Analyzing (Online) Survey Data with Closed and Open-Ended Questions

16

Surveys often contain closed questions with predefined responses and open questions for free text responses to collect both quantitative and qualitative data. The result of a survey like this is almost always a data matrix in a "Cases × Questions" format, which provides one row for each respondent and contains at least one column per question. The issue then arises as to how a data matrix of this kind, which could easily contain several hundred or thousand cases in the rows and a few hundred columns, may be imported—and meaningfully analyzed—in MAXQDA. Since many surveys are conducted online these days, you might also want to import surveys from tools such as LimeSurvey, Qualtrics, or SurveyMonkey. Whether the survey was conducted online or on paper, the great advantage of analyzing survey data with MAXQDA is that qualitative data imported as text and quantitative data imported as case variables can not only be separated but also analyzed in an integrated manner, that is, in the form of a mixed methods analysis.

In This Chapter

- Importing and automatically coding survey data
- Tips for importing data from online tools such as LimeSurvey and SurveyMonkey
- Analysis strategies for responses to open questions
- Analysis strategies for the integration of qualitative and quantitative data

Preparing and Importing Survey Data

A data matrix with the results of a survey can be imported into MAXQDA from the Excel format, which is offered as an export option by many online data collection tools and statistics programs. Before importing an Excel file with survey results, you

U. Kuckartz, S. Rädiker, *Analyzing Qualitative Data with MAXQDA*,
https://doi.org/10.1007/978-3-030-15671-8_16

should check carefully whether any responses need to be removed and whether the remaining responses are plausible. This especially applies to online surveys where the data has not been entered by the analysts. For online surveys, test runs of the finished questionnaire are often carried out, the responses of which must be removed, along with any duplicate entries that may have resulted from multiple submissions by the same person. Valuable information about entries that need to be removed can often be found in the time the responses were submitted, which is usually recorded in a separate column by the online tool in question. Moreover, it is particularly important in the case of online surveys to decide on the extent to which data from people who did not complete the survey in its entirety should be taken into account. The widely used online tool "LimeSurvey" provides valuable information in this regard, i.e., on which page a person ended the survey and the total time taken to complete it. Plausibility checks primarily involve checking whether a person entered made-up information in order to participate in a promised prize draw or because their motivation declined toward the end of the survey. You can often recognize fake data where "no answer" responses have been provided for multiple questions, where the same response option has been selected for every matrix question, or where open questions were not answered at all or only very briefly.

Apart from cleaning up the data records (rows) and variables (columns), there are usually no further precautions to be taken for the import into MAXQDA. However, the import does require that there be a column in the data that can serve as document name in the MAXQDA project. Furthermore, you should always use a unique identifier for each individual case, if possible, to ensure that the correct allocation of cases to respondents can be guaranteed at all times both within MAXQDA and for later exports of the data. This could be, for example, a case ID that was assigned automatically by the online survey tool or a unique name that you assigned to the case yourself. Table 16.1 shows an excerpt from the data matrix of an online survey of students at Marburg University regarding their quality of life; the matrix comprises a total of 1178 cases (rows) and contains 9 columns with answers to open questions, in addition to numerous variables for standardized information.

Table 16.1 Responses to open and standardized questions in a table format

Case ID	What recreational opportunities/facilities do you think are missing? (Open question)	How would you define a good quality of life? (Open question)	Age	Area of residence in Marburg
1007	Good and inexpensive gym near me. Club with good music	When you can find a balance between your responsibilities (studies), socializing (friends), and personal...	27	Marburg district
1008	[empty, because no response]	Time, the space and courage to do things that the person really...	27	Marburg city center
...	...	...	...	...

► **Tip** If the data matrix is available as an SPSS file, you can also save it in an Excel format from within SPSS via ***File*** > ***Save As***. Here you should select the option that the value labels are to be exported instead of the values. This ensures that, for a variable like "Area of residence in Marburg," for example, not only the numbers in the data matrix are included in the file but also the assigned response specifications as text. Unfortunately, SPSS does not export the variable labels themselves to the Excel file but rather the often significantly shortened variable names in the headings, which is why you should then manually replace these with suitable short versions of the questions before importing them into MAXQDA.

You can start the import into MAXQDA by going to ***Import*** > ***Survey Data*** > ***Import Data from Excel Table*** and then selecting the Excel file containing the data matrix. MAXQDA analyzes the file and then opens a dialog box that lets you adjust the import settings (Fig. 16.1).

In the upper area of the dialog box, you can select which columns contain the names of the document groups and documents. It is helpful to select a grouping variable that can form the document groups, within which the documents based on each case can be organized in the "Document System." For the quality of life data set above, for example, you could organize the data according to area of residence.

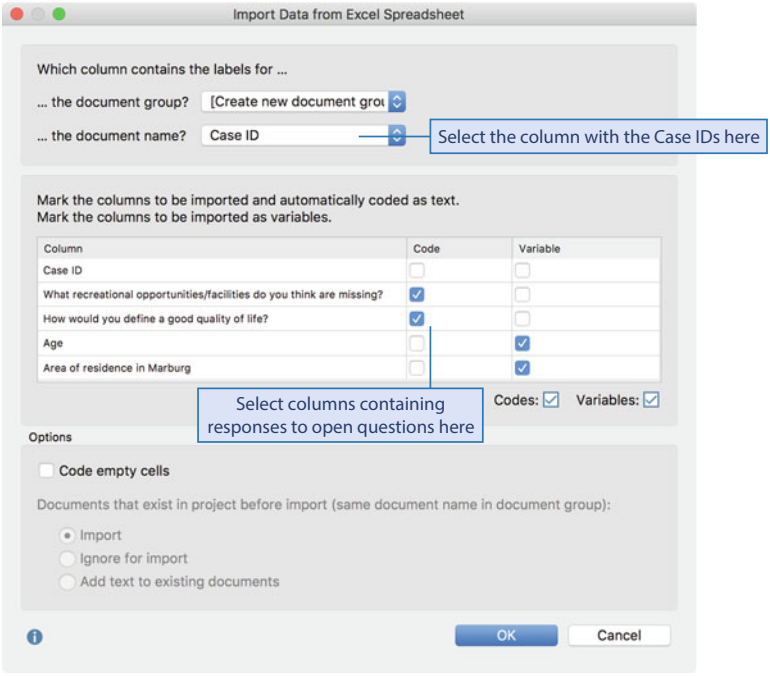

Fig. 16.1 Settings for importing survey data

However, since the standardized data can later be used to compile documents into any number of document sets to conduct group comparisons, it is usually sufficient to accept the default settings and allow MAXQDA to automatically create a single new document group for all the imported documents. That said, for performance reasons ideally document groups should not contain any more than 1000 documents. For the document name, as shown in Fig. 16.1, it is essential that you select the case ID to make sure it can be associated with the original data set. This also ensures that each document name occurs only once.

During the import, MAXQDA automatically creates a separate document for each case, i.e., for each row in the Excel table. The responses to the open questions form the content of the respective case, and they are automatically coded with the respective column heading, usually the shortened version of the question. One possible result is illustrated in Fig. 16.2: in the "Document System," each row of the Excel table has been converted into a separate text document. One of these text documents is open in the "Document Browser," indicating that two open responses have been coded with a question each, for which two codes were automatically added to the "Code System." In addition, the imported variable values are always visible in the Data Editor for Document Variables.

Returning to the import dialog box in Fig. 16.1: in the middle area, MAXQDA lists all the columns from the Excel table. For each column, you have to decide whether the content of the column is to be included in the document as pre-coded text or imported as a document variable. By default, MAXQDA places a check mark in the column "Variable" for all questions, because usually there are more standardized quantitative answers in a data set than responses to open questions. What is important for the import is to change the check mark to the "Code" column for all *open questions*. In addition to the open questions, you can also check the "Code" column for some standardized information, such that the contents of the column are both included in the document as text and imported as variables. This is

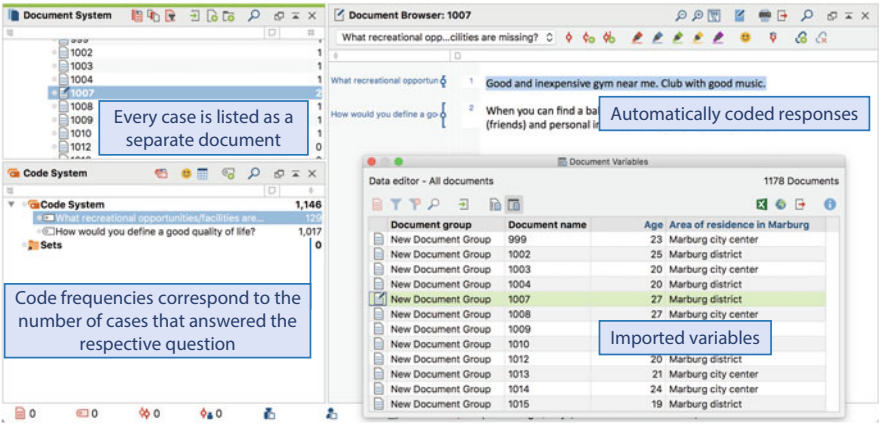

Fig. 16.2 Imported survey data in MAXQDA

especially helpful if you not only want access to the variable information for a document via its name's tooltip or via the Data Editor for Document Variables but want to be able to see within the text itself how old the respondent is, for example, and where he or she lives in Marburg.

The option ***Code empty cells*** should normally remain unchecked, because otherwise it ensures that cells without content are also coded with the respective column header. The great advantage of deselecting this option is that, after you have completed the import, you will be able to see immediately from the code frequencies in the "Code System" in how many cases a code has been assigned, i.e., how many people actually answered each open question. You can see an example of this in the "Code System" in Fig. 16.2, which illustrates the view in MAXQDA just after survey data has been imported: while only 129 people responded to the question about recreational opportunities/facilities, 1017 people responded to the question about quality of life.

The options in the lower area of the dialog box in Fig. 16.1 are intended for studies involving several survey cycles and, among other things, enable you to attach the contents of a data matrix to texts which you already imported at an earlier point in time. Remember that the document group and document name must match in both data sets, which is why the imported data should always contain these two columns to ensure that the documents can be correctly assigned.

The import process is quite fast for 1000–2000 respondents. The precise time required depends on the number of open questions the survey contains and how many variables need to be generated, so you may need to allow for a few minutes waiting time for, say, 5000 cases and 10 open questions; after all, imports of this size may generate up to 50,000 code assignments. There is no limit to the number of cases in MAXQDA, but for technical reasons, it is best to avoid exceeding a total of 200,000 code assignments in MAXQDA projects—a number you should keep in mind when dealing with several thousand cases and multiple open questions.

Importing Survey Data from LimeSurvey

LimeSurvey is a convenient and free open-source software tool for conducting online surveys, which you can easily install on your own web server. This has the great advantage that you can determine the server location yourself and are not dependent on countries with questionable regulations when it comes to data privacy and data security.

Once all the data has been collected in LimeSurvey, you can log in to your LimeSurvey account and save the data as an Excel file via the ***Export*** > ***Export results to application*** button in the "Responses & Statistics" section. You can define numerous settings for the export:

- *Format*: Should be set to Microsoft Excel.
- *Range*: IDs of the cases you want to export—this will usually be all the cases.
- *General* > *Completion state*: Usually all rows and not only the filled-out rows should be exported, because you can easily remove duplicate or implausible data rows in Excel.

- *General* > *Export language*: For surveys conducted in several languages, the export can be restricted to one of the languages used.
- *Headings*: Here simply leave the default settings as they are, because it is usually helpful if the full question text is included in the heading.
- *Responses*: Here we also recommend sticking to the default settings. The complete responses will then be entered in the individual cells, instead of numerical codes that may be difficult to interpret.
- *Columns*: Here you can specify which information you want to export. There is a risk here of clearing your entire previous selection of columns with one wrong click, so it is much easier to export all the columns and later, in Excel, delete those that are not important, such as the response times for individual questions.

Importing Survey Data Directly from SurveyMonkey

SurveyMonkey is currently one of the world's most widely used online survey tools, offering a very large selection of ready-made questions from a variety of subject areas. The company is headquartered in California and offers a free basic plan with limited functionality and a limited number of cases, as well as several fee-based plans with additional question formats and unlimited cases. MAXQDA allows you to import responses directly from a SurveyMonkey survey, saving you the intermediate step of exporting the data into Excel. The import procedure is as follows:

- Step 1: Go to ***Import*** > ***Survey Data*** > ***Import Data from SurveyMonkey***.
- Step 2: The SurveyMonkey website will then open in your web browser, where you will need to log in with your SurveyMonkey account and grant MAXQDA access to the survey responses.
- Step 3: Once you have successfully granted it access, MAXQDA will display an overview of all the surveys in your SurveyMonkey account, from which you can select one by clicking on it. The options also allow a random selection of cases if you only want to analyze a subset of them.
- Step 4: Just as with the import via Excel, you can determine which elements of the data are to be imported as open questions and which ones as document variables. MAXQDA will have already made an automatic selection based on the question types.

► **Please Note** SurveyMonkey's basic plan does not allow you to import data directly into MAXQDA or export it into Excel. This requires a paid plan. You can find an overview of all the plans and their features here: https://www.surveymonkey.com/pricing/details/. The "Extract data" row in the "Partner Integrations and APIs" section tells you whether a given plan lets you import data directly into MAXQDA or not. The "Analysis & Reporting" section lists the export options for each plan.

Analyzing Survey Data

Exploring the Data

Once the survey data has been imported into MAXQDA as described above, you can begin your analysis. Usually you will want to get an overview of the survey data first, both of the scope of the free text responses, which are much shorter in comparison to interview transcripts, and of the standardized variable information available. Kuckartz, Ebert, Rädiker, and Stefer (2009) describe one possible procedure for exploring survey data using the example of an online evaluation and also provide tips on how to use MAXQDA in this context. The procedure they describe for exploring qualitative data makes use of many of MAXQDA's data exploring capabilities, as described in Chap. 5: case summaries in bullet point form can be prepared for a random selection of about 5–10% of cases, taking into account particularly relevant standardized data. To do this, open the List of Document Variables via the ***Variables*** ribbon tab, and choose a selection of variables as tooltip variables so that the information stored in them is displayed in the small preview when you hover your cursor over the document in the "Document System" (see Fig. 16.3; details on tooltip variables can be found in Chap. 10). Case summaries are stored in the document memo and are given a title that briefly describes the case in question and may also contain certain quantitative information, such as age and area of residence. Analysis ideas, theses, hypotheses, and ideas for helpful analysis categories can be noted in a free memo or more easily in a memo at the top level of the "Document System." This exploration can be complemented with a lexical search for words of interest or with word frequency analyses, which you can open in MAXQDA via the word cloud function: simply right-click on a document group to generate a word cloud containing the most frequently-used words within the texts in this group, as described in detail in Chap. 5.

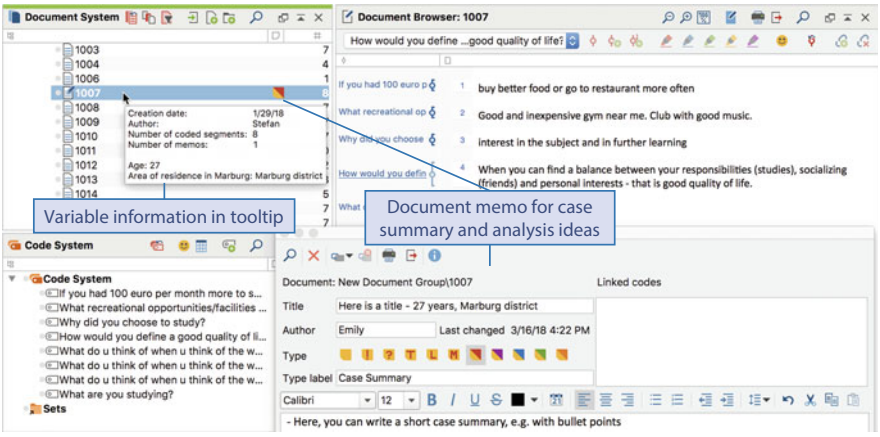

Fig. 16.3 Data exploration screen (with more imported questions than in the previous figure)

To explore the quantitative data from surveys, stored in the form of document variables, once you have completed the import, you can perform a basic count in which frequency tables are generated for each closed question containing their respective responses and absolute as well as percentage frequencies. You can perform a basic count of this kind in MAXQDA via ***Variables*** > ***Document Variable Statistics*** and browse through the frequency tables question by question. A more efficient method is available in MAXQDA's add-on Stats module, in which you can create frequency tables for document variables via ***Descriptive Statistics*** > ***Frequencies***. You can save meaningful groupings of cases based on their variable values in the form of document sets using MAXQDA Stats. But you can only create new memos to record further analysis ideas once you have closed Stats, since all other MAXQDA functions are locked while Stats is open.

Categorize Responses to Open Questions

Once you have explored the survey data, the issue arises as to how you can meaningfully analyze the comparatively short responses to open questions. As a rule, each question is analyzed individually, and the respective response texts are systematized and described using thematic categories such that, among other things, conclusions can be drawn about frequently and less frequently mentioned topics after the analysis has been completed. The categories for the analysis of responses can be formed using both an a priori approach or based directly on the data. For example, forming categories based on the data is the ideal option if you want to avoid pre-categorization and you want to use the responses themselves as your starting point, if the range of possible responses is unknown, or if the questions to be answered in the survey have a rather more exploratory character. A priori category formation, on the other hand, is to be recommended if information on the survey topic is available in advance, if only specific aspects of the responses are of interest and the responses contain information that does not need to be coded, or if the analysis categories are to be derived directly from the survey questions.

MAXQDA offers you a function especially designed for the analysis of responses to open questions and the associated formation of categories. If you go to ***Analysis*** > ***Categorize Survey Data,*** a window will open into which you can drag a code with the responses to an open question from the "Code System," and this will in turn open the analysis view. In Fig. 16.4 you can see the view of the Categorize Survey Data function for the question "How would you define a good quality of life?". In the left window area, the "Code System" is reduced to the codes corresponding to question you want to analyze, and in the "Responses" column, you can see the texts written by all the respondents for that question. You can use your mouse to select response texts, or parts of them, and drag them onto a code, after which that code will be displayed in the "Assigned Codes" column. The fourth column "Comment" lets you record notes, comments, and reminders for each individual response. Whether you have decided to form your categories a priori, based on the data, or by applying a combination of these two approaches, you can click on the ***New code*** icon to add new subcategories either in advance or as your analysis progresses.

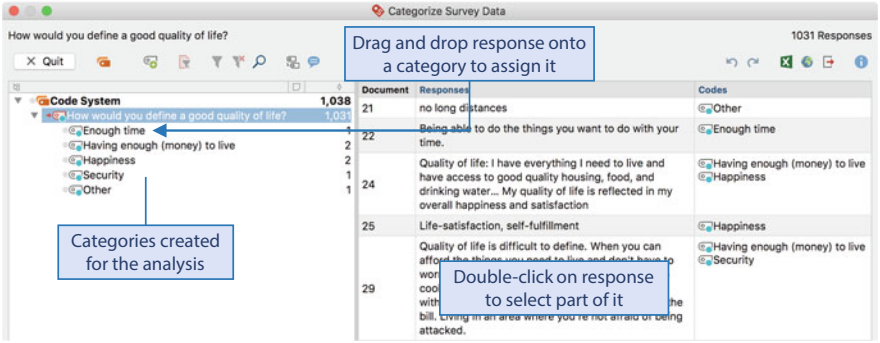

Fig. 16.4 Analysis view for categorizing responses to open questions

Sometimes responses to a given question also address topics relating to other questions than the one answered—topics which then also need to be coded here. By activating the option ***Display all codes*** in the toolbar, you can briefly display the complete "Code System" and perform such "external coding" and then hide the unnecessary codes again by deactivating this option.

► **Tip** The shortcuts you can set by right-clicking on a code and opening its properties menu are also available for coding survey responses. By right-clicking on a response and selecting ***Create and Assign New Code***, a new subcode can be added directly below the code associated with the open question and assigned to the selected response.

Once all the responses have been categorized, you can generate an overview of the individual categories and their code frequencies. What is helpful here is that the subcodes can be sorted by frequency—you can do this simply by right-clicking on the code of the open question and then selecting the corresponding function. To get an overview of the coded segments of a particular category, click on the code in the displayed code system; only the responses corresponding to the selected category will then be displayed in the "Responses" column.

Analyses of Interrelationships and Mixed Methods Analyses

While the "Categorize Survey Data" window is open, MAXQDA's other functions are locked. Only once you close the window, when you have completed the categorization process, are further functions available once again for analyzing the interrelationships between categories and for mixed methods analyses involving document variables. Since several hundred, even several thousand cases are often analyzed in surveys, strategies that are based on the frequencies of code assignments can produce very interesting results. Although individual cases are not the focus in this situation, you can still review them at any time and compile responses based on groups or topics. The following are some particularly noteworthy strategies:

Group comparisons based on imported variables—Using the ***Mixed Methods > Crosstab function***, you can compare code frequencies and responses for different groups of respondents, for example, those living in Marburg's city center and those living in the countryside. These groups may, for instance, have completely different expectations regarding leisure activities in and around Marburg.

Group comparisons based on categories applied to open responses—Groups can also be formed that are based on the coded responses resulting from the categorization process. In the simplest case, at least two groups can always be compared, namely, those who gave a specific response and those who did not. In the Marburg student study, for example, you could compare respondents who mentioned "friends" when asked about quality of life with those who did not write about "friends." The first step here is to add the code "Friends" as a new document variable, which you can do via the function ***Transform into Document Variable*** in the code's context menu. The newly formed variable then indicates how often the code "Friends" was assigned to each case, which means that this variable can also be used to generate a crosstab. Additionally, the ***Mixed Methods > Statistics for QUAL Groups*** function lets you create a so-called joint display (Guetterman, Creswell, & Kuckartz, 2015), in which you can compare the statistical values of groups formed for that purpose. A joint display of this kind can be used to answer questions such as "How old on average are the respondents who named friends as an aspect of quality of life?". If you use the MAXQDA add-on module Stats for statistical analysis, you can dispense with the intermediate step of converting codes into variables and simply use code frequencies as grouping variables to generate descriptive statistics and crosstabs, and for performing variance analyses.

Combinations of categories—Which categories often co-occur in the same case? To answer this question, you can use the functions listed under ***Analysis > Code Configurations***. In a simple configuration analysis, several codes are selected—usually the subcategories of an open question formed in the course of your analysis, where you have used a "code all that apply" approach. MAXQDA then presents a list containing the combinations of subcategories and displays the number of cases in which each combination occurs. For an open question about the respondents' understanding of quality of life, this function could be used to investigate whether the aspects "friends," "family," and "leisure time" are often mentioned in combination. In the case of complex configuration analyses, at least two top-level codes of different open questions must be selected. MAXQDA then analyzes the frequency of all possible combinations of their respective subcodes and presents the occurring combinations, together with their respective frequencies, in a table. With regard to the study involving Marburg students, this technique could be used to analyze how strongly the lack of recreational opportunities and facilities correlates with certain notions about quality of life. Finally, both simple and complex code configurations can also be analyzed to create a typology with types of respondents.

References

Guetterman, T., Creswell, J. W., & Kuckartz, U. (2015). Using joint displays and MAXQDA software to represent the results of mixed methods research. In M. T. McCrudden, G. J. Schraw, & C. W. Buckendahl (Eds.), *Use of visual displays in research and testing: Coding, interpreting, and reporting data*. Charlotte, NC: Information Age Publishing.

Kuckartz, U., Ebert, T., Rädiker, S., & Stefer, C. (2009). *Evaluation online: Internetgestützte Befragung in der Praxis*. Wiesbaden: VS Verlag für Sozialwissenschaften.