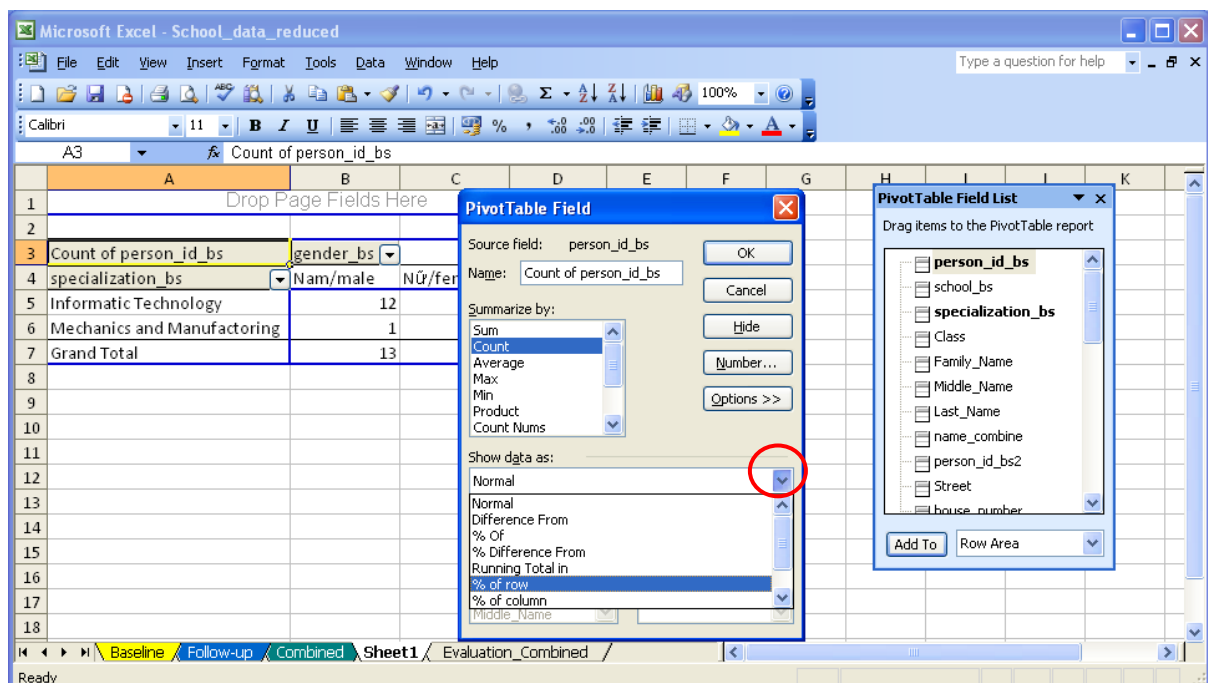
Instead of absolute numbers, depending on the question you would like to answer with the statistic, the share of females and males in each course of study might be of interest. This would allow for a ranking of courses according to the share of females or males studying in the courses.

Survey Data Management



Screenshot 96: Creating a PivotTable

For displaying percentages instead of absolute figures, go to the “Field Settings...” of the variable “person_id_bs” again. Then click on “Options” and on the drop down menu below “Show data as:” (highlighted with a red circle in the screenshot below). You can select row and column percentages in this drop down list. In our case, we select “% of row” and click on “OK”.

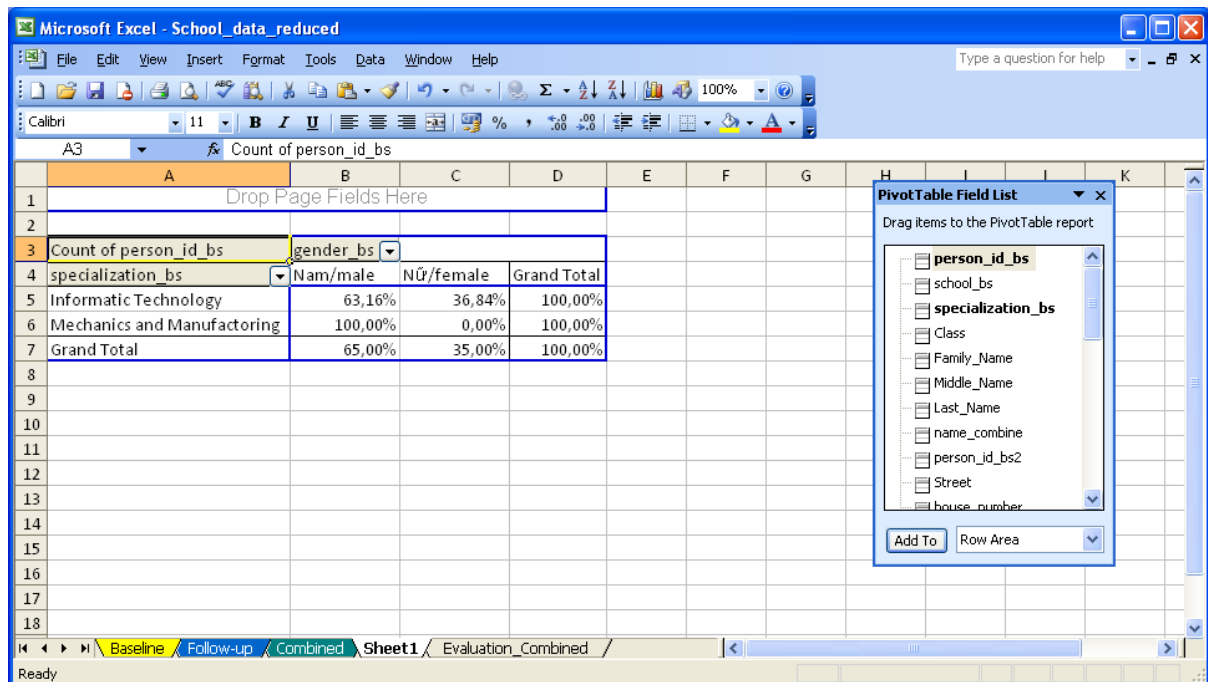


Screenshot 97: Creating a PivotTable – selecting statistics

The resulting table displays the share of male and female students in each course of specialization. Instead of course of specialization you could also add any other categorical variable to gain information on the gender distribution (an example would be the city of origin of the students to see, whether males or females exhibit the greater mobility when it

Survey Data Management

comes to more remote cities. In this context, one could also add the course of studies instead of the gender variable. This would answer the question whether certain courses of study are preferred by students with a more remote city of origin, giving hints on the own competitiveness).



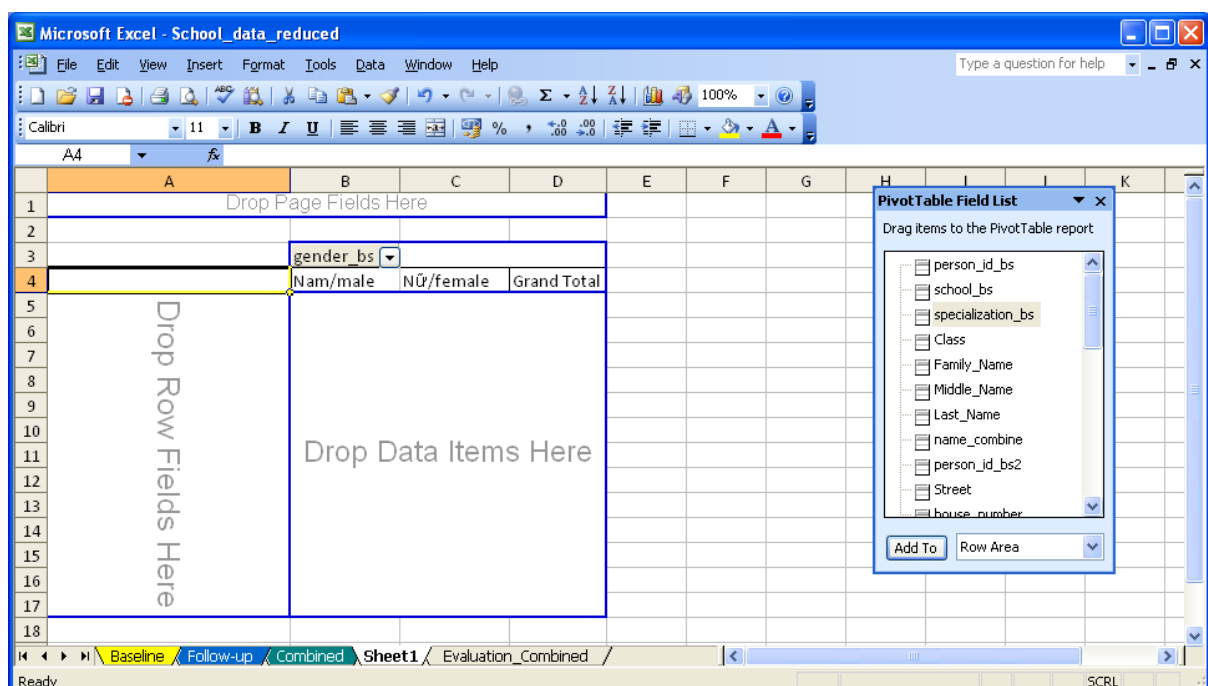
The screenshot shows a PivotTable in Microsoft Excel. The PivotTable is titled 'Count of person_id_bs' and is located in the range A3:D7. The row field is 'specialization_bs' and the column field is 'gender_bs'. The PivotTable shows the following data:

specialization_bs	Nam/male	NŪ/female	Grand Total
Informatic Technology	63,16%	36,84%	100,00%
Mechanics and Manufacturing	100,00%	0,00%	100,00%
Grand Total	65,00%	35,00%	100,00%

Screenshot 98: Creating a PivotTable

As you will see in a later chapter, you can add the same variable multiple times into the data field of the table. This allows you to display percentages and counts at once.

It may be also interesting to tabulate the average rating of training quality by gender. To create such a table, please remove the variables `specialization_bs` and `person_id_bs` by selecting the variables in the table with the mouse and dragging them out of the table.



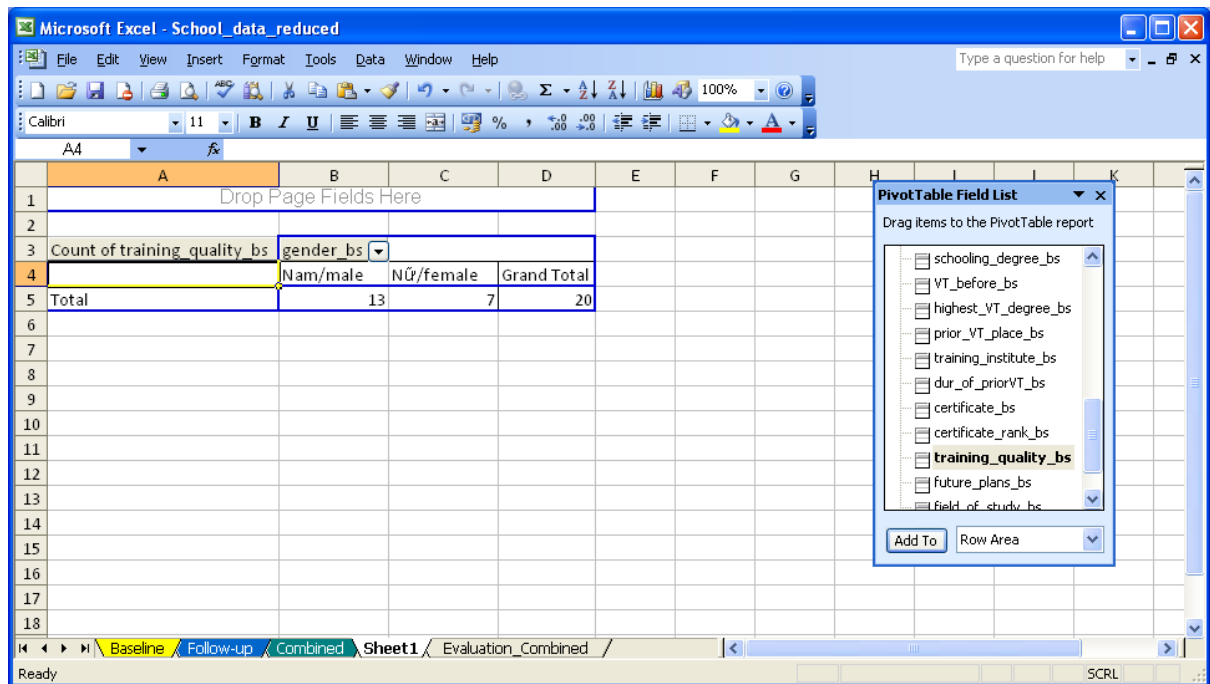
The screenshot shows the PivotTable field list in Microsoft Excel. The field list is titled 'PivotTable Field List' and is located in the range H3:I18. The field list shows the following variables:

- person_id_bs
- school_bs
- specialization_bs
- Class
- Family_Name
- Middle_Name
- Last_Name
- name_combine
- person_id_bs2
- Street
- house_number

The variables 'specialization_bs' and 'person_id_bs' are selected for removal. The PivotTable area is empty, showing 'Drop Page Fields Here' and 'Drop Data Items Here'.

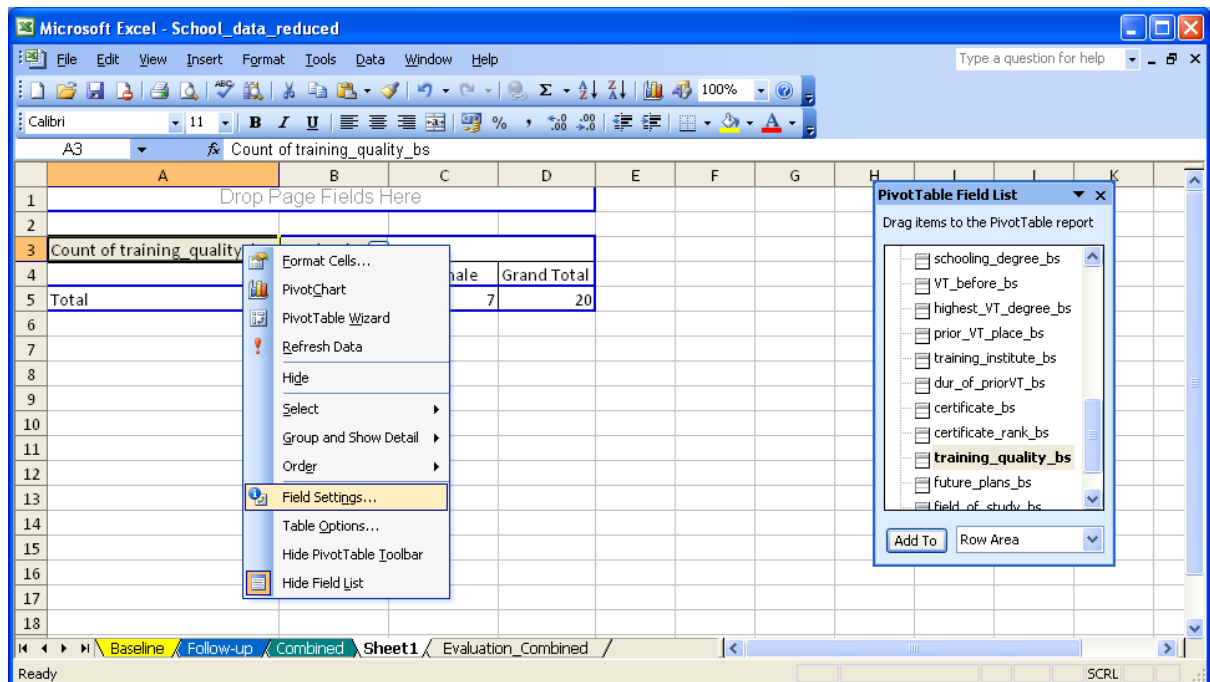
Screenshot 99: Creating a PivotTable

Now you can add the variable training_quality_bs with the quality rating of training by dropping it into the data field in the middle of the table.



Screenshot 100: Creating a PivotTable

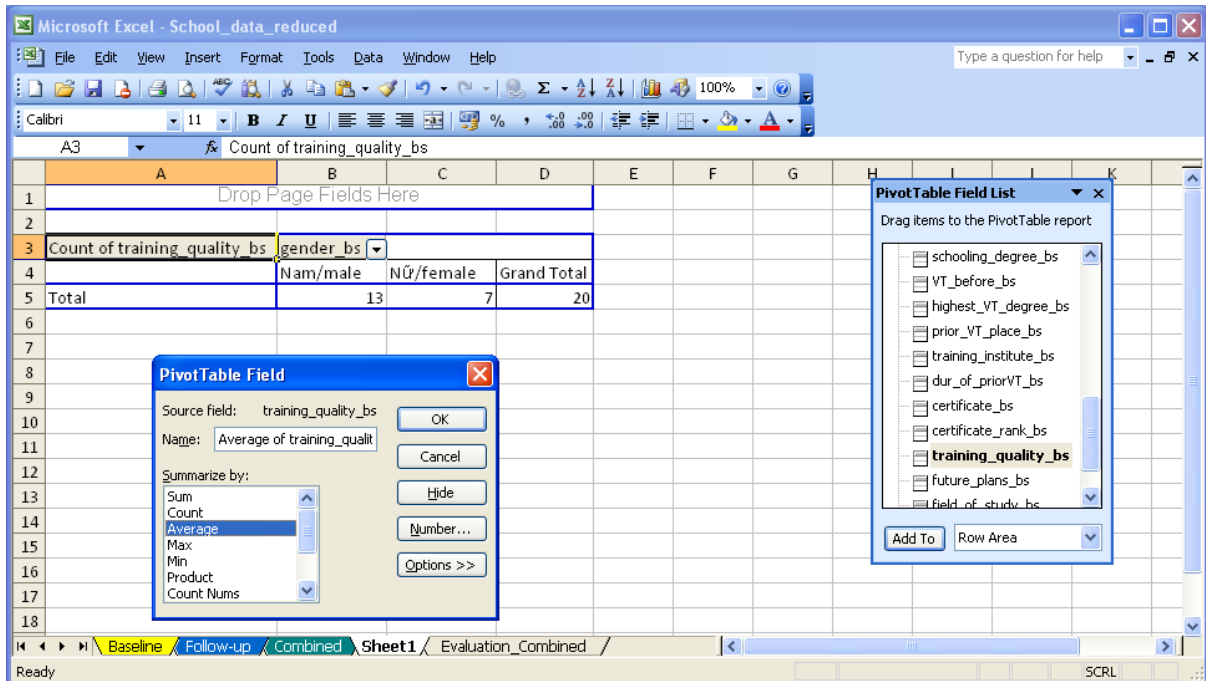
In order to change the statistics from count to mean, please do a right click on the variable name in the left upper part of the table and select “Field Settings...” as in the previous example.



Screenshot 101: Creating a PivotTable – selecting statistics

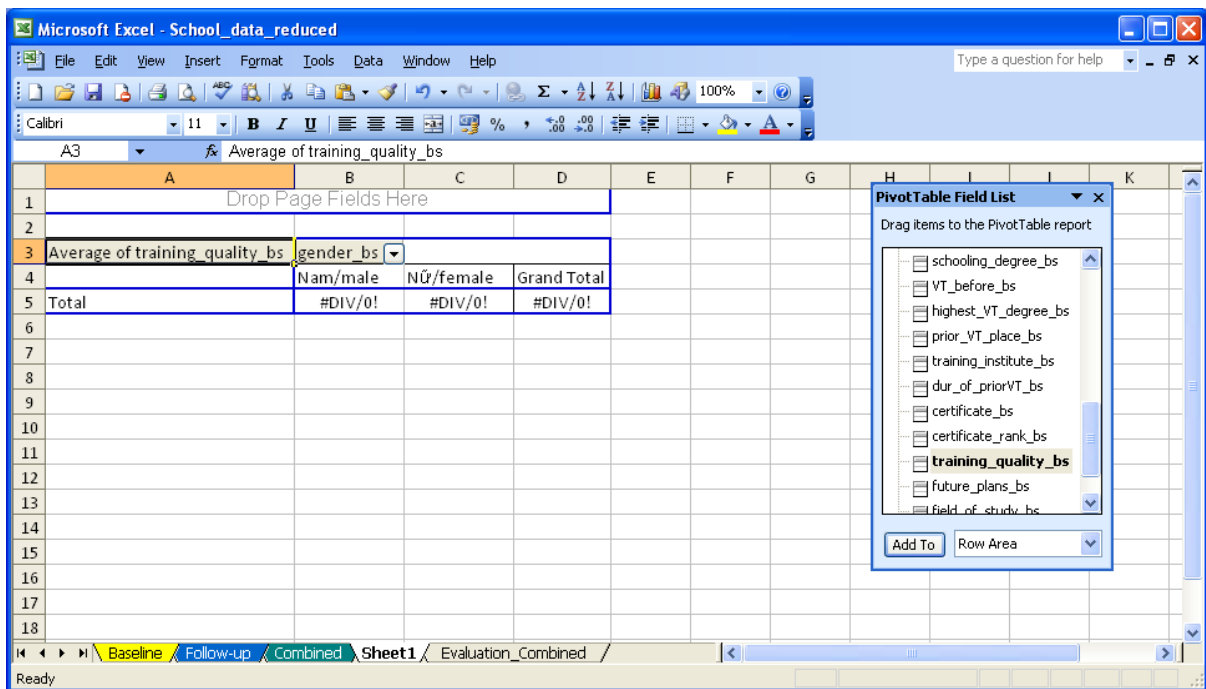
Survey Data Management

A dialogue box will appear where you can select “Average” as a statistic. Please proceed by clicking “OK”.



Screenshot 102: Creating a PivotTable – selecting statistics

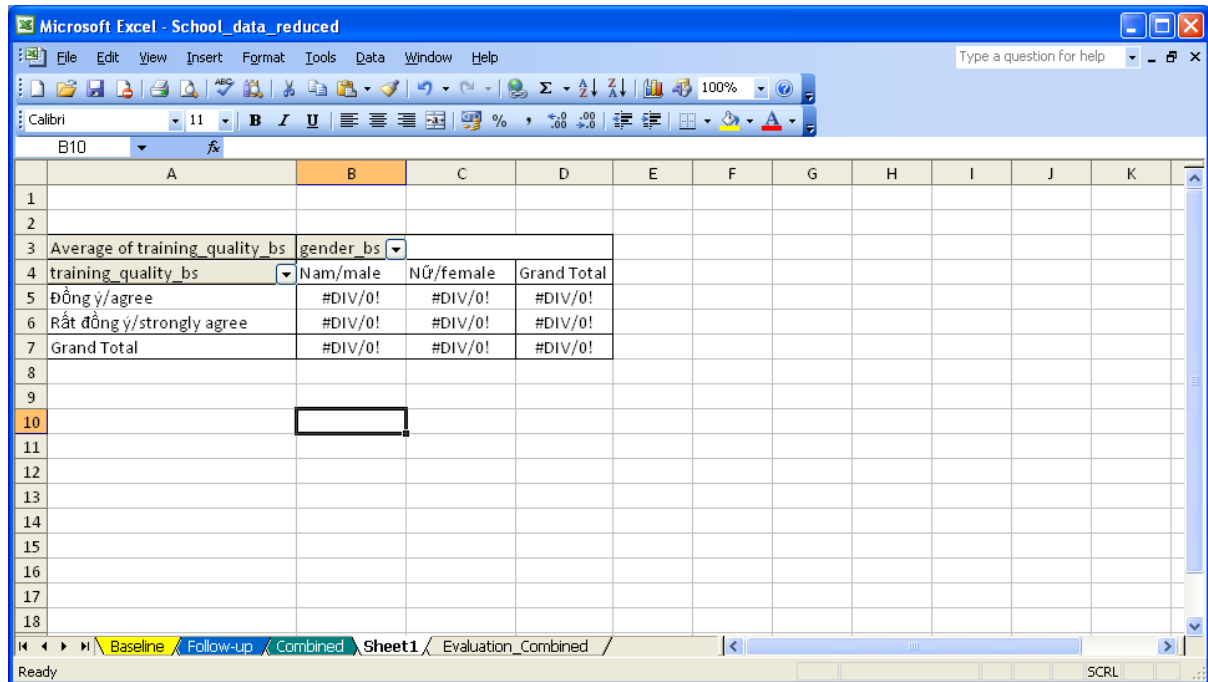
As you can see, the table displays an error, which has its roots in the coding of the variable.



Screenshot 103: Creating a PivotTable – Problems

The coding of a variable can be easily checked by adding the variable to the row field of the table. The recoding will be explained in the next chapter.

Survey Data Management

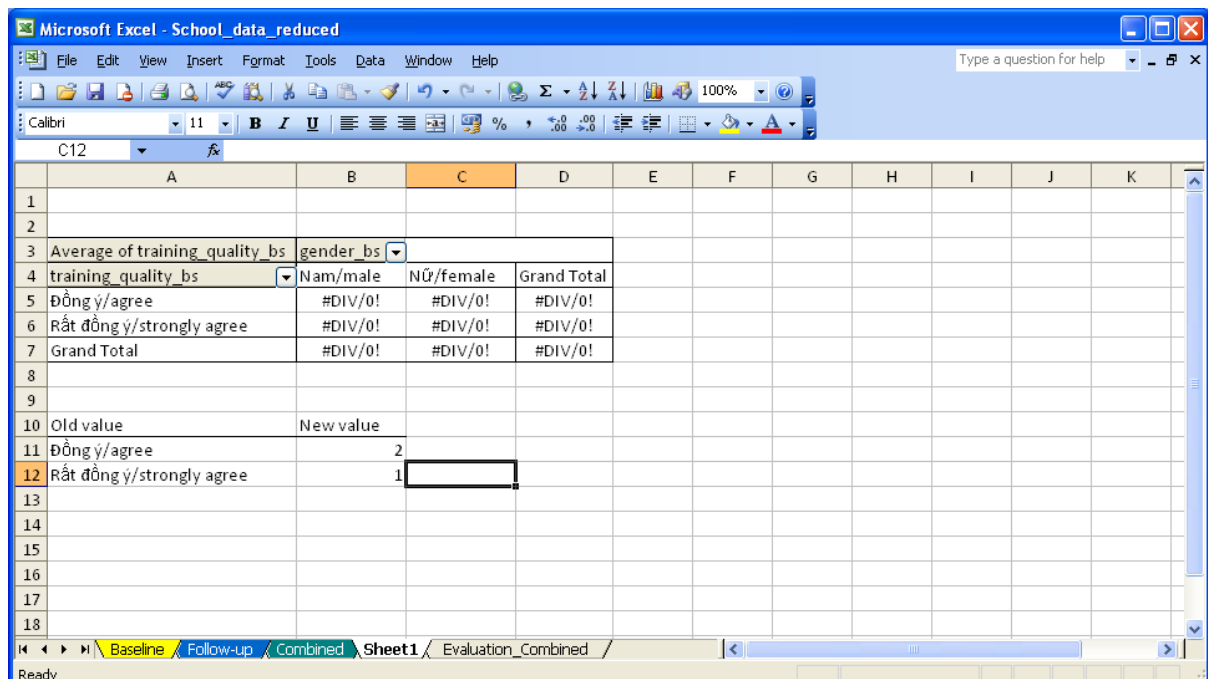


Screenshot 104: Creating a PivotTable – Problems

9.2 Recoding of data

As we have seen in the previous example, it is in some cases necessary to recode the data to compute certain statistics. We will recode the variable `training_quality_bs` in the following example in order to be able to create a table with average values of this variable.

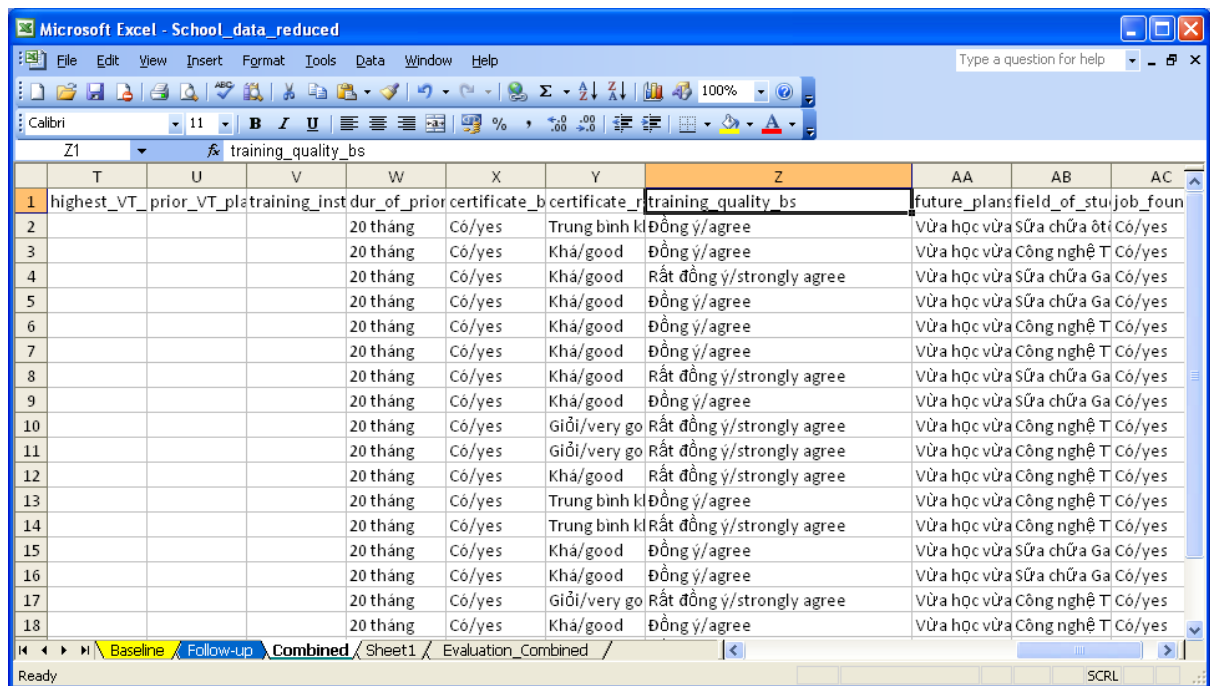
To prepare the recoding, we first have to create a table that contains the original values of the variable and the new values that will be used for recoding. This can be done by copying the values from the PivotTable. To copy them, select the cells and press the buttons “Ctrl” and “c” on your keyboard simultaneously. You can then paste these values into empty cells below the PivotTable.



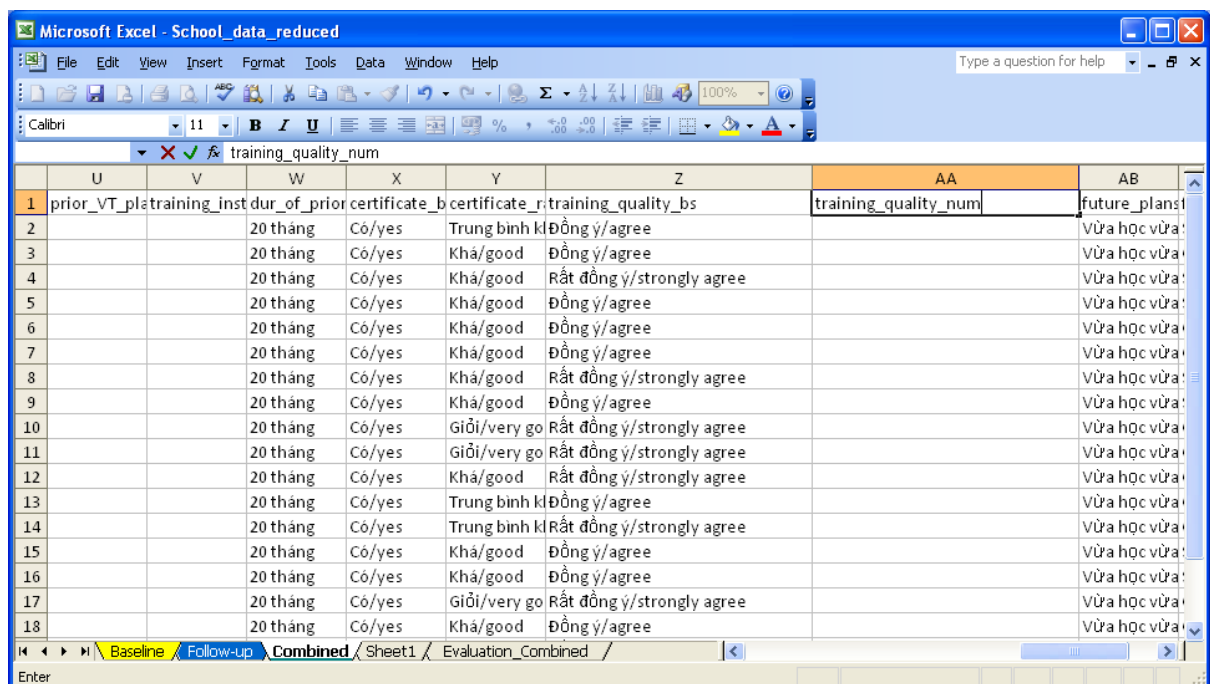
Screenshot 105: Recoding variables

On the right hand side of these original values, we will write the new values needed for recoding. To have the new values on the right hand side is important, because we will use the VLOOKUP function from the last chapter for this task. We recode “strongly agree” with 1, as it is the best category. For the second best “agree” we assign a 2.

After you have created the table for the recoding, please click on the green “Combined” spreadsheet in the Excel worksheet and go to the variable “training_quality_bs”. There, you add a new column and write a new variable name into the first row. As we will recode a text variable to a numeric one, we will call the new variable “training_quality_num” in this example.

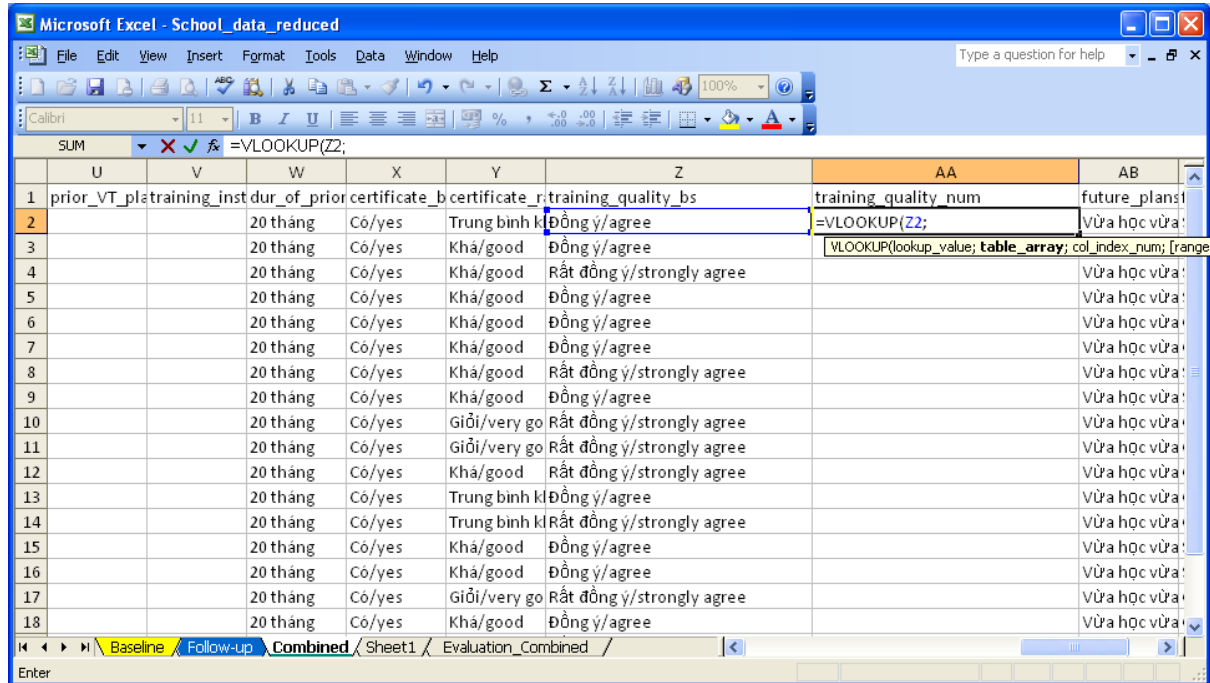


Screenshot 106: Recoding variables



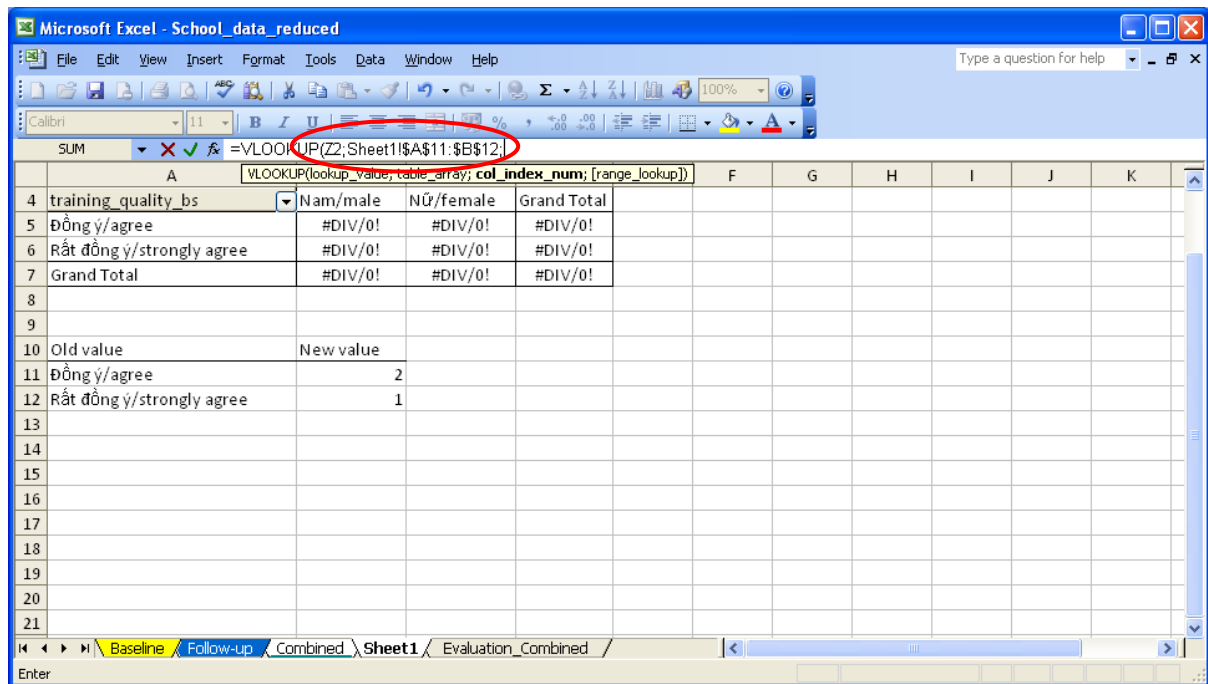
Screenshot 107: Recoding variables – inserting a new column

Similarly to the case with the new identifier, we can now ask Excel to look up the original coding of the variable and return the new coding using the VLOOKUP function. We start typing “=VLOOKUP(Z2;” into the first cell. Since the table, where Excel has to look up this value, is in another spreadsheet, we switch the spreadsheet by clicking on “Sheet1”.



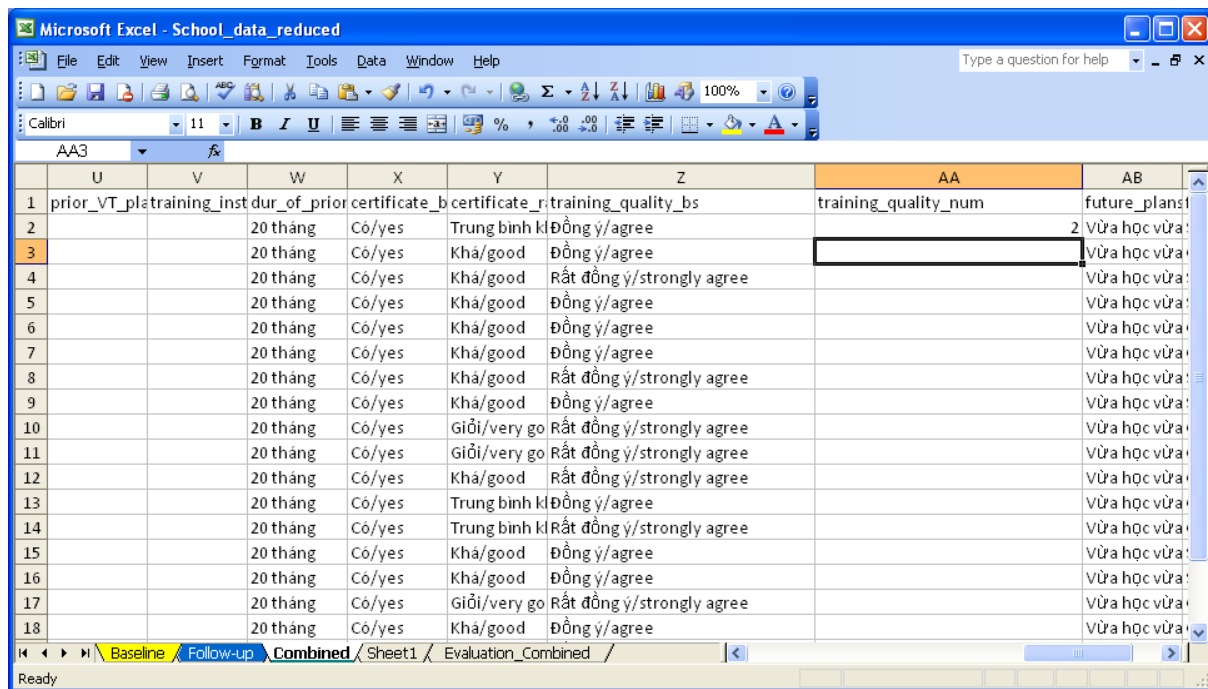
Screenshot 108: Recoding variables using VLOOKUP

Now you can select the table for the recoding with your mouse. Make sure that you enter the dollar signs for making the reference to the selected table fix before switching back to the original worksheet. In case you insert the dollar signs using the key “F4” on your keyboard, Excel will switch back to the original spreadsheet automatically. In case you insert them by hand, you continue working in the spreadsheet “Sheet1” after you typed the semicolon. The formula should now look like “=VLOOKUP(Z2;Sheet1!\$A\$11:\$B\$12;”



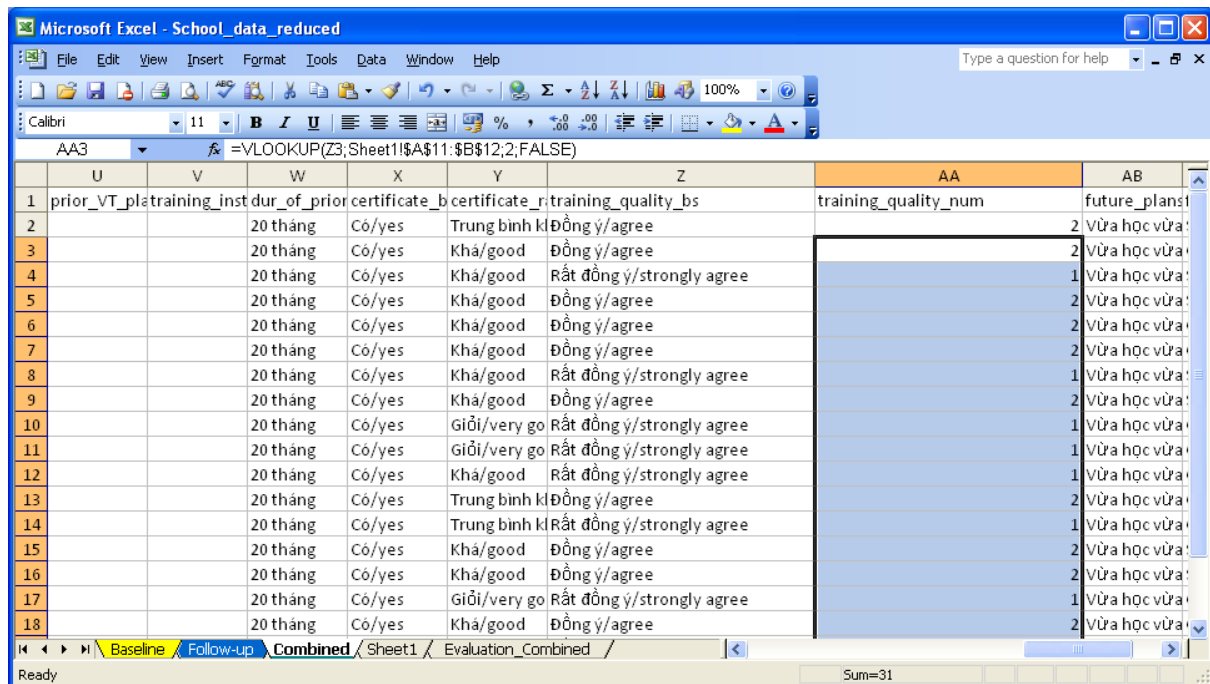
Screenshot 109: Recoding variables using VLOOKUP

In the next step, we finish the formula by telling Excel to return the second column of the recoding table and by requiring that only exact matches are to be made. The complete formula should look like this: “=VLOOKUP(Z2;Sheet1!\$A\$11:\$B\$12;2;FALSE)”. When you press enter, Excel should return “2” in the first cell.



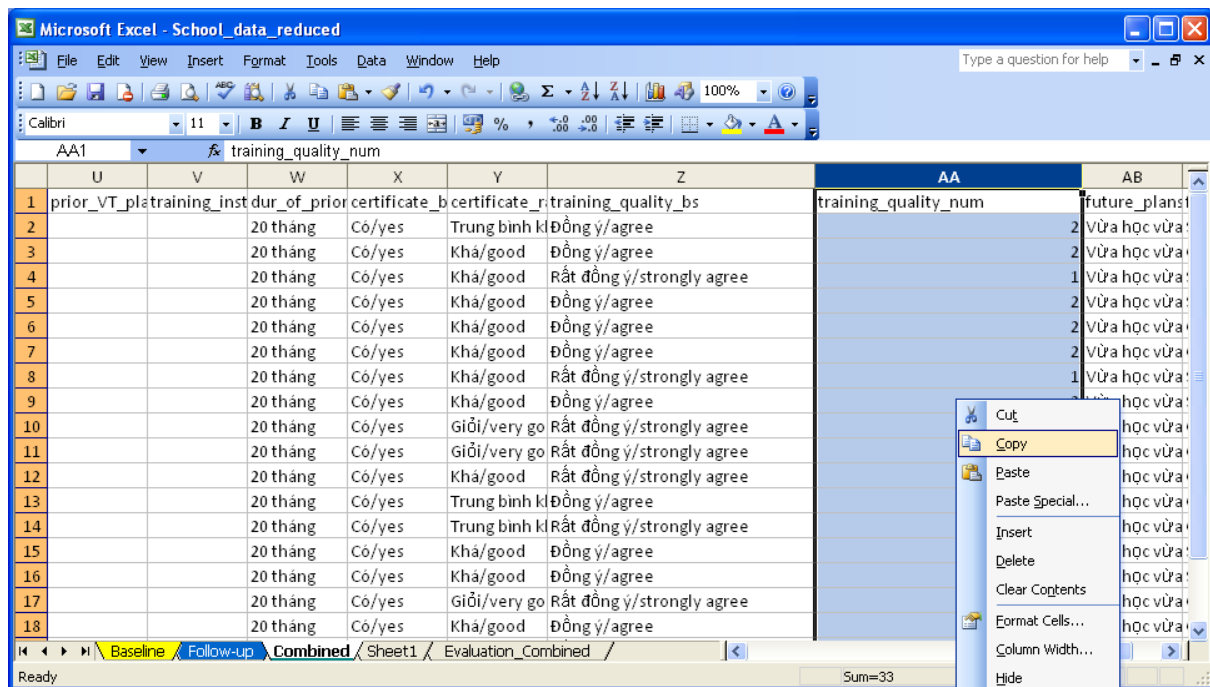
Screenshot 110: Recoding variables using VLOOKUP

You can now copy the cell with the VLOOKUP-formula to all other cells and VLOOKUP will automatically recode the other entries too.

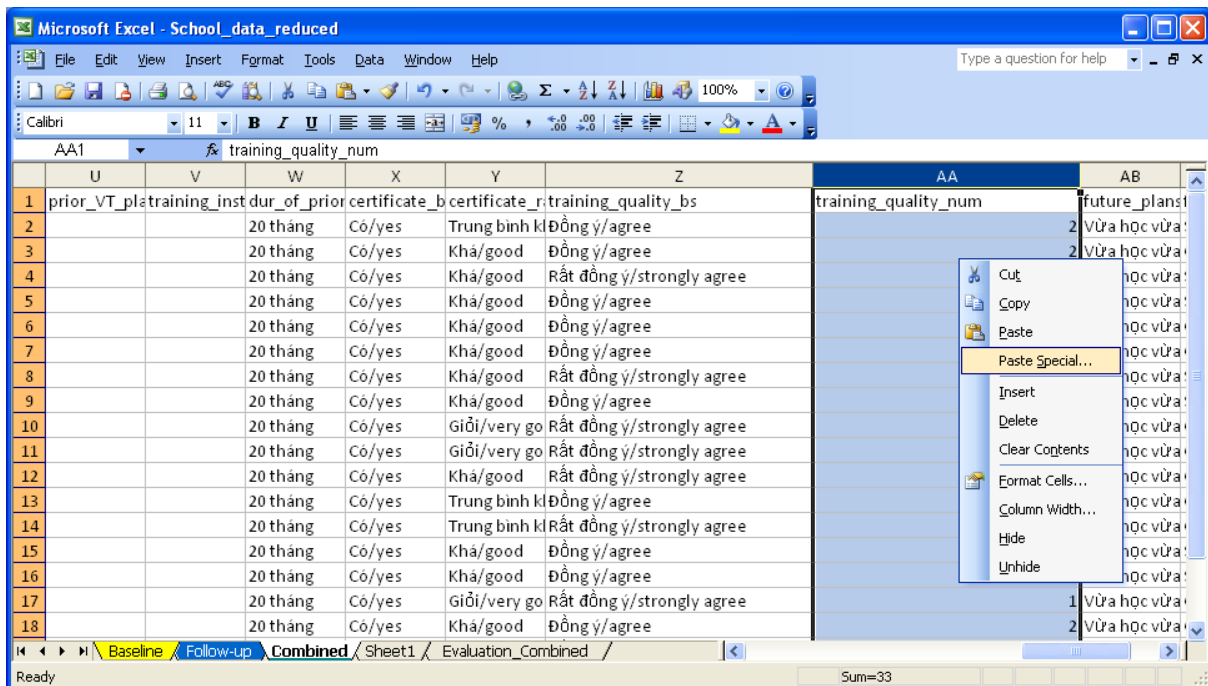


Screenshot 111: Recoding variables using VLOOKUP

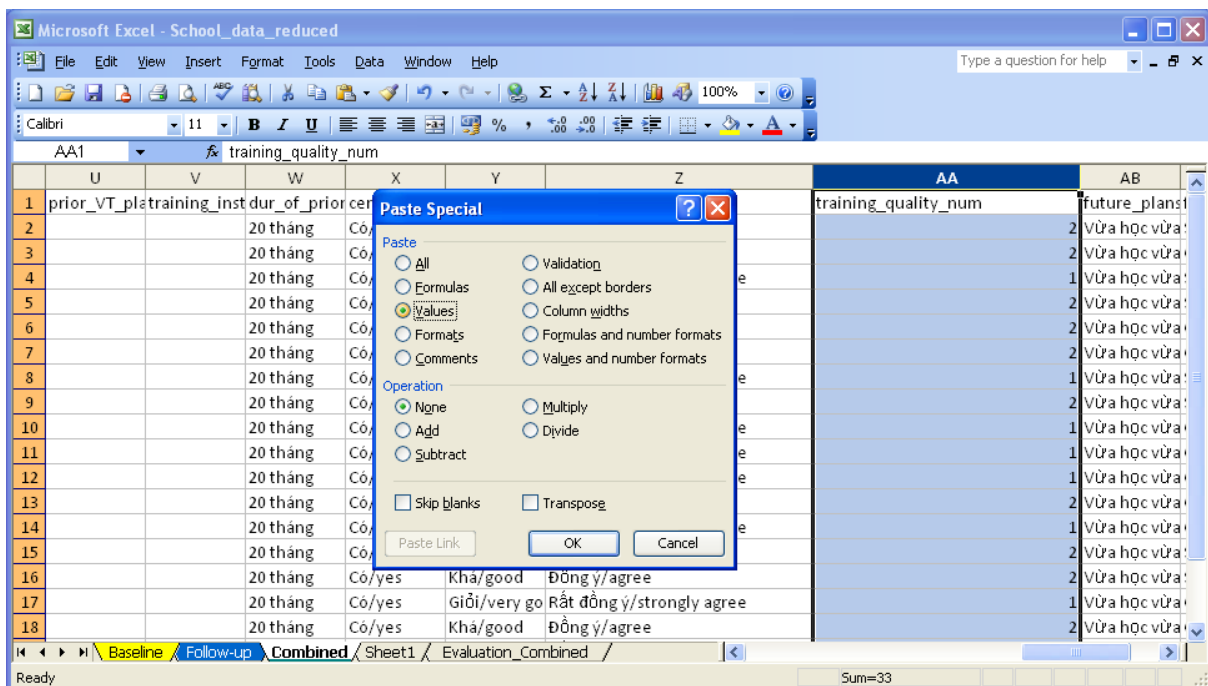
Before we move on to the next chapter, we should make sure, that the newly recoded variable will not change when we remove or alter the recoding table. Therefore, please copy the values of the recoded variable and follow the “Paste Special...” procedure to remove the underlying formula.



Screenshot 112: Recoding variables – removing underlying formulas



Screenshot 113: Recoding variables – removing underlying formulas

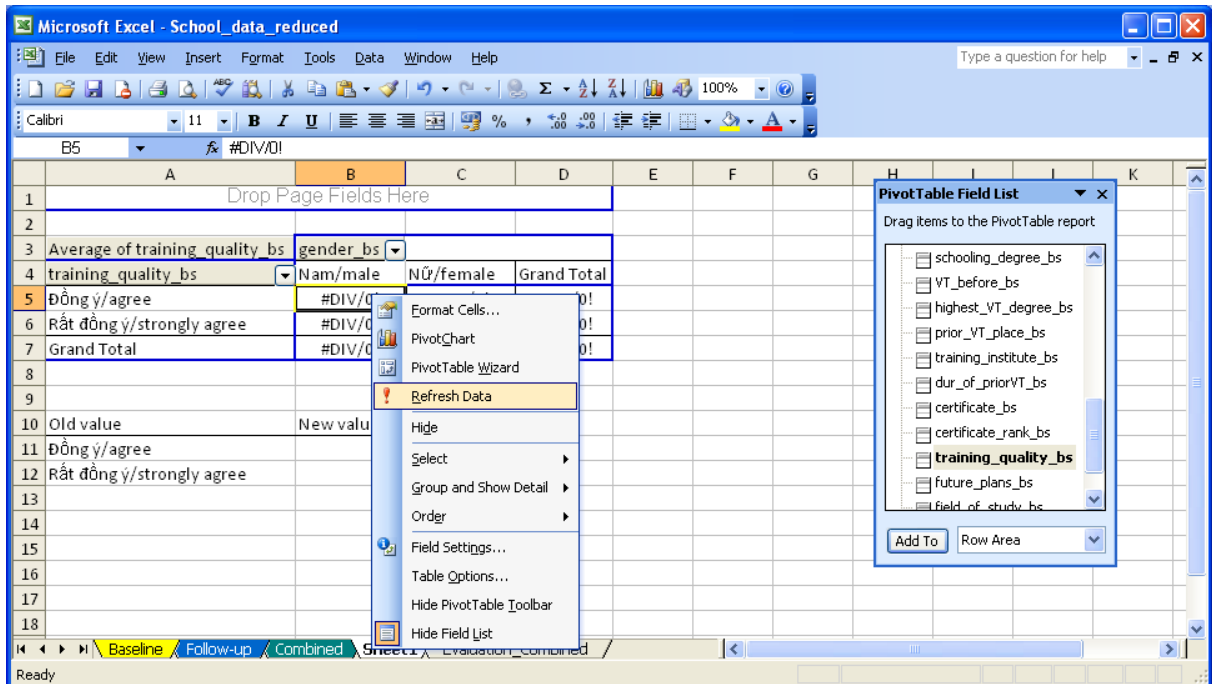


Screenshot 114: Recoding variables – removing underlying formulas

9.3 Creating averages and other statistical information

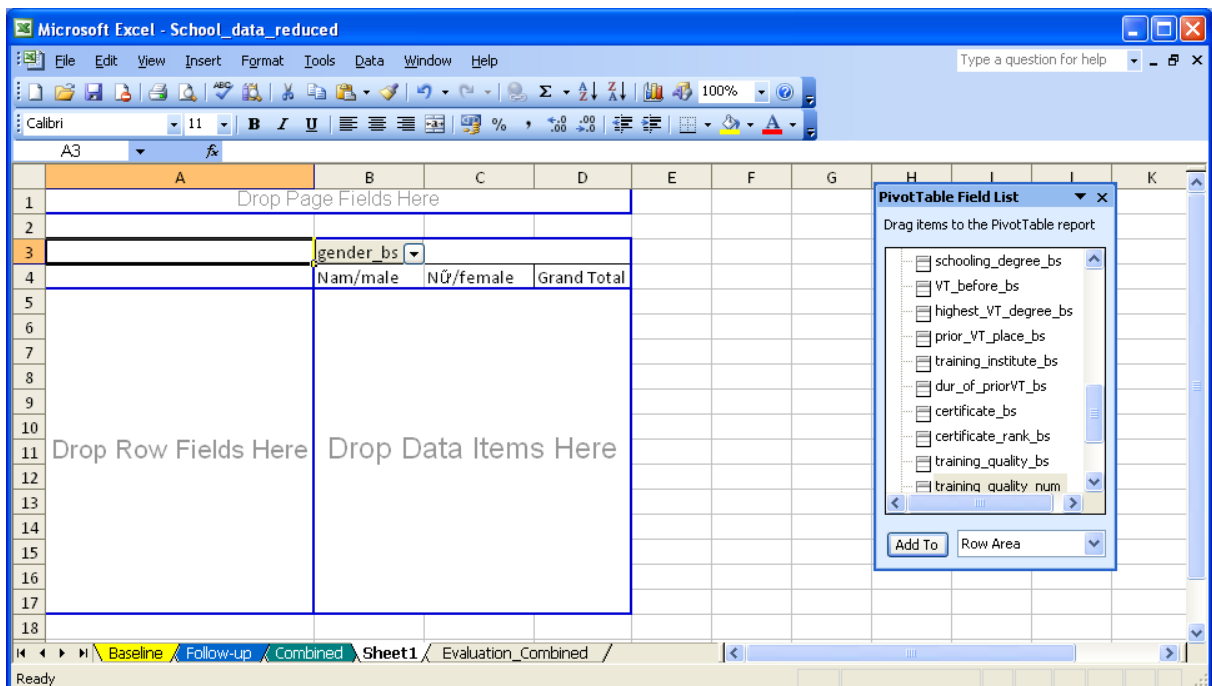
After we have done the recoding, let us now return to the PivotTable in spreadsheet “Sheet1”, where we had the problem with the coding of the training quality variable. As the new variable is not in the variable list of the PivotTable, we first have to refresh the list. Perform a right click on any part of the PivotTable and select “Refresh Data”.

Survey Data Management



Screenshot 115: Creating PivotTables

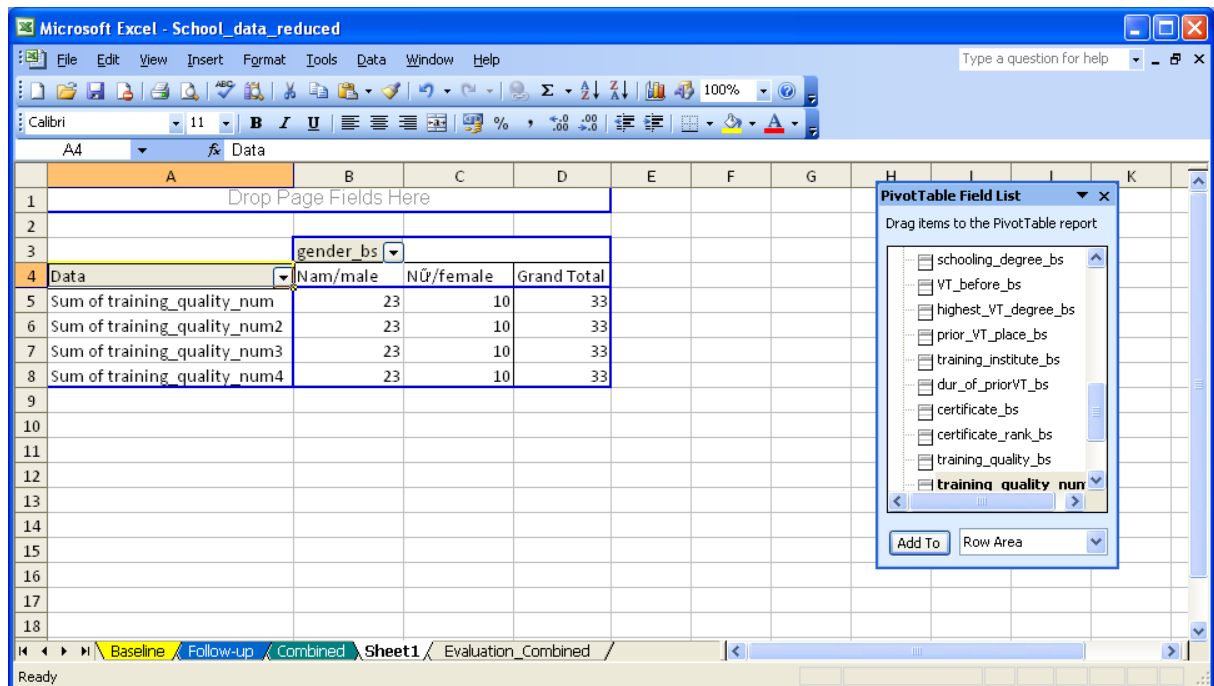
The new variable should now show up in the list. You can exchange the old with the new training quality variable by selecting the old training variable in the table and dragging it out of the table.



Screenshot 116: Creating PivotTables

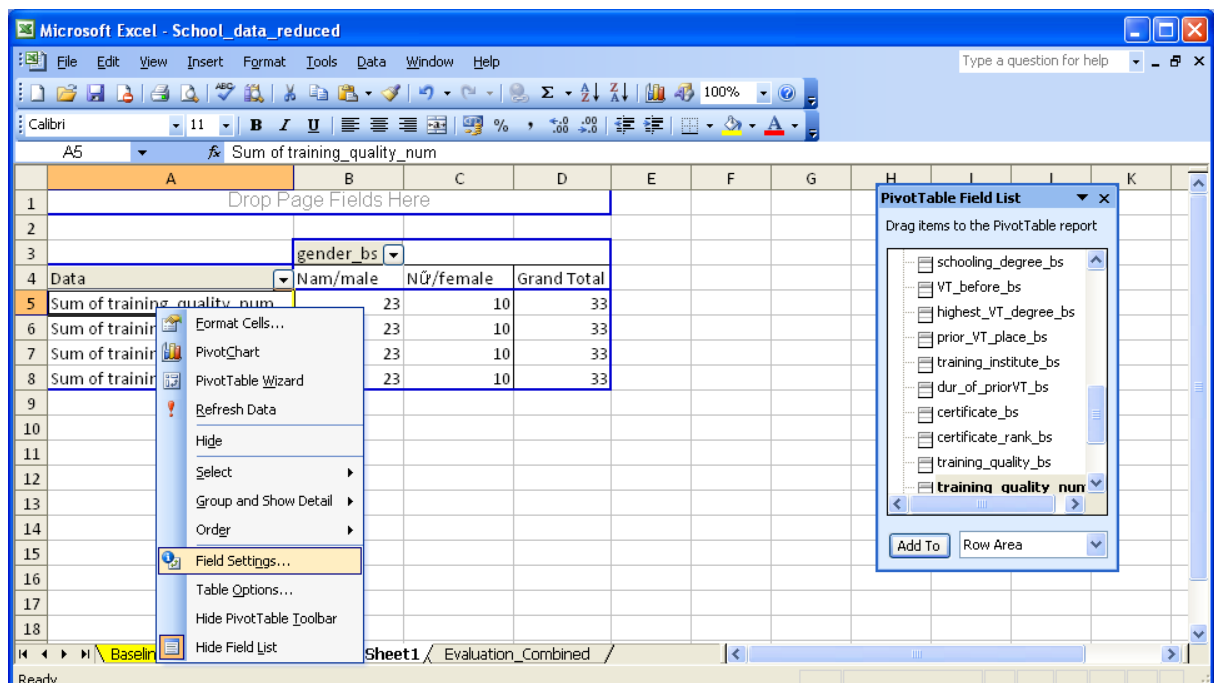
After the old variable is removed, please drag and drop the new variable “training_quality_num” four times into the data field of the table.

Survey Data Management



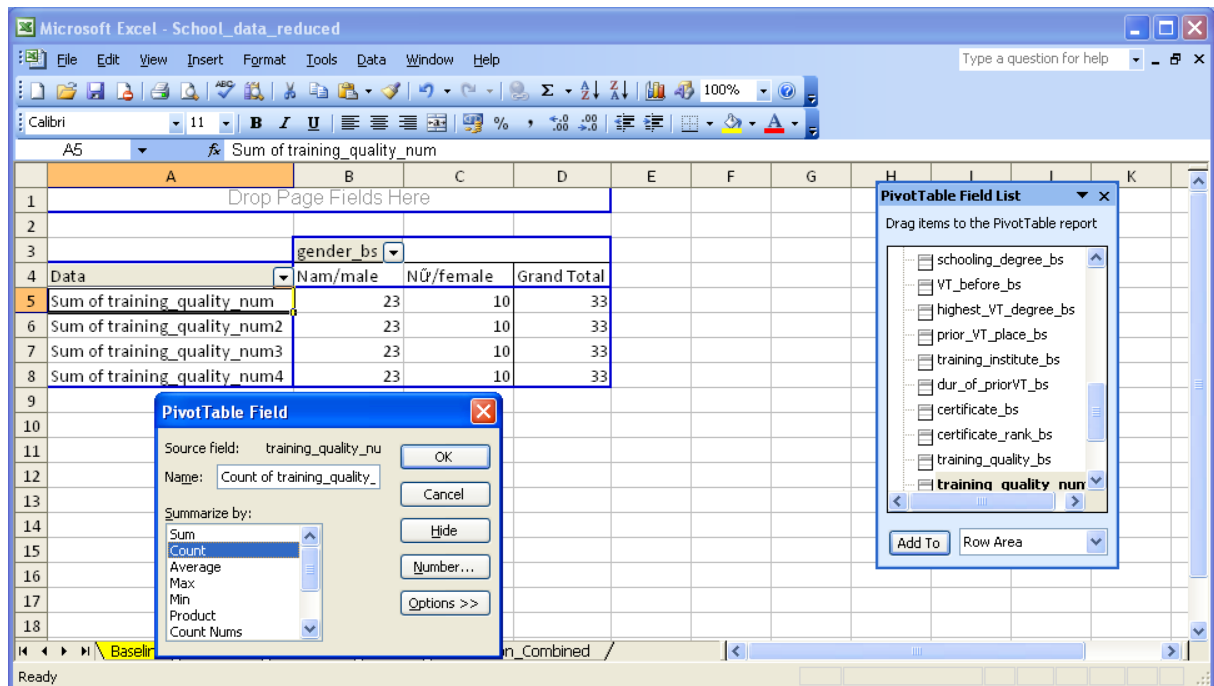
Screenshot 117: Creating PivotTables

This action should result in a table similar to the one above. Now you can change the statistics that are displayed for males and females by doing a right click on each of the four variables in the table and selecting “Field Settings...”. In the field settings, you can now select all relevant statistics for each of the variables separately. As meaningful statistics you could select “Count”, “Average”, “Min” and “Max”.



Screenshot 118: Creating PivotTables – selecting statistics

Survey Data Management



Microsoft Excel - School_data_reduced

Drop Page Fields Here

	gender_bs		
Data	Nam/male	NÜ/female	Grand Total
Sum of training_quality_num	23	10	33
Sum of training_quality_num2	23	10	33
Sum of training_quality_num3	23	10	33
Sum of training_quality_num4	23	10	33

PivotTable Field List

- schooling_degree_bs
- VT_before_bs
- highest_VT_degree_bs
- prior_VT_place_bs
- training_institute_bs
- dur_of_priorVT_bs
- certificate_bs
- certificate_rank_bs
- training_quality_bs
- training_quality_num

PivotTable Field

Source field: training_quality_num

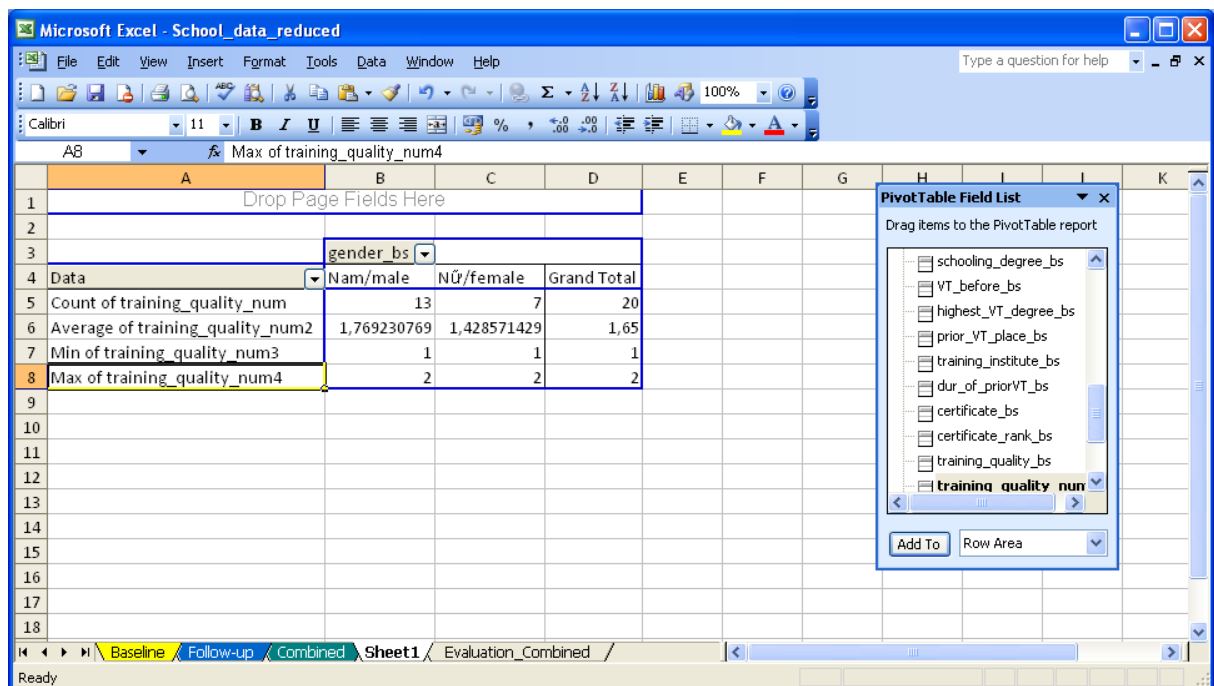
Name: Count of training_quality_num

Summarize by:

- Sum
- Count
- Average
- Max
- Min
- Product
- Count Nums

Screenshot 119: Creating PivotTables – selecting statistics

After selecting the suggested statistics, the final table should look similar to the table in screenshot 120.



Microsoft Excel - School_data_reduced

Drop Page Fields Here

	gender_bs		
Data	Nam/male	NÜ/female	Grand Total
Count of training_quality_num	13	7	20
Average of training_quality_num2	1,769230769	1,428571429	1,65
Min of training_quality_num3	1	1	1
Max of training_quality_num4	2	2	2

PivotTable Field List

- schooling_degree_bs
- VT_before_bs
- highest_VT_degree_bs
- prior_VT_place_bs
- training_institute_bs
- dur_of_priorVT_bs
- certificate_bs
- certificate_rank_bs
- training_quality_bs
- training_quality_num

Screenshot 120: Creating PivotTables – selecting statistics

Before applying the discussed statistical methods, the question to be answered by the statistic has to be clearly defined. Depending on the question, some statistical methods will turn out to be more adequate than others. For example, let's assume that you would like to know whether grades differ by course of study. One could do a simple cross tabulation of grades and courses of study, but such a table is likely to be large and confusing. Instead, a way to condense the information on grades by means of averages would allow a better

overview. Such a table would be similar to the one in screenshot 120, but with course of study instead of gender and the average of certificate rank instead of the average of training quality.

In summary, the question to be answered by a statistic should always be defined before applying a statistical method.

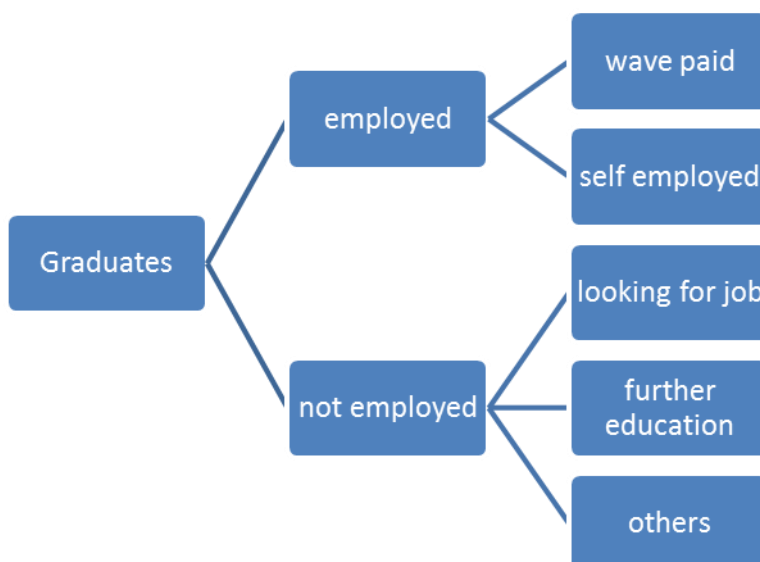
10 Reporting and interpretation of the results

The reporting of the results of a tracer study can be organized in different ways, following the needs of the addressee (e.g. institution's management). Among the different ways of presentation, which range from the mere reporting of tables to a detailed report that incorporates information on the tracer study and on the institutional framework embedding it, we present a middle course. Such a report would mention the data collection, give an assessment of data-quality, present selected key statistics and evaluate them in the context of targets set at school or other levels. The structure of such a report could be as follows: (1) introduction, (2) data collection and data quality, (3) descriptive statistics (4) evaluation of target achievement and (5) conclusion.

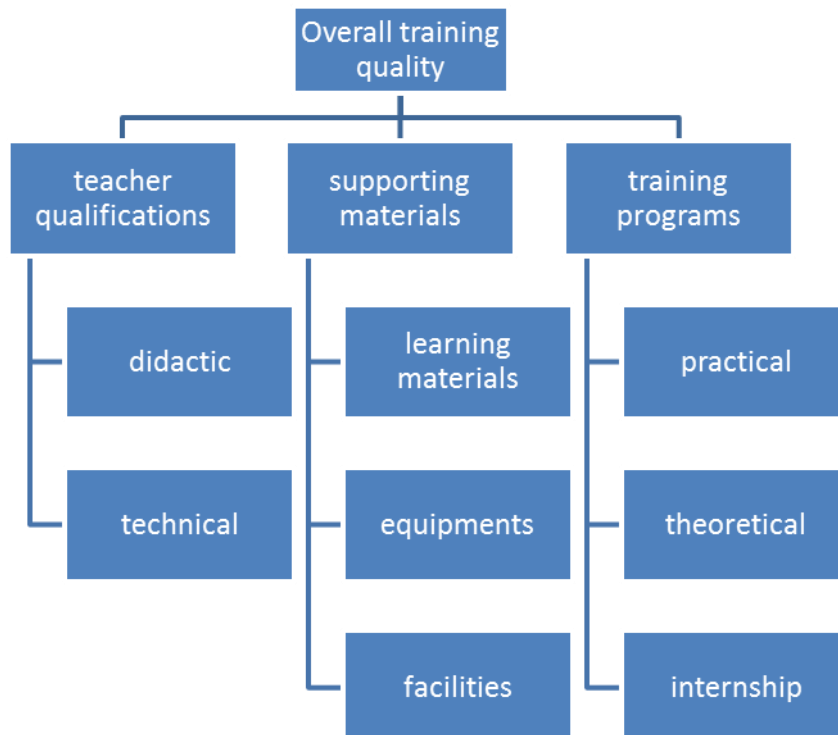
The introduction should sketch the purpose, the content and the structure of the report. It may incorporate also some broader information on the environment the study was implemented in.

The chapter on data-collection and data-quality may comprise information on the sampling procedure, the mode of data-collection (telephone, mail, etc.) and the point in time and duration of data-collection. Any difficulties that were arising during collection should also be mentioned here. These difficulties may be closely connected to data problems, such as the frequent misunderstanding or the low response rate to a question.

The descriptive statistics presented should incorporate demographic information (e.g. gender and age) and information on the distribution of students across the fields of specialization. Other descriptive statistics, such as the labor market status of graduates should also be presented and discussed on the background of the regional labor market. The descriptive statistics on the labor market status can be grouped into different categories depending on the outcome of interest:



An equally important part of this chapter should be a section dealing with the training evaluation. Statistics on training evaluation may be split up by gender, by occupation or other distinguishing factors of importance. Within the scope of the Tracer study, evaluation of different aspects of training quality can be drawn from statistics such as:



On the background of these statistics, the achievement of targets (e.g. percentage of graduates in employment) that were set at school or another institutional level should be evaluated and the report should highlight the main findings of the study. No matter, whenever these targets are met or missed, a detailed analysis of factors supporting and hindering the target achievement should follow.

The major points of the target-achievement evaluation plus the results from the other preceding chapters should be summarized in the conclusion.

Appendix 1 - Baseline questionnaire

Code of trainee:

Address and Logo of the vocational training institute

TVET Monitoring Questionnaire – baseline

Dear graduate,

This questionnaire is the first step of a tracer study that will help to understand the labor market situation of vocational training graduates and will improve VT students learning situation. Our school would like to learn from the experiences you made during your training – we therefore kindly ask you to provide us with some information, which will help us improve the quality of our training.

Thank you very much!

(If not mentioned different please, tick only one box per question)

A. PERSONAL INFORMATION

Full name: Class:

Speciality/Occupation: Faculty:

1. Your gender male female

2. Your date of birth __/__/19__

3. The highest education degree that you achieved (excluding vocational training):
 primary school lower secondary upper secondary
 college university other:

4. Did you receive any vocational training before you enrolled in *(Name of the TVET institute)*?
 no (please continue with question 9) yes

5. Which is the highest vocational training degree that you achieved before you enrolled in *(Name of TVET institute)*?
 elementary VT secondary VT technical college training other:
 certificate of job -guidance training course in secondary school

6. Where did you receive vocational training prior to enrolling in *(Name of the TVET institute)*?
 training at Vocational Training Center secondary VT school
 VT college other training institution.....

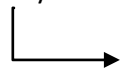
7. Could you please state the name of this training institute?
.....

8. Could you please state the duration of this training? __/__ [Months/Years]

Survey Data Management

9. Regarding your training at (*Name of the TVET institute*): Did you receive a certificate/degree?

yes no not yet



Please state how this certificate/degree was ranked:

pass pass plus good very good excellent

10. How do you assess the training quality of this vocational training institution?

very good good fair not good totally not good

B. FUTURE PLANS

11. What are your plans after graduation?

continue studying continue studying and taking up work
 take up work (please continue with question 13) other:

12. In which field do you want to continue studying?

Qualification level to be reached:

secondary college university other:

13. Have you already found a job (as employee or entrepreneur)?

no (please continue with question 16) yes

14. How adequately did the training at the vocational training institution equip you knowledge and skills for the job's requirements at your current workplace?

very adequate adequate not adequate totally not adequate

15. Please indicate the firm's name, address and sector:

Name of firm:

Address: House No.....Street.....Community.....

District.....City/Province

Sector:

agriculture – forestry - fishery

industry production

construction

trade

services (hotel, restaurant, bank ...)

state office (ministry, province committee,...)

education/ training

others, please specify: _____

C. CONTACT INFORMATION

16. Please give us your contact information in for participating in the tracer study:

a. **Your current address:** House

number.....Street.....

Community.....District.....City/Province.....

Landline: ____/____

Survey Data Management

Mobile phone: ____/____

Email: _____

- b. Your parent/friends address:** House number.....Street.....
Community.....District.....City/ Province
- Landline: ____/____
- Mobile phone: ____/____

All given answers will be kept confidential, it will not be possible for persons or institutions other than the school that sent you this questionnaire to identify whether you participated in the survey or which answers you gave. Any reference that could be used to identify you, such as your name or your address, will be removed from the dataset at the end of this tracing period.

Appendix 2 - Follow-up questionnaire

Please put the code of graduate before sending

Address and Logo of the TVET institute

Dear Graduate,

We would be most grateful if you could kindly fill out the following questionnaire about your job-experience since you graduated from [Name of the TVET institute]. The provided information will be of great importance for our TVET institute to improve the training quality.

If you have any question or need any explanation to answer this questionnaire, please feel free to contact us:

Contact person:

Name:

Phone number:

Email:

Thank you very much

(Please tick only one box per question)

A. Personal information

1. Your gender male female
2. Your date of birth ___/___/19___
3. Regarding your training at [name of the TVET institute], please state how your certificate/degree was ranked:
 pass pass plus good very good excellent

B. Labor market experience and education

4. Please indicate your current labor market status:
 Wage paid self employed not employed -> continue with **question 15**
5. How many hours per week do you work on average? _____ hours
6. How long did it take you to start your current (self) employment after graduation?
 less than 1 month 1 - 3 months more than 3 months
7. How did you find this job?
 placement by vocational training institute personal contacts (family, friends)
 placement by a job agency job advertisement in the internet
 job advertisement in newspaper / radio / TV job fare
 other, please specify: _____
8. The training that you received from [name of the TVET institute] provided the skills and knowledge required in your actual job?
 strongly agree agree somewhat agree disagree strongly disagree
9. How is your job related to the training you received at [name of the TVET institute]?
 closely related somewhat related not related

Survey Data Management

10. Please state your monthly salary. In case you are self-employed, please state the monthly income that is generated by your self-employment. Write either the exact amount or tick the appropriate category (VND).

Exact amount: _____ VND

Category: <1 million 1-2 million 2-4 million 4-8 million >8 million

11. How many employees work in the enterprise you are currently employed in? In case of self-employment, please state the total number of employees in your business including yourself.

1 2-10 10-200 200-300 more than 300

12. Did you receive any on-the-job-training since you started working in this firm?

Yes No (please continue with question 12)

→ Please state the total length of the on-the-job-training you received in this firm?
< 2 weeks 2-5 weeks 1-2 months 2 – 5 months > 5 months

13. Assume that your firm's hierarchy / positions range from 1, which is the lowest possible position in your firm, to 10, the highest position. Please indicate where you locate your job position in this hierarchy by crossing one of the numbers:

1 2 3 4 5 6 7 8 9 10

14. Please indicate name, address and phone number of the company you work in:

Company name: _____

Address: _____

Tel: _____ Email/website: _____

Please indicate the sector of the company:

- | | |
|---|--|
| <input type="checkbox"/> agriculture, forestry, fishery | <input type="checkbox"/> construction |
| <input type="checkbox"/> industry production | <input type="checkbox"/> services (hotel, restaurant, bank ...) |
| <input type="checkbox"/> education/ training | <input type="checkbox"/> state office (ministry, province committee,...) |
| <input type="checkbox"/> trade | <input type="checkbox"/> others, please specify: _____ |

Please answer this section (from Qs. 15 to Qs. 17) if you are currently NOT employed.

15. What is the reason for your non-employment?

- could not find a job -----=> continue with **question16**
 further education ----- => continue with **section C**
 other reason (please specify):-----=> continue with **section D**

16. For which jobs did you apply?

- Jobs related to my trained occupation
 Jobs unrelated to my trained occupation

17. In your opinion, what were the reasons that you did not find a job?

(you can tick more than one option)

- training did not fit job needs lack of foreign language skills
 lack of computer skills lack of working experience
 lack of information about the job
 other, please specify: _____

C. Further education

18. Are you currently in further education?

Yes No (please continue with **section D**)

→ If yes, please indicate the name of the training institution and the kind of further education you are currently participating in:

Survey Data Management

Name of training institution: _____

Name of the course: _____

Level of the course: university college secondary other

D. Training evaluation

How much do you agree to the following statements regarding the training at [**Name of the TVET institute**]?

19. The teaching staff of the vocational training institute was qualified and up to the training requirements
 strongly agree agree somewhat agree disagree strongly disagree

20. The curriculum and subject syllabi are updated to be relevant to practical relevance
 strongly agree agree somewhat agree disagree strongly disagree

21. Please consider theoretical and practical parts of the training course separately, how well did they prepare you for working life?
Practical courses: very good good average bad very bad
Theoretical courses: very good good average bad very bad

22. In your opinion, where would you see highest need of improvement?
 learning materials (books, etc.) teachers' didactic qualification
 equipments (machines, tools, etc.) teachers' specialized theoretical knowledge in his/her field
 facilities (class rooms, library etc.) teachers' practical skill in his/her field
 internship / practical in-company training

23. Overall evaluation: The vocational training institute is a reputable training institution
 strongly agree agree somewhat agree disagree strongly disagree

24. Overall evaluation: I am satisfied with the quality of training that I received in the vocational training institute
 strongly agree agree somewhat agree disagree strongly disagree

Data-protection statement:

All given answers will be kept confidential, it will not be possible for persons or institutions other than the school that sent you this questionnaire to identify whether you participated in the survey or which answers you gave. Any reference that could be used to identify you, such as your name or your address, will be removed from the dataset at the end of this tracing period (approx. after one year).

Programme Reform of TVET in Viet Nam

