From Coding to Categorical Variables:
The New MAXQDA Function "Transform into a Categorical Variable"

December 2010

Prof. Dr. Udo Kuckartz, Department of Education,

Philipps-Universitaet Marburg, Wilhelm-Roepke-Str. 6 B, 35032 Marburg/Germany;

Email: kuckartz@staff.uni-marburg.de;

Website: www.methoden-evaluation.de

Table of contents

l

Evaluative Categories in Content Analysis

In many research projects, forms of evaluative content analysis are used. The standard steps are to: (1) define evaluative categories, usually with ordinal variables, (2) code text segments, and (3) analyze the data descriptively and statistically. A good example of this form of analysis process is found in Philipp Mayring's chapter on qualitative content analysis found in "A Companion to Qualitative Research" (Flick, et al., 2004), which describes scaled variations of structure content analysis.

In one of Mayring's detailed examples from a study on student teachers, a category called "self esteem" is created with three options: "high," "medium" and "low" (see figure 1). These categories were developed from the material – one can see from the following figure that the categories are not only precisely defined, but also empirically supported with the help of the anchor examples in the material.

The coding process, which is standard procedure for content analysis, has the researcher working through the entire data set and assigning evaluative codes to appropriate text segments that have to do with "self esteem." This means that every single text segment that connects to self esteem will be assigned the code "high," "medium" or "low" on the basis of the coding guidelines established.

At the end, each case (in the case of interviews, a case would be an interviewed person) can be analyzed as a whole and given a summary characterization as having "high," "medium" or "low" self esteem. Cases characterized as having high self esteem can then be compared with those with low self esteem. Code frequencies can also be compared and used in combination with other categories in crosstabs.

Category	Definition	Examples	Coding Rules
C1: high self confidence	High subjective conviction to have successfully coped with the situational demands, which means - to be clear about the demands and their coping possibilities, - to have a positive, hopeful feeling in handling the situation, - to be sure to have coped with the demands on ones own efforts.	"Of course there had been some little problems, but we solved them all, either I myself or the student gave in, depends who made a mistake. Everyone can make mistakes." (17, 23) "Sure there had been problems, but in the end we had a fine relationship. We got it all together." (27, 33)	All three aspects of thew definition have to point to "high" self confidence no aspect only "middle" Otherwise C2: middle self confidence
C2: middle self confidence	Only partly or fluctuating conviction to have successfully coped with the situational demands	"Quite often I found it hard to maneuver through the problems, but finally I made it." (13, 45) "Time by time everything got better , but I couldn't tell if it was me or the circumstances." (77, 20)	If not all aspects of definition point to "High" or "low"
K3: low self concept	Conviction to have badly coped with the situational demands, which means - not to know what the situation exactly demands, - to have a negative, pessimistic feeling in handling the situation, - to be sure that ones own efforts had no effect on improving the situation.	"that stroke my self confidence; I thought I'm a nothing – or even less than that." (5, 34)	All three aspects of definition point to low self confidence, no fluctuations recognizable

Fig. 1 *Coding guidelines for the category "self esteem" (from Mayring 2000: 16)*

Principles of Application in MAXQDA

The method for evaluative content analysis can be done in MAXQDA in the following way. First, the category "self esteem" is created as a code with the subcodes "high," "medium" and "low." The definitions of these codes along with anchor examples can be created as code memos.

Now the material can be worked through, meaning that each document is read line by line. The text segments that have to do with self esteem are identified and then coded with the appropriate code (e.g. "high" self esteem). After an entire text has been worked through in this way, the researcher will have one of the following situations:

- Text segments about self esteem were all coded with the same subcode (e.g. "medium" self esteem). In this situation, the entire case can said to have a medium level of self esteem.
- Text segments about self esteem were coded with various subcodes, but one of those subcodes obviously occurs more often (e.g. three with "high" self esteem and one with "medium" self esteem). In this situation it makes sense to give the whole case that level of self esteem that is coded most often.
- Text segments about self esteem were coded with various subcodes, and none of them clearly occur more often than the others (e.g. two with "medium" self esteem and two with "high" self esteem). In this situation, a quick categorization cannot be made, so the coded segments should be compared to one another by the coders, who then make a decision about which categorization is more appropriate.
- There was no text segment found to connect with self esteem. The entire text had no clues about the person's level of self esteem. In this case, it is not possible to assign a category to the case, and it must be handled as a missing value.

The MAXQDA Function "Transform into a Categorical Variable"

After coding the appropriate text segments, the "self esteem" code can be transformed into a categorical variable by right-clicking on the code and selecting the appropriate option in the menu that appears.
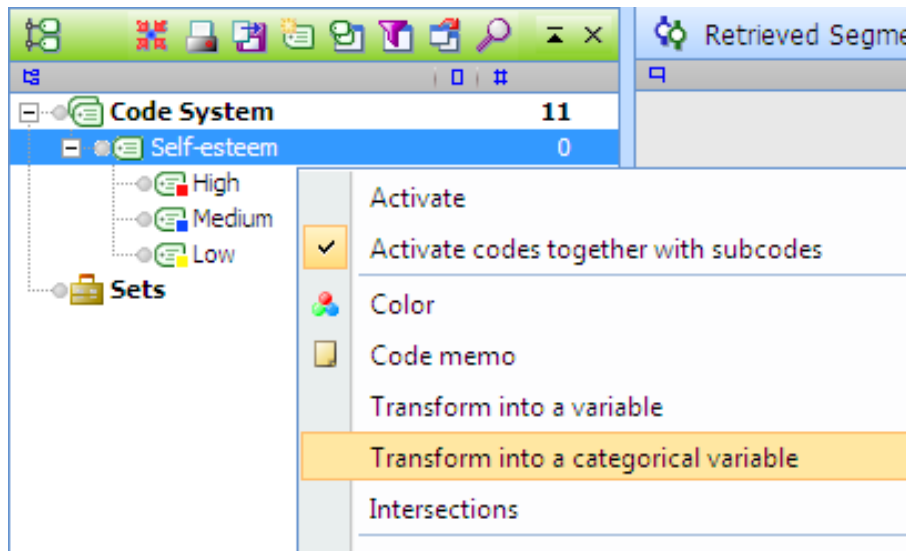
Fig. 2 *"Transform into a categorical variable" option in the context menu*

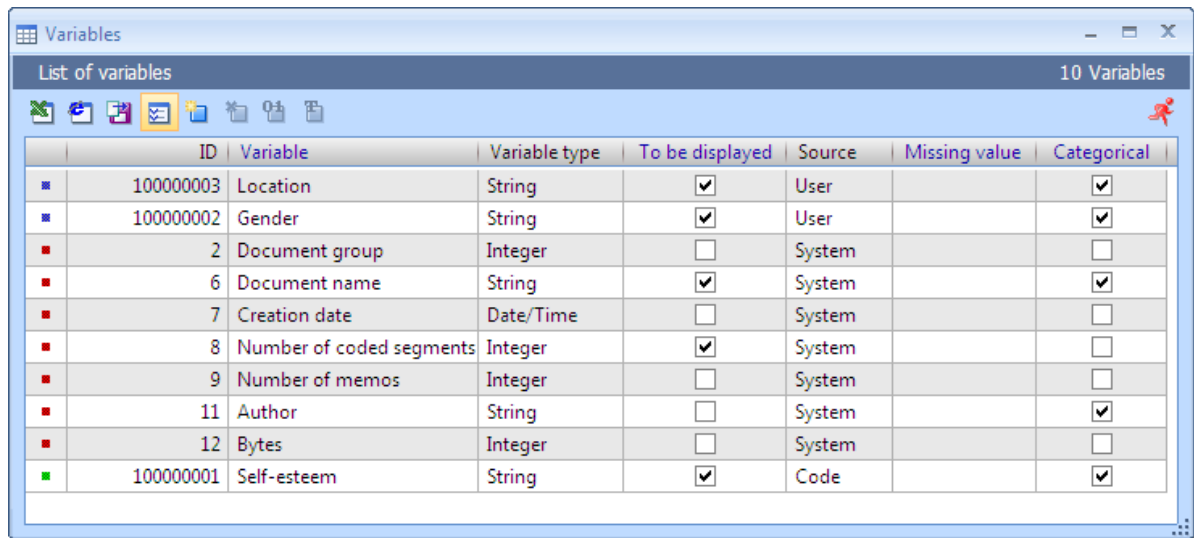After this option is selected, MAXQDA performs the following actions:

1. A new categorical variable is created in the List of Variables with the name of the code that it was created from (in this case, "Self esteem").

2. All cases (documents) in the **Document System** are evaluated according to the rules explained above. That means that each case is assigned as having "high," "medium," "low," or "undefined" self esteem or left blank. If only subcode is used more than the others in a case, that case receives that label. If there are two or more subcodes used the same numbers of time, it is labeled "undefined." If none of the subcodes are used at all, no value is assigned.

Fig. 3 *Spreadsheet with the new categorical variable "Self esteem"*

Dynamic Properties of Categorical Variables

In MAXQDA's **List of Variables**, the categorical variables have a special status. One recognizes them in the list, because they have a green square in the first column und are created from a "code" as seen listed below in the "Source" column; categorical variables are defined as string variables, and the strings are taken from the Code System (in this case "High," "Medium" or "Low").

	ID	Variable	Variable type	To be displayed	Source	Missing value	Categorical
	100000003	Location	String	☑	User		☑
	100000002	Gender	String	☑	User		☑
	2	Document group	Integer	☐	System		☐
	6	Document name	String	☑	System		☑
	7	Creation date	Date/Time	☐	System		☐
	8	Number of coded segments	Integer	☑	System		☐
	9	Number of memos	Integer	☐	System		☐
	11	Author	String	☐	System		☑
	12	Bytes	Integer	☐	System		☐
	100000001	Self-esteem	String	☑	Code		☑

Fig. 4 *The newly-created variable "Self esteem" in the List of Variables*

Categorical variables are dynamic, which means they are updated automatically when new codings are made in the documents. This is also the case for documents that are imported after the categorical variable has already been created; when one codes this new document, the variable label changes accordingly.

Tip for MAXQDA

It makes sense to assign difference colors to the various subcodes (e.g. red for "high" self esteem, blue for "medium" and yellow for "low"). One can then right-click in the coding stripe column, check the box for "Color-coded text", and very quickly find the coded passages in the Document Browser.

Categorical Variables in the Context of Mixed Methods Functions in MAXQDA

Categorical variables lend themselves very well to use with MAXQDA's mixed methods functions. With the "Logical activation" function, for example, one can choose to activate only those cases (documents) with a certain variable value. This is helpful for answering

research questions such as "How do student teachers with low self esteem experience their situation in the school system? How do they approach disciplines issues?"

The Crosstabs function offers an aggregated overview of the number of codings in certain categories in the Code System in each of the three levels of the self esteem variable. The self esteem characterizations are shown in the columns on the x-axis, and the other specified categories are shown in the rows on the y-axis. Using the self esteem example, the Crosstabs function could count the number of times that student teachers with low self esteem talk about discipline issues in comparison to the number of times student teachers with high self esteem talk about the issue. With a single mouse click, one can also call up the document segments counted in each cell in the **Retrieved Segments** window.

One can also use the Quote Matrix to see a detailed table of the document segments, each column holding those segments that occur in documents with a specific categorical value. In our self esteem example, one column could hold those statements about a certain topic that come from student teachers with high self esteem, and the other column could display those statements from teachers with low self esteem.

The Typology table uses categorical variables similar to the way the Crosstabs function does; a table is created with the categorical variable values in the columns. In this case, however, the variables are analyzed rather than the categories. One could look, for example, what percentage of people with high self esteem are men and what percentage are women, or whether good grades in teacher certification exams seem to connect in any way to self esteem, etc.

Literature

Creswell, John W. and Plano Clark, Vicki L. (2010). Designing and Conducting Mixed Methods Research (Chapter 7, 203 ff.). London: Sage.

Kuckartz, Udo (2010). Einführung in die computergestützte Analyse qualitativer Daten. 3., akt. u. erw. Aufl., Wiesbaden.

Mayring, Philipp (2000). Qualitative Content Analysis [28 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(2), Art. 20, http://nbn-resolving.de/urn:nbn:de:0114-fqs0002204.

Mayring, Philipp (2004). Qualitative content analysis. In U. Flick, E.v. Kardorf & I. Steinke (Eds.), A companion to qualitative research (pp. 266 - 269). London: Sage.

Mayring, Philipp (2007). On Generalization in Qualitatively Oriented Research [23 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *8*(3), Art. 26, http://nbn-resolving.de/urn:nbn:de:0114-fqs0703262.