Stefan Rädiker, Udo Kuckartz

Analyzing Open-Ended Survey Questions with MAXQDA

Step-by-step

Guide

MAXQDA PRESS

Citation: Rädiker, Stefan & Kuckartz, Udo (2020). Analyzing open-ended survey questions with MAXQDA. Step-by-step. Berlin: MAXQDA Press.

ISBN 978-3-948768027
https://doi.org/10.36192/978-3-948768027

MAXQDA Press is a division of VERBI Software. Consult. Sozialforschung. GmbH
Invalidenstr. 74, 10557 Berlin, Germany, info@maxqda-press.com, www.maxqda-press.com

Content

Introduction

The motives for using open-ended questions in surveys are manifold. According to Porst (2014), open-ended questions should be used when not much is known about the topic under investigation, when the possible range of answers is very large, or when there is a risk that the preset answers might steer respondents too much in a particular direction. Through answering open-ended questions, respondents can express their opinions, views and experiences in their own words, and their free-text answers provide valuable qualitative data to supplement the standardized quantitative data. There are numerous suggestions for the analysis of free-text responses; many of these approaches focus on coding the data (e.g Popping, 2015; Züll & Mohler, 2001) and in recent years increasingly on (semi-)automatic coding (Roberts et al., 2014; Schonlau & Couper, 2016; Schonlau, Gweon, & Wenemark, 2019; Senderovich & Maysuradze, 2015), with the primary aim of transforming the originally qualitative data into quantifiable data. Much less frequently, as in Fielding, Fielding, and Hughes (2013), there is discussion of how the qualitative free-text responses can be analyzed together with the quantitative standardized survey responses.

In this paper, an approach is presented in which the answers to open-ended and closed-ended survey questions can be analyzed both separately and integratively using the MAXQDA software package. MAXQDA is a so-called *QDA software* (Qualitative Data Analysis Software), which provides functions for the analysis of qualitative data as well as a variety of functions for processing mixed methods data. The software allows the open-ended and closed-ended questions to be analyzed separately in one software; for example, the answers to open-ended questions can be thematically coded and frequency tables and statistical characteristics can be calculated for the standardized answers. Moreover, integrative analyses can be carried out in which the two types of data are related to each other and are used for the creation of so-called joint displays (Creswell & Plano Clark, 2018, pp. 227–232; Guetterman, Creswell, & Kuckartz, 2015). Furthermore, it is possible to switch back and forth between a cross-case analysis and a holistic analysis of individual cases at any time.

The aspect of integration encompasses different dimensions in the approach presented here: on the data level, free-text answers are integrated with standardized data (qualitative + quantitative) and the answers to various open-ended questions are integrated with each other (qualitative + qualitative). At the level of analytical methods, word-based and category-based procedures as well as individual case studies are linked with cross-case procedures.

The following presentation follows a typical analysis process – step-by-step – and is based on preliminary work that was methodologically reflected on the example of a teaching evaluation using a mixed methods approach (Kuckartz, Ebert, Rädiker, & Stefer, 2009). An online questionnaire was used to obtain feedback

from students on a university course in social science statistics. Standardized answers to closed-ended questions were used to collect information about participation behavior and assessments of learning materials, lecturers, personal learning outcomes and the course as a whole. In addition, age, gender and school leaving certificate grades in mathematics served as differentiating variables. The questionnaire also contained seven open-ended questions or requests, for example:[1]

- Please describe how you have worked through the content of this course.
- Please describe your feelings about statistics at the beginning of the semester. Have these changed during the semester?
- What did you particularly like about the whole course (lecture/exercise/tutorial)?
- What did you dislike?
- What suggestions for improvement do you have?

In the following, we will show how the answers to such open-ended and closed-ended questions are imported *together* into the MAXQDA software so that the reference of the data parts is preserved for an integrative analysis perspective (Section 1). Already during data exploration (Section 2) it is possible and useful to adopt such an integrative perspective and to combine the free-text answers and the standardized answers – especially for within-case explorations. The subsequent manual and automatic coding (Section 3) focuses primarily on open responses and forms the basis for numerous category-based analysis strategies (Section 4). A research project will not always go through all the steps and analysis procedures proposed in this paper. For this reason, the various analysis options are summarized in tabular form and will be presented together with typical analysis questions in Section 5. The article concludes with a critical discussion of the surplus value of (integrative) survey data analysis with MAXQDA (Section 6).

1 Data Preparation and Data Import

Usually, the data of a survey exist in the form of a rectangular matrix, with each case occupying one row and the questions or answer options forming the columns. It is exactly such a table, containing both the answers to the open-ended questions and the answers to the closed questions, that MAXQDA expects for the import of survey data. The table must be available as an Excel file, which in the case of online surveys can usually be exported directly from the questionnaire software, such as Qualtrics.com or LimeSurvey.org. Statistical pro-

1 The original study was conducted in German. All presented study-related information and data have been translated from German to English by the authors.

grams such as SPSS, STATA, and R also offer the possibility of saving a data matrix in Excel format. When creating the Excel file, you must decide whether the values of numeric variables or their respective value labels are to be exported. As a rule, it is recommended to output the value labels for nominal-scaled variables in order to get along without a code plan (e.g. “female”, “male” and “divers” instead of “1”, “2”, “3”) and for ordinal-scaled variables, to add the numerical value beforehand the value itself to maintain the order of the possible answers in case of alphabetical sorting (“1 = always”, “2 = often”, “3 = rarely”, “4 = never”, etc.). Fielding et al. (2013) suggest to import interval-scaled characteristics categorized into groups for the analysis of free-text responses in QDA software. In fact, such categorized characteristics are frequently used, especially in integrative approaches, for example in the formation of groups for joint displays (Kuckartz & Rädiker, 2019, pp. 171–186). However, since the grouping of the interval-scaled data leads to a loss of information and MAXQDA also provides special functions for categorizing such data, it is preferable to transfer the values directly so that later calculations of characteristic values such as median, mean value and standard deviation remain possible.

When the data is imported into MAXQDA (Import > Survey Data), the free-text answers and the standardized data are split up (Fig. 1): each case becomes a separate text document containing the answers to the open-ended questions. The headings of the questions for the free-text answers are adopted as code names and the answers are automatically assigned to the corresponding codes in the text document. The number behind each code in the “Code System” window indicates the number of cases from which answers to the open-ended question are available. The answers to closed-ended questions are – as known from statistical programs – available as a rectangular data matrix. This matrix can be accessed via Variables > Data Editor for Document Variables.

Tab. 1 shows an excerpt from a simple data matrix as it is optimally structured for import into MAXQDA. The first row of the import table contains a shortened version of the questions or statements as headings, because this is usually clearer for further analysis and secondly because MAXQDA cuts off the contents of the first row after 63 characters during import. The table should contain a column with a case ID, which allows a unique identification and assignment of cases. This ID is used in MAXQDA to link the free-text answers with the standardized data, because this ID is used as the name/label of a case. If, after the analysis in MAXQDA, further statistical analyses are to be carried out with other statistical software, the case ID provides the possibility for matching the cases. It is not necessary to import all variables with standardized data. Especially in surveys with many variables, it may be appropriate to select those characteristics that promise added value for integrative analyses. If it becomes apparent in the course of the analysis that further standardized data is required, this can be imported subsequently and assigned correctly via the case ID.

When the data is imported into MAXQDA (Import > Survey Data), the free-text answers and the standardized data are split up (Fig. 1): each case becomes a separate text document containing the answers to the open-ended questions. The headings of the questions for the free-text answers are adopted as code names and the answers are automatically assigned to the corresponding codes in the text document. The number behind each code in the "Code System" window indicates the number of cases from which answers to the open-ended question are available. The answers to closed-ended questions are – as known from statistical programs – available as a rectangular data matrix. This matrix can be accessed via Variables > Data Editor for Document Variables.

Tab. 1: Structure of the import table with answers to open-ended and closed-ended questions

Case ID	**Work through the content [open-ended question]**	**Liked [open-ended question]**	**Course proceeds according to a clear structure [1 = not true; 4 = true]**	**Overall rating [1 = very good; 6 = insufficient]**
…	…	…	…	…
4	I only go through the slides again after the lecture, if I don't understand something, I check the Bortz [text book]. Then I try to solve the tasks in the reader. For further questions, I ask in the tutorial.	The lecture is very theoretical. That's why I think it's good that in the exercise, you have to apply the theoretical knowledge in practice! The tutors told us important and valuable things beyond the lecture and exercise. I therefore think the whole course was a success.	1	2
…	…	…	…	…

Note. This is an extract from a data matrix with n=194 cases, 7 open-ended and 18 closed-ended questions/items; the information in square brackets is for explanation only. Original text in German (with lots of spelling errors); translated by the authors.

When importing data from SurveyMonkey.com, there is a special feature that provides direct access to its API so that data can be imported directly without having to go through Excel (Import > Survey Data). For panel

studies it is also possible to add data of the same persons at a later date: further answers to open-ended questions are appended to the text of a case below and further closed questions are added as new columns in the data matrix.

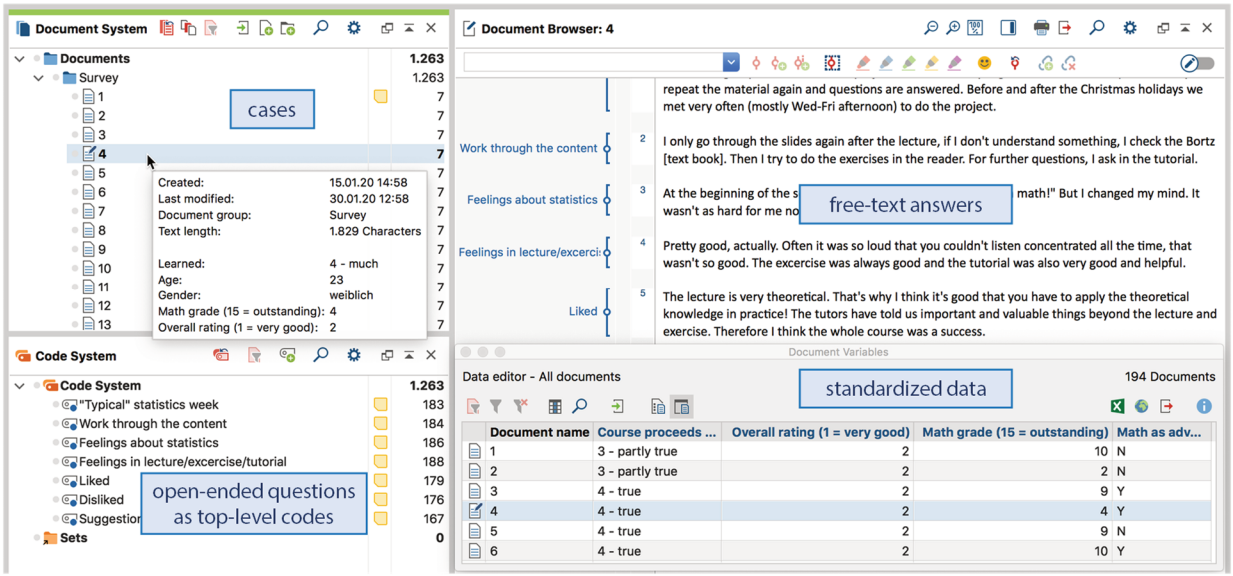


Fig. 1: Survey data directly after import into MAXQDA

2 Data Exploration

The first step of the analysis is to explore the data in order to become familiar with them, to record initial findings and to develop suitable starting points for analysis for the type of data. In this step, initial assumptions about relationships and recurring patterns can also be noted, which can be checked in the further course of the analysis. Data exploration can be directed at the free-text answers and/or the standardized answers, or a combination of the two types of data. Instead of only distinguishing between data types, it is more appropriate to distinguish between within-case and cross-case exploration.

2.1 Within-Case Data Exploration

Within-case data exploration means taking a holistic perspective and looking at the free-text answers of single respondents together with their answers to a selection of closed questions. This approach may at first seem unusual for primarily quantitatively-oriented researchers, since individual cases play a less important role in their work than in qualitatively-oriented research. However, as some examples will show below, it may well be a worthwhile approach.

First, cases must be selected for exploration, since the sample is usually so large that not all cases can be subjected to a more in-depth analysis due to time constraints. Kuckartz et al. (2009, pp. 67–68) suggest a random selection, since they see the danger in a theory-based selection that the gaze may be too narrowed and guided by one's own pre-categorization. Kuckartz et al. recommend selecting 5-10% of cases, which is certainly a good guideline. However, especially with large samples, the number should be limited to a manageable number for reasons of time (for example, a maximum number of 20 cases should be specified). Regardless of the number of cases, the following applies: if further case studies provide little new information, then it is not worth continuing this procedure.

After the selection of cases, background variables are chosen that can be displayed in MAXQDA parallel to the free-text responses of a case directly at the mouse pointer (Fig. 1). These variables can be selected by clicking on the gear symbol in the "Document System" window and enabling the Display favorite variables in tooltip option. From this mixed-data information, integrative case summaries can be written, in the present example those that focus on learning behavior (Kuckartz et al. 2009, p. 70):

Person 26: Learned a lot, independent learner, positive judgement - woman, 18 years, math grade 9 of 15 pts.

- Assesses the course with "very good" [c]
- Has learned a lot in the course [c]
- Never attended the tutorial [c, o]
- Has attended the course regularly and has acquired the material very independently [c, o]
- As the course progressed, she became more and more confident that she would pass the final exam [o]
- Regards the textbook used as incomprehensibly written [o]
- She evaluates the lecturer friendly and helpful, the examples useful and the exercises helpful [o]
- In the lecture it was too loud for her [o]

The summaries are given a title containing the main points of the case in condensed form. The title can also be supplemented with standardized information: in this case, gender, age and mathematics grade upon leaving school. In order to explain the sources of the example summary, each bulleted point is supplemented

with letters in brackets: answers to closed-ended questions are represented by [c], those from open-ended question by [o], and those from both by [c, o].

By creating such integrative summaries for several cases, initial important analysis results and assumptions about connections can already be made. These are worth analyzing in more detail in the further course of the project. In this way we were able to identify three different learning types among the students: firstly, the form of a high degree of autonomy in learning expressed in case 26 above; secondly, a guided three-stage learning; and thirdly, a tendency towards "disinterested" learning, which is strongly characterized by the goal to "pass the exam anyhow" (Kuckartz et al., 2009, p. 71).

The results of case-oriented data exploration can be recorded in MAXQDA in document memos – these can be used to store both the case summaries themselves and the assumptions and hints about relationships that arise during the exploration. Different symbols for the memos help to differentiate between different types of notes and to be able to access important thoughts and analysis ideas later. Should central or easily citable formulations in the free-text answers stand out during exploration, these can be highlighted in color with an electronic highlighter or inserted in memos with a reference to the source and are thus available for analysis and reporting. By sorting the data matrix with the standardized answers, it is also possible to obtain targeted access to the free-text answers of selected persons who have given a particularly high or low value for a variable, such as those who rated the course as very well or very poorly. In this way – in the spirit of "serendipity", i.e., the accidental finding of what was not originally sought (Merton & Barber 2003) – new discoveries and assumptions about connections can be made, providing valuable starting points for later cross-case analysis.

2.2 Cross-Case Data Exploration

Usually, the free-text answers will be analyzed question by question in cross-case analysis, but it is quite conceivable to apply the procedures presented in the following to several free-text answers simultaneously. First, it is advisable to examine some formal aspects in advance:

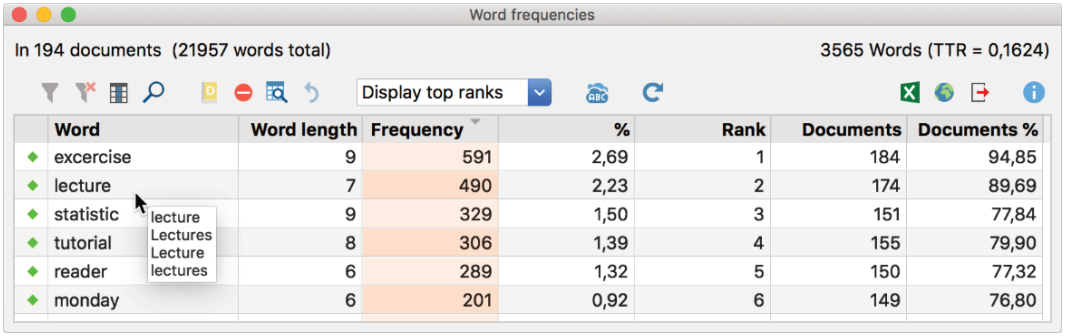
- How long are the answers and what kind of answers are there, i.e. do they consist only of single words, bullet points with short phrases or whole sentences?
- How much do the answers differ with regard to these aspects?

By activation of all documents and one question code MAXQDA lists all answers to a question in the "Retrieved Segments" window, which makes it easy to answer these kinds of qusetions. In general, it can be assumend that the length of answers depends primarily on the type of question asked: answers to "other" or

"comment" questions, and questions about associations of terms are shorter than questions about motives and reasons.

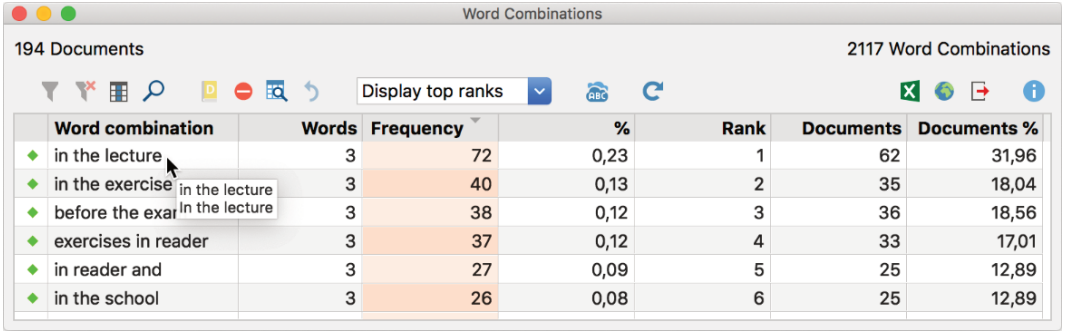
Word-based procedures are suitable for the cross-case exploration of free-text answers. Two basic approaches can be distinguished here. In *data-driven* analysis, the most frequently occurring word frequencies and word combinations are examined. This approach is also labelled as *inductive* (although this term, which comes from research philosophy, is not quite accurate, it is often used in the context of text analysis.) In *theory or concept-driven* (*deductive*) analysis, words and word combinations of interest are searched for directly. For example, certain terms may already have attracted attention during case-oriented exploration, the use and frequency of which are now being investigated in more detail.

For word-based analysis, MAXQDA provides frequency tables with single words or with 2- to 5-word combinations (MAXDictio > Word Frequencies and MAXDictio Word Combinations). Fig. 2 shows the most frequent-words for the free-text answers to the question on how the students worked through the content of the course. The table can be sorted by each column, i.e., alphabetically, according to length, frequency, and ranking of the word combination. The most common word in the analyzed free-text answers is "excercise". The last column presents the information that almost 95% of the respondents used this word. When analyzing word frequencies, it is necessary to use stop lists to exclude words that occur frequently but are useless for analysis. Stop lists usually include articles, conjunctions and prepositions. However, stop lists should be used with caution and the excluded words should be checked for analytical importance. For example, the conspicuously frequent use of words that introduce justifications (such as "because") can also provide important clues. When analyzing single words as well as combinations of words, the function for lemmatizing the result table can be used. Lemmas are the basic forms of words, i.e., in the results table the inflections of a word are summarized in one row. Words that have been aggregated in a row are listed at the mouse pointer (Fig. 2).



Word	Word length	Frequency	%	Rank	Documents	Documents %
excercise	9	591	2,69	1	184	94,85
lecture	7	490	2,23	2	174	89,69
statistic	9	329	1,50	3	151	77,84
tutorial	8	306	1,39	4	155	79,90
reader	6	289	1,32	5	150	77,32
monday	6	201	0,92	6	149	76,80

Fig. 2: Frequency table for single words (using lemmatization and a stop list with excluded words)



Word combination	Words	Frequency	%	Rank	Documents	Documents %
in the lecture	3	72	0,23	1	62	31,96
in the exercise	3	40	0,13	2	35	18,04
before the exa	3	38	0,12	3	36	18,56
exercises in reader	3	37	0,12	4	33	17,01
in reader and	3	27	0,09	5	25	12,89
in the school	3	26	0,08	6	25	12,89

Fig. 3: Frequency table for 3-word combinations (without usage of lemmatization or stop list)

The interactivity of MAXQDA's result tables is helpful for the analysis: a double click on a line with an interesting word combination in Fig. 3 lists all occurrences of the word combination with its surrounding context; a further click on one of these opens the corresponding case and displays all free-text answers of the person – and, if desired, their standardized answers. By using the QDA software, the case as a whole with all its data is always accessible, even in the case of cross-case analysis. For individual terms, a list including an adjustable number of words before and after can also be requested at any time (MAXDictio > Keyword-in-context).

Important in this context is the question of what defines a word or phrase as interesting and useful. Fielding et al. (2013) provide valuable information in this regard:

> 'Useful' words will occur with quite high frequency (indicating some commonality amongst respondents) but not so frequently that they lack any power to discriminate between respondents. They will

> have a link of some sort to the topic under investigation and they should not have too many separate distinct meanings (to avoid ambiguity). (p. 3267)

In addition to the tabular analysis of word frequencies and keywords in context, the use of MAXQDA's visualizations is also suitable for cross-case exploration. Word clouds can be used to display the most common words from the free-text answers to one or more questions (Visual Tools > Word Clouds). Numerous options are available for word cloud design, such as shape, color, alignment, etc. so that the cloud can be prepared for a presentation of results. Furthermore, the Interactive Word Tree (available in ribbon tab MAXDictio) allows the exploration of frequently recurring words and phrases in free-text answers in a visual way. The most common word represents the root of the word tree and the word that follows this word most often forms the topmost, strongest branch, the second most common word forms the second topmost branch, and so on. In order to visualize the answers to a single open-ended question with the Word Tree, it is necessary to first save them in a separate document:

1. Activate all documents and the code of the open-ended question so that all answers to the question are displayed in the "Retrieved Segments" window.
2. Select all answers with Ctrl+A (Windows) or cmd+A (Mac) and copy them to the clipboard.
3. Create a new text document and paste the contents of the clipboard.

The results of the cross-case data exploration of the free-text answers – including the generated graphics – can also be recorded as memos in MAXQDA, e.g., directly at the top-level category as a code memo for the respective open-ended question. The results of searches for individual words or word combinations can be saved directly as codes (see Section 3.2).

The standardized data are also evaluated across all cases. With the help of MAXQDA Stats a univariate basic count with frequency tables, diagrams, and summarized characteristic values (mean value, standard deviation) can be generated. The results are either recorded directly in MAXQDA Stats in the output viewer or copied into memos.

3 Coding of Responses to Open-Ended Questions

The indications of important formulations, topics and differentiation criteria obtained through data exploration provide a good basis for the subsequent coding of the free-text responses. Coding usually refers to the process of assigning a free-text answer or a part of it to a thematically appropriate category. In this context, categories serve the purpose of classification, reduction, abstraction and attribution of meaning, i.e., categories are used to index, describe and explain the data material. Categories are the central tool for the analysis

of text data (Kelle & Kluge, 2010; Saldaña, 2015). In the following we distinguish between three different forms of coding:

- manual coding, in which the free-text answers are all read and manually assigned to categories;
- automatic coding using search words, in which the hits of an automatic text search are assigned to a category;
- and coding using a dictionary that counts how often predefined search terms, grouped into categories, occur in the free-text answers.

3.1 Manual Coding

The manual coding of free-text responses generally follows (more or less consciously by the researchers) the principles of a structuring qualitative content analysis (Kuckartz, 2014; Mayring, 2014; Schreier, 2012). The process of manual coding basically begins with the creation of a category system in which several categories are usually arranged in a hierarchical structure. The creation of the category system can follow either a concept-driven (deductive) or a data-driven (inductive) approach. Kuckartz et al. (2009, pp. 78–79) provide indications as to when which method is suitable: the concept-driven approach is suitable if information on the topic is already available or if concrete questions and assumptions, such as those that have emerged during data exploration, are to be checked. The data-driven approach, on the other hand, would be appropriate if the spectrum of possible answers is unknown and no preliminary categorization is to be made. In most projects, both procedures are mixed, i.e., some categories are created by a concept-driven approach and supplemented or refined by categories that arise when working through the data material.

MAXQDA offers a special function called “Categorize Survey Data” for the manual coding of free-text responses, which is suitable for all described methods of category formation (Analysis > Categorize Survey Data). The function offers an interactive environment where the free-text answers to an open-ended question are listed one below the other (Fig. 4). Directly to the right of the answers, a separate column shows the categories already assigned. The left pane contains the categories created for the analysis, which can be assigned to a response or parts of it.

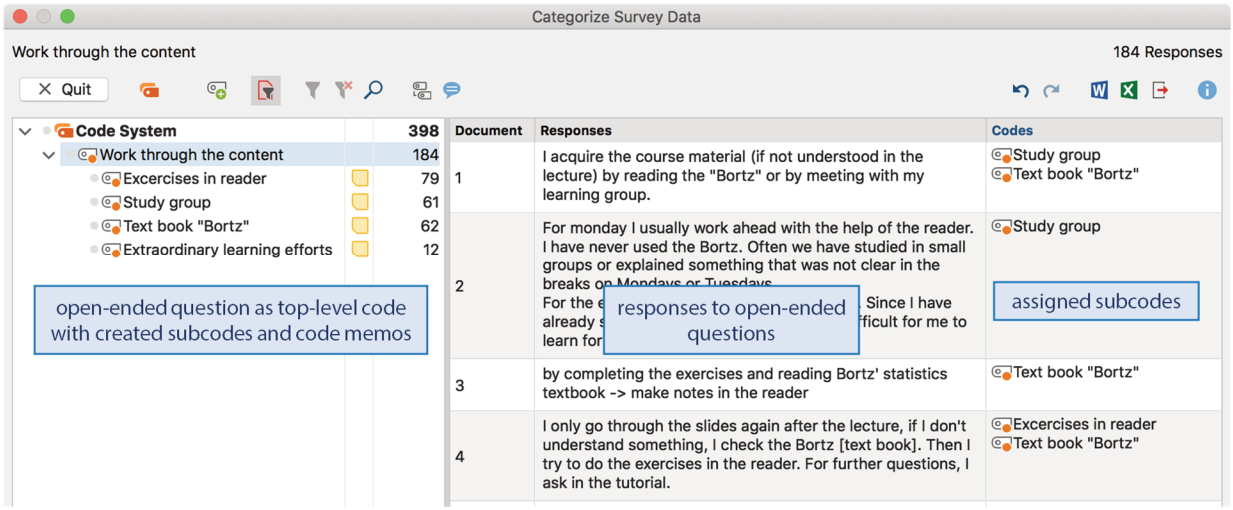


Fig. 4: The *Categorize Survey Data* work environment in MAXQDA

Qualitative data analysis usually follows an iterative and circular process and adjustments to the category system are particularly necessary at the beginning of the coding process. Therefore, in MAXQDA's work environment several categories can be combined into one, and – vice versa – a very broad category can be differentiated into further categories in the on-going analysis. To check which free-text responses have already been assigned to a category, the category can be selected in the left pane. MAXQDA then automatically filters the view and displays only the free-text responses of the selected category in the right pane.

For reasons of quality assurance, the method of qualitative content analysis requires us to add definitions to each category. These definitions describe when the category is used and how it distinguishes itself from other categories (Kuckartz, 2014; Schreier, 2012). These category definitions are stored in memos, which are directly assigned to the corresponding codes. In addition, it is important to define coding rules that control the exact coding procedure. The rules are to be adapted to the respective material and situation, whereby the following rules are usually suitable for the analysis of open-ended questions:

Scope of segment (coding unit) – Individual statements or 'units of sense' are coded. Since free-text answers can vary greatly and range from individual keywords to elaborate continuous texts (e.g. in the data of Waldherr et al. , 2019), the claim to always code whole sentences would not be appropriate. If most answers are only a few lines long, you can specify that the whole answer is coded always.

Repeated information – If a person names several relevant aspects in an answer, then several categories are assigned, too. However, a person may name the same aspect several times. In this case, the category should only be assigned once in order to be able to determine in a later stage of analysis how many people named an aspect (although MAXQDA provides explicit functions to count codes only once per case).

Separated information – Particularly in the case of open-ended questions with similar topics, it happens that respondents refer to previous answers and simply write "see above" or "see question 3". In our example this was the case with the question "What suggestions for improvement do you have?", because some people had already given appropriate suggestions while answering the question "What did you dislike?". In such cases, it would be fatal not to code relevant answers because they are related to another open-ended question. In MAXQDA's "Categorize Survey Data" window, it is possible to also show the categories for the analysis of other open-ended questions in order to be able to code "wrongly placed" answers for the correct question.

Missing answers – By default, MAXQDA does not code empty answers to open-ended questions during import, so that the number of people who have (not) answered a question can be read from the number of encodings after import. However, it may be that some people have expressed their non-answer by means of a dash or something similar. Code assignments to such symbols should be deleted or alternatively, the symbol should be coded with a category "missing". Answers such as "none" might be problematic, because depending on the topic and content of the open-ended question it has to be decided whether they should be interpreted in the same way as empty answers.

An important contribution to the quality of coding open-ended questions is the verification of intercoder agreement, which is also an essential requirement in qualitative content analysis. It is therefore not surprising that this topic has been discussed for a long time (e.g. by Montgomery & Crittenden, 1977; Carey, Morgan, & Oxtoby, 1996). Teamwork functions are available in MAXQDA allowing two coders to work simultaneously on the same data. The agreement of two coders is checked with a specially designed interactive function, which at the same time allows the optimization and correction of the disagreements (Analysis > Intercoder Agreement).

3.2 Automatic Coding of Search Hits

Automatic coding starts with a search for a word or phrase that is interesting and useful. This word or phrase can either be significant from a theoretical point of view or it can be a result of the data exploration phase. Fielding et al. (2013) describe an iterative process for creating a category system using automatic coding. The process – slightly supplemented – looks as follows and can be implemented one-to-one in MAXQDA with MAXDictio > Word Frequencies:

1. Create a word frequency table and select a word of interest.
2. Compile all text passages with the word of interest and check the results.
3. If the word is not useful for the analysis, select another word.
4. If the word is useful, search for synonyms and other spellings in the alphabetically sorted word frequency table (especially in online surveys, words are often misspelled, contain mixed-up letters, etc.)
5. Perform an automatic coding of all identified terms. Use the central search word as code name.
6. Create a category definition as a code memo that contains the search terms used and a note that the category has been assigned automatically.
7. Check the assignments, particularly whether the opposite of what you are looking for was meant (since the automatic coding does not take negations into account).

Usually it is of little use in step 5 to code only the search word itself, instead you will use the MAXQDA function to code several words before and after the search hit, the surrounding sentence, or the paragraph. In our study we searched for "fear" (of statistics) in all answers to open-ended questions and automatically coded the results, because we were interested in the extent to which students were confronted with this feeling and if so, what exactly the fear related to. The result was a kind of "compendium of fear".

Automatic coding processes can be integrated into a manual coding process to save time. If, for example, statements with certain terms are repeatedly assigned to the same category during manual coding, they can be automatically coded together with the surrounding sentence or paragraph in the remaining data.

3.3 Automatic Coding Using a Dictionary

The first step in automatic coding with a dictionary is the creation of the dictionary consisting of linear or hierarchical categories (MAXDictio > Dictionary). Each category contains thematically matching search terms or search phrases. The text is coded by counting how often a category – defined by its search terms – appears in the free-text answers to a question. This means that a slightly different coding term is used in this procedure than in qualitative content analysis: as a result of the coding process, there are not coded segments, but a frequency table of categories. In MAXQDA, each category of a result table can be converted into a document variable (click on symbol Transform results into document variables). This variable indicates how often the category has been assigned in a case, i.e. how often its search words have been found per case. It can be used for further bivariate and multivariate (integrative) analyses. For example, the correlation between the frequency of a category and the age of the respondents can be calculated.

4 Category-Based Integrative Analysis

Once the coding of the free-text responses is complete, the analysis work is by no means over. On the contrary, it is the prerequisite for numerous analysis procedures for open survey questions, which can be divided into three areas in principle:

1. The in-depth analysis and description of individual categories and their subcategories, comparable to a univariate analysis in statistics. (What does a typical statistics week among students look like? How do they work through the content of the course and acquire the presented knowledge? Which feelings about statistics dominate them?)
2. The analysis of relationships between categories, comparable to the analysis of correlation and difference hypotheses in quantitative research. (To what extent does the way of acquiring knowledge correlate with the way the students participate in the lecture, exercise, and tutorial part of the course? Do the persons who tend to express negative feelings about statistics differ in their acquisition of knowledge from those with positive feelings?)
3. The analysis in which qualitative free-text answers are combined with quantitative standardized answers in the spirit of mixed methods (Creswell & Plano Clark, 2018). (Do older respondents write differently about the tutorial than younger ones? To what extent do the category frequencies differ between those with a good and those with a bad grade in mathematics?)

In the following, we describe different procedures in MAXQDA, with which all three types of analysis can be implemented.

4.1 Compilation of coded segments

It is possible to compile the coded free-text answers in the "Retrieved Segments" window according to different criteria, i.e. to carry out a so-called "text retrieval". The retrieval can contain coded segments from one category or from several categories, whereby only single or several cases can be considered. Quantitative information can be used for the selection of cases. For example, the retrieval can be reduced to statements made by people with good grades in mathematics (Mixed Methods > Activate Documents by variable). In the same way, the assignments of categories, that have been generated by coding the free-text answers, can be used for case selection (such as by compiling only the statements on knowledge acquisition of persons with positive feelings towards statistics).

Fielding et al. (2013, p. 3273) distinguish between 'single sort' and 'multiple sort' retrievals. While 'single sort' requires only one category to be assigned to a text passage in order to be compiled, 'multiple sort' re-

trievals require that several categories are assigned to the same text passage. This means that in a 'multiple sort' retrieval, several codes must overlap at one text passage. For example, if data exploration shows that the textbook plays a role even during joint learning meetings of the students, it would be interesting to search for text segments for which both the categories "textbook" and "learning meeting" were assigned.

4.2 Statistics for subcategories

When evaluating open-ended questions, there is often the desire to generate statistics of (sub)categories in order to answer the question of how many of the respondents mentioned certain aspects. MAXQDA can create a corresponding table of category frequencies for each open-ended question separately (Codes > Subcode Statistics). Cases where none of the counted categories have been assigned can be classified as missing.

Instead of a frequency table or a bar chart (Fig. 5) generated from it, the frequencies of categories can also be displayed as a word cloud, in which more frequently occurring categories are visualized in a larger font (Codes > Code Cloud). Such a presentation only highlights the relative differences between frequencies, which is better for those open-ended questions where a frequency table would pretend to be inappropriately accurate.

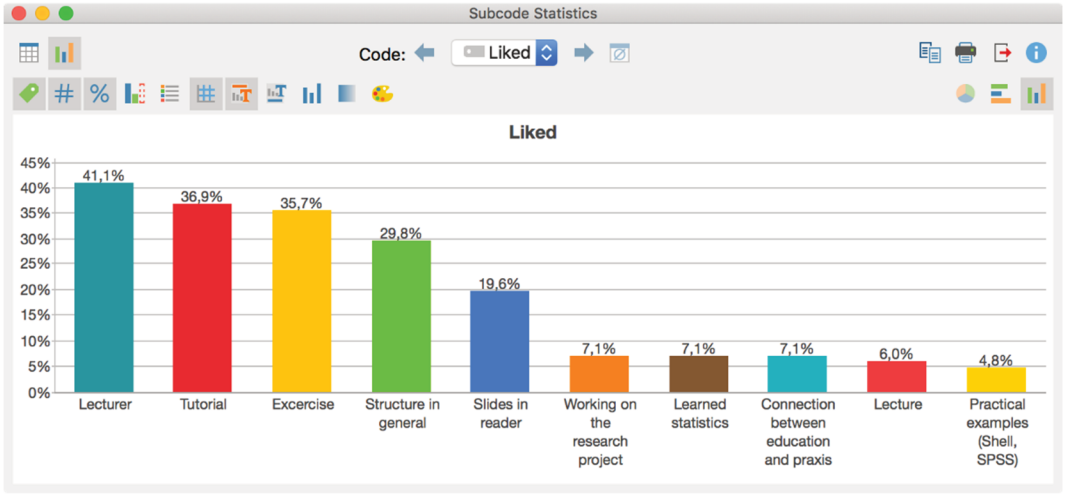


Fig. 5: Bar chart showing the frequency of subcategories

4.3 Pairwise occurrence of categories

While multiple sort retrieval is a method of compiling statements in which a selected number of categories have been assigned, it is also possible to systematically analyze the co-occurrence of categories in the data. Three types of analysis are available for the definition of "co-occurrence" of two categories:
(1) both categories must overlap, i.e., they are assigned to parts of the same text passage,
(2) it is sufficient if both categories occur at a defined distance from each other,
(3) it is sufficient if both categories were assigned somewhere in the person's responses.
The function Visual Tools > Code Relations Browser creates a matrix showing how often two categories occur together (Fig. 6). The more often this is the case, the larger the squares are displayed; however, it is possible to switch to number display at any time. The similarity matrix between the categories is transformed into a distance matrix and is used to create a so-called "Code Map" (Fig. 7). This map uses classic multidimensional scaling to place categories on a surface so that similarly used categories are closer together and easier to identify than in the matrix. In addition, the categories can be colored in groups using a hierarchical cluster analysis with average linkage.

Both the matrix and the resulting Code Map are interactively linked to the cases and the original data. Individual codes can be removed and by double-clicking on a matrix cell the corresponding text segments are listed, so that a content check and examination of the multiple coded answers can be conducted.

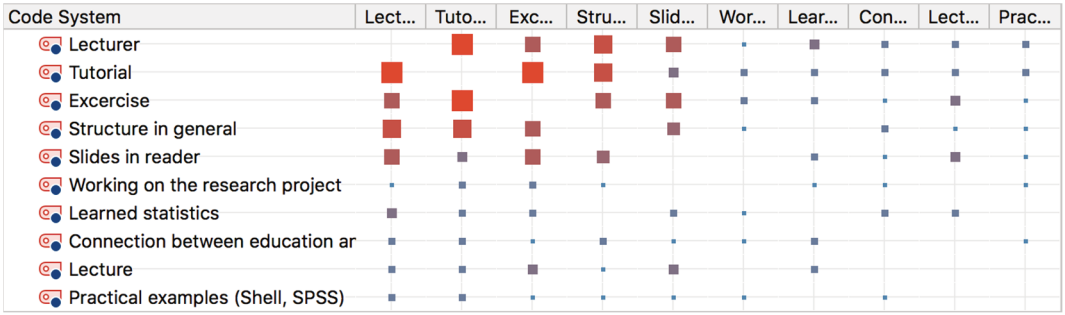


Fig. 6. *Code Relations Browser* showing the paired occurrence of categories

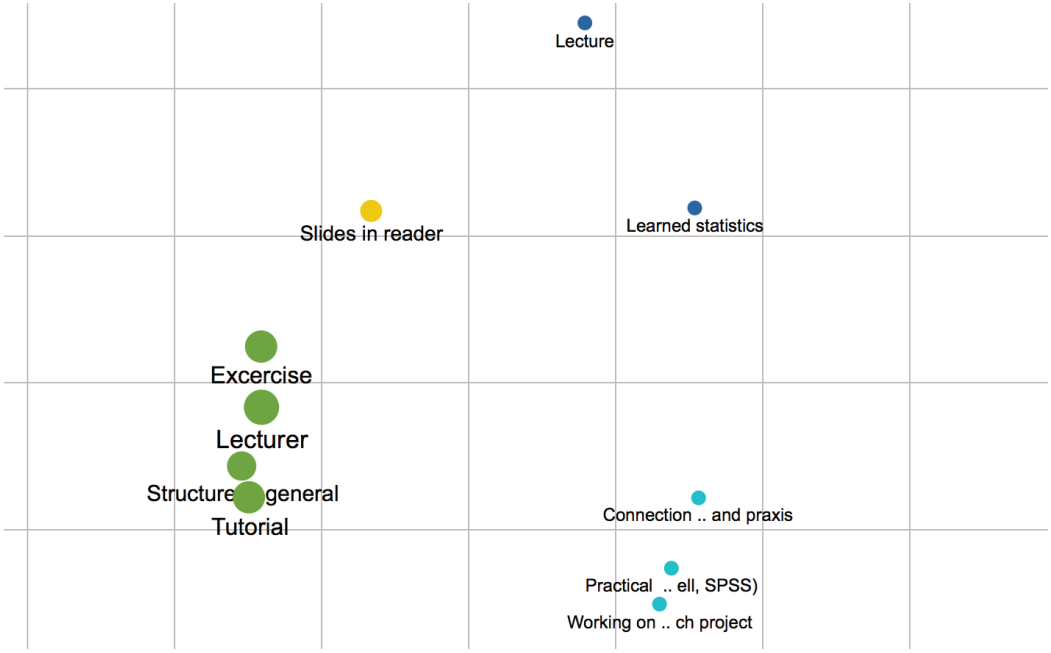


Fig. 7: *Code Map* for placement of categories on a surface based on multidimensional scaling

4.4 Combinations of categories

The analysis described in 4.3 only considers the paired occurrence of categories. Often, however, questions must be answered that relate to the co-occurrence of several categories within a case. For example, in our evaluation study, students indicated that they work through the content using "exercises in the reader" (category 1), the "study group" (category 2), and "textbook" (category 3). The question immediately arose as to what extent these strategies were combined. To answer such questions, the function Analysis > Code Configurations can be used, which can be applied to individual categories as well as to the subcategories of different categories. Fig. 8 shows the result of such an analysis, in which all combinations of categories occurring in the data are listed and sorted according to their frequency. Only seven people, corresponding to 3.6% of the respondents, mentioned all three learning strategies.

	Excercises in reader	Study group	Text book "Bortz"	Documents	Percent	Number of codes
◆				53	27,32	0
◆	■			36	18,56	1
◆		■		22	11,34	1
◆	■		■	20	10,31	2
◆			■	19	9,79	1
◆	■	■		19	9,79	2
◆		■	■	18	9,28	2
◆	■	■	■	7	3,61	3
Σ				194	100,00	12

Fig. 8. *Code Configurations* for the analysis of the co-occurrence of more than two categories

4.5 Identification of groups of cases

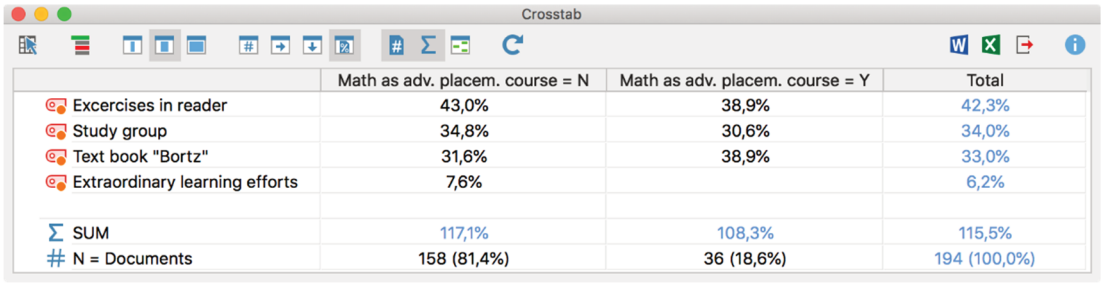
While the Code Map places categories on a surface according to their similarity, the function Visual Tools > Document Map does this for cases. Here, too, the method of multidimensional scaling is applied. The use of this statistical method was also described by Jackson and Trochim (2002) for the analysis of open-ended questions with respect to individual statements. The Document Map is particularly interesting because both the qualitative categories and the standardized answers can be used to determine similarities between cases. It is therefore possible, for example, to include the categories developed for the coding of knowledge acquisition together with age and overall assessment in the calculation of the case positions on the map. The Document Map allows visual identification of groups, by coloring them based on a cluster analysis. The cluster memberships of the cases can be saved as values of a newly formed document variable and this variable can then be used for further analysis, such as for selective retrievals and comparisons of the identified groups.

4.6 Group comparison of category frequencies and category contents

In order to answer the question of the extent to which groups of cases differ with regard to the assignment of categories, cross tabulations can be created in MAXQDA (Mixed Methods > Crosstabs). Fig. 9 shows a crosstab comparing the frequencies of the categories for acquisition of knowledge for two groups – those with and without an advanced placement course in secondary education. The display has been changed to percentage of cases to support an easy comparison of the groups. Like the Code Map and the Document Map, the cross table can be regarded as a joint display in the context of mixed methods data analysis (Creswell & Plano Clark, 2018; Guetterman, 2018; Kuckartz & Rädiker, 2019). The columns are generated using the (quantitative) standardized survey data and the contents in the cells using the (qualitative) coded open-ended questions.

The cross table is also interactively linked to the original data; concrete statements can be output at any time, forming the basis for the category frequencies displayed.

In the sense of quantification (Sandelowski, Voils, & Knafl, 2009; Vogl, 2017) the frequencies of individual categories can be transformed into variables that indicate whether and how often a category has been assigned within a case (Mixed Methods > Quantitizing). A variable created in this way can be used as a grouping criterion for the cross table. Similarly, the cluster memberships created by the use of the Document Map are also suitable for forming the groups to be compared. With the help of MAXQDA Stats, the correlations from the cross table can be analyzed statistically in greater depth, for example to calculate a measure of association for the strength of a correlation or to calculate a chi-square test.



	Math as adv. placem. course = N	Math as adv. placem. course = Y	Total
Excercises in reader	43,0%	38,9%	42,3%
Study group	34,8%	30,6%	34,0%
Text book "Bortz"	31,6%	38,9%	33,0%
Extraordinary learning efforts	7,6%		6,2%
Σ SUM	117,1%	108,3%	115,5%
# N = Documents	158 (81,4%)	36 (18,6%)	194 (100,0%)

Fig. 9. *Crosstab* for group comparison of category frequencies

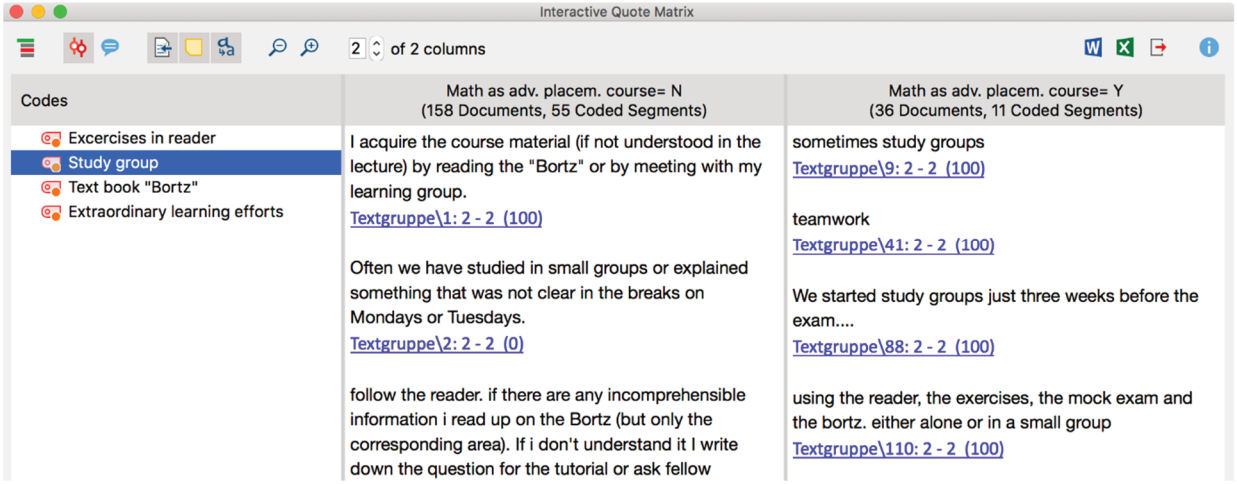


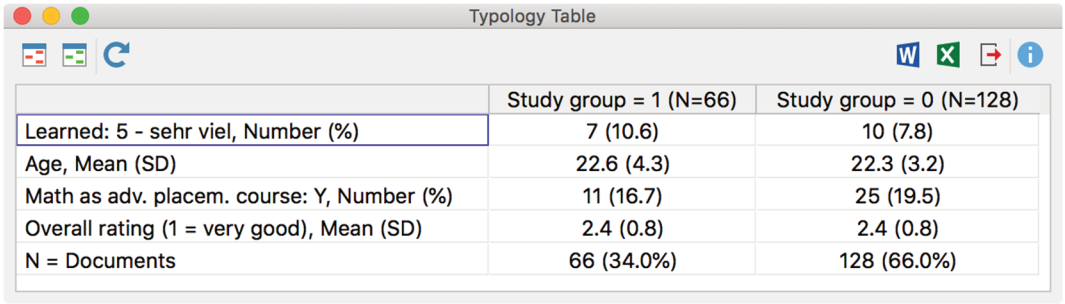
Fig. 10: *Interactive Quote Matrix* for group comparison of category contents

For "simple" open-ended questions on term associations, which usually provide short and sometimes identical answers from respondents, it may be sufficient to compare category frequencies between groups. However, for somewhat more complex open-ended questions, qualitative differences in the answers are also important. If the category "textbook" was coded for two persons, this does not necessarily mean that both persons used the textbook in the same way for working through the content of the course. It is therefore possible to display the answers behind the numbers in a so-called Interactive Quote Matrix (access via ribbon tab Mixed Methods). Fig. 10 shows the same group comparison as in the cross table above: for each group, the coded segments for each category are compared synoptically, making it possible to compare the category contents of the groups. Interactive means in this context that it is possible to switch between categories, to scroll through the columns independently in order to juxtapose any comment from one group with any comment from another, and to "jump" to the answers to other questions of a person at any time (Rädiker, 2018).

It should be noted that a group comparison can be carried out not only for category frequencies and category contents, but also for word frequencies and word combinations. These do not require prior coding and can therefore immediately be used in the exploration phase. Such comparisons allow the investigation of the extent to which older and younger respondents differ in their wording to describe their feelings about statistics.

4.7 Statistics for qualitative groups

Frequently, the coding of free-text responses results in a grouping of cases. These can be simple binary groups where a category has been coded or not (people meeting in learning groups vs. people *not* meeting in learning groups) or more complex typologies, such as those consisting of different learning types. The joint display "Statistics for Qualitative Groups" (from ribbon tab Mixed Methods) calculates statistical parameters of the standardized variables for such groupings and thus allows a comparison of the groups. Fig. 11 shows such a display: for categorical variables proportional values are calculated in percent, for metric variables mean values and standard deviations can be calculated. To make the table easier to interpret, the lowest and highest values per line can be colored.



	Study group = 1 (N=66)	Study group = 0 (N=128)
Learned: 5 - sehr viel, Number (%)	7 (10.6)	10 (7.8)
Age, Mean (SD)	22.6 (4.3)	22.3 (3.2)
Math as adv. placem. course: Y, Number (%)	11 (16.7)	25 (19.5)
Overall rating (1 = very good), Mean (SD)	2.4 (0.8)	2.4 (0.8)
N = Documents	66 (34.0%)	128 (66.0%)

Fig. 11: *Statistics for Qualitative Groups* compares statistical characteristics for two groups, that have been crated on basis of the coding of open-ended questions

5 Overview of Helpful Functions and Procedures

The approach presented in this paper covers many different analysis procedures for data exploration, coding and subsequent category-based analysis with MAXQDA. It can be assumed that, depending on the question and objective of a research project, different priorities will be set in the use of the procedures. For this purpose, Tab. 2 lists all the procedures presented in this article in a systematic form, including a short description and – if suitable – with a typical analysis question that can be answered with the respective function. The table allows the analyst to select the most suitable functions for a specific project. The functions can be roughly divided into word-based and category-based procedures, but there are also mixed forms such as automatic coding. While the word-based methods are primarily used for data exploration, the category-based methods are used for the subsequent (integrative) analysis.

Tab. 2: Functions and procedures for the analysis of answers to open-ended and closed-ended survey questions with MAXQDA

FUNCTION	DESCRIPTION/CENTRAL ANALYSIS QUESTION
Data Exploration	
Standardized variables as quick info per case	*What standardized answers has a person made?* Simultaneous display of free-text answers and selected standardized answers directly at the mouse pointer.
Word frequencies Word combinations Word cloud	*Which words and phrases occur (how) frequently in the free-text answers?* Identification and verification of frequently used expressions in the free-text responses using tabular and visual frequency analysis.
Keyword-in-context Interactive Word Tree	*In which contexts do words and word combinations occur (how often)?* Identification and checking of phrases and expressions in which a search word occurs in the free-text answers.
Coding	
Categorize survey responses	Interactive environment for manual assignment of categories to free-text responses.
Automatic coding of search hits	Assignment of categories to sentences or paragraphs with selected search terms.
Automatic coding using a dictionary	Frequency analyses of categories that are underlaid with search terms (quantitative content analysis). Storage of the results as variable information per case or as coding of sentences or paragraphs with the search terms.
Category-based analysis and presentation of results	
Coding Query	*What did selected persons answer to an open-ended question?* Compilation of the coded free-text answers as a list.
Statistics for Subcodes Code Cloud	*How often do categories occur in an open-ended question?* Frequency table or word cloud for the categories in an open-ended question.
Code Relations Browser Code Map	*Which pairs of categories often appear together?* Visual analysis of pairwise combinations of categories as a similarity matrix and using multidimensional scaling on a surface; interactive coloring of groups using cluster analysis.
Code configurations	*Which categories often occur together?* Tabular listing of multiple combinations of categories.

Document Map	*Which groups of cases can be identified?* Visual analysis of case similarities with respect to categories of free-text responses and standardized characteristics using multidimensional scaling on a surface; interactive coloring of groups using cluster analysis.
Transforming code into variables (quantization) Cross table Stats	*How are category frequencies distributed among the cases and what differences between groups of cases can be identified in this respect?* Group comparisons based on category assignments to free-text responses; univariate, bivariate and multivariate analysis with category frequencies and standardized characteristics.
Interactive Quote Matrix	*What differences in free-text answers can be identified between groups of cases?* Synoptic overview of free-text responses for groups, which are formed on the basis of the standardized characteristics.
Statistics for qualitative groups	*To what extent do groups formed on the basis of the coding of the free-text responses differ in terms of statistical characteristics?* Tabular comparison of mean values and proportional values.

6 Benefits of the (Integrative) Analysis of Survey Data with MAXQDA

It is hardly conceivable to conduct integrative analyses of survey data without software. The main value in the use of MAXQDA lies in the fact that both the qualitative free-text answers and the quantitative standardized answers can be analyzed in parallel and in relation to each other without having to switch back and forth between software packages. The different types of functions and procedures, as summarized in Table 2, make it possible to carry out both word-based analyses more aimed at quantification and analyses that are more focused on content and individual cases. The more data there are to be analyzed, the more word-based procedures are required. If, say, a data set has more than 20 open-ended questions and/or more than 5,000 cases, it can be very challenging to analyze this mass of data using only manual procedures in the software. Depending on the content of the questions, procedures that are more strongly geared towards automation and support (semi-)automatic classification of free-text answers, category formation or word-based analyses are suitable for this purpose (Deneulin & Bavaud, 2016; Popping, 2012; Schonlau & Couper, 2016; Schonlau et al., 2019; Senderovich & Maysuradze, 2015; Waldherr et al., 2019).

Kuckartz & Rädiker (2017) list numerous benefits of using QDA software for the analysis of qualitative data, including faster speed, integration of different types of data, efficient support of teamwork, and the use of

visualizations for analysis and presentation of findings. These advantages also apply to the analysis of surveys with closed-ended and open-ended questions with MAXQDA. Additionally, a few more benefits should be highlighted that are particularly of interest for the (integrative) analysis of free-text responses:

- Because the answers to all open-ended questions of a person are always available, cross-over coding can be carried out without problems. Answers that refer to another question can be coded with the categories of the correct question. This ensures that no information in the free-text responses is left out.
- It is particularly relevant for replication and longitudinal studies that category systems for manual coding, which have proved their worth in a study, can be transferred from one project file to the next project file, including the category definitions. The same applies to dictionaries used in quantitative content analysis. In the case of longitudinal studies, additional data on the same persons can also be added.
- A case always remains accessible as a whole with all its associated qualitative and quantitative data, e.g., when using interactive result tables that allow the analyst to jump directly to the free-text answers of a person. The constant availability of an overall picture of a person supports the reliable coding of statements of a person. Fielding et al. (2013) emphasize the following:

  Whichever of these techniques or functions is used, the essential benefit of using a CAQDAS program to carry out the analysis is that the original words and phrases are always just a couple of mouse clicks away. This could help another researcher verify an analysis by reviewing the application of an important code or after regenerating an output report, and it helps the primary researcher to stay in close proximity to the source material even in the normally abstracted world of a survey. (p. 3272)

Finally, if there is the desire or the necessity to use special statistical procedures in a project, it is possible at any time to export the generated category frequencies as a rectangular matrix “Cases x Categories” and import them into software packages for statistical analysis such as SPSS, STATA, or R.

References

Carey, J. W., Morgan, M., & Oxtoby, M. J. (1996). Intercoder agreement in analysis of responses to open-ended interview questions: Examples from tuberculosis research. *CAM Journal*, *8*(3), 1–5. doi: 10.1177/1525822X960080030101

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: SAGE.

Deneulin, P., & Bavaud, F. (2016). Analyses of open-ended questions by renormalized associativities and textual networks: A study of perception of minerality in wine. *Food Quality and Preference*, *47*, 34–44. doi: 10.1016/j.foodqual.2015.06.013

Fielding, J., Fielding, N., & Hughes, G. (2013). Opening up open-ended survey data using qualitative software. *Quality & Quantity*, *47*(6), 3261–3276. doi: 10.1007/s11135-012-9716-1

Guetterman, T. (2018). *Visual joint displays with graphs and images to represent mixed methods integration.* Presented at the MMIRA International Conference 22.08.2018, Vienna.

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.

Jackson, K. M., & Trochim, W. M. K. (2002). Concept mapping as an alternative approach for the analysis of open-ended survey responses. *Organizational Research Methods*, *5*(4), 307–336. doi: 10.1177/109442802237114

Kelle, U., & Kluge, S. (2010). *Vom Einzelfall zum Typus: Fallvergleich und Fallkontrastierung in der Qualitativen Sozialforschung* (2nd ed.). Wiesbaden: VS Verlag für Sozialwissenschaften.

Kuckartz, U. (2014). *Qualitative text analysis: A guide to methods, practice & using software.* Thousand Oaks, CA: SAGE.

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

Kuckartz, U., & Rädiker, S. (2017). Computergestützte Analyse qualitativer Daten (CAQDAS) in der psychologischen Forschung. In G. Mey & K. Mruck (Eds.), *Handbuch Qualitative Forschung in der Psychologie* (pp. 1–22). doi: 10.1007/978-3-658-18387-5_55-1

Kuckartz, U., & Rädiker, S. (2019). *Analyzing qualitative data with MAXQDA. Text, audio, and video.* Schweiz, Cham: Springer Nature Switzerland.

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution.* Retrieved from https://nbn-resolving.de/urn:nbn:de:0168-ssoar-395173 [18.04.2020]

Montgomery, A. C., & Crittenden, K. S. (1977). Improving coding reliability for open-ended questions. *Public Opinion Quarterly*, *41*(2), 235. doi: 10.1086/268378

Popping, R. (2012). Human or machine coding of open-ended questions. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *115*(1), 79–88. doi: 10/gctq8w

Popping, R. (2015). Analyzing open-ended questions by means of text analysis procedures. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *128*(1), 23–39. doi: 10/gctq8v

Rädiker, S. (2018). *Interactive joint displays in MAXQDA for mixed methods data analysis.* Presented at the MMIRA International Conference 22.08.2018, Vienna.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., … Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. doi: 10.1111/ajps.12103

Saldaña, J. (2015). *The coding manual for qualitative researchers* (3rd ed.). Thousand Oaks, CA: SAGE.

Sandelowski, M., Voils, C., & Knafl, G. (2009). On quantitizing. *Journal of Mixed Methods Research*, *3*(3), 208–222. doi: 10.1177/1558689809334210

Schonlau, M., & Couper, M. P. (2016). Semi-automated categorization of open-ended questions. *Survey Research Methods*, *10*(2), 143–152. doi: 10.18148/srm/2016.v10i2.6213

Schonlau, M., Gweon, H., & Wenemark, M. (2019). Automatic classification of open-ended questions: Check-all-that-apply questions. *Social Science Computer Review*, 1–11. doi: 10.1177/0894439319869210

Schreier, M. (2012). *Qualitative content analysis in practice*. Thousand Oaks, CA: SAGE.

Senderovich, N., & Maysuradze, A. (2015). Interactive coding of responses to open-ended questions in russian. In P. Klinov & D. Mouromtsev (Eds.), *Knowledge engineering and semantic web* (pp. 195–209). Cham: Springer. doi: 10.1007/978-3-319-24543-0_15

Vogl, S. (2017). Quantifizierung: Datentransformation von qualitativen Daten in quantitative Daten in Mixed-Methods-Studien. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *69*(S2), 287–312. doi: 10.1007/S11577-017-0461-2

Waldherr, A., Wehden, L.-O., Stoltenberg, D., Miltner, P., Ostner, S., & Pfetsch, B. (2019). Induktive Kategorienbildung in der Inhaltsanalyse: Kombination automatischer und manueller Verfahren [83 paragraphs]. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 20(1), Art. 19. doi: 10.17169/fqs-20.1.3058

Züll, C., & Mohler, P. P. (2001). *Computerunterstützte Inhaltsanalyse: Codierung und Analyse von Antworten auf offene Fragen (GESIS How-To-Reihe)*. 14.

This guide describes a procedure for analyzing responses to open survey questions with MAXQDA. The procedure does not consider the open-ended questions in isolation from the other data collected, but rather interlinks this qualitative data with the standardized, quantitative data in the sense of a mixed methods approach. The procedure consists of three steps:

1. within-case and cross-case exploration
2. coding the answers
3. category-based analysis

The procedure combines qualitative and quantitative data as well as word-based and category-based procedures and is demonstrated by an example of a mixed methods evaluation of a university course with an online survey. Following this approach, the information on an individual case is always accessible and can be included in a cross-case analysis.

Dr. Stefan Rädiker is a consultant and trainer for research methods and evaluation (www.methoden-expertise.de).

Dr. Udo Kuckartz is a Professor Emeritus of Methods of Social Research in Education at the Philipps-Universität Marburg, Germany.

MAXQDA PRESS

www.maxqda-press.com

ISBN 978-3-948768-02-7