

Qualitative Content Analysis is the method that helps you summarize the meaning of qualitative data using a coding frame. It helps you to boil down a high amount of information to its most important core while it only picks up distinct concepts (mutually exclusive) and covers all aspects present in the data (collectively exhaustive). Qualitative Content Analysis reduces and summarizes the data which makes it different from other qualitative methods that aim at enriching or interpreting data.

The first main advantage of Qualitative Content Analysis is that it can be applied to a wide range of data: songs, speeches, social media posts, pictures, interviews, newspaper articles, journal entries and others. The second main advantage is that it can help you to translate qualitative data into quantitative data so that you can apply statistical analysis. It is particularly useful to solve specific problems where only the core information is needed especially in areas such as Marketing & Sales, Human Resources, Brand Management, Product Development, Quality Management and Qualitative Benchmarking.

## Table of Contents

- [Application & Impact: Summarize the core of what is meant](#)
  - [Marketing & Sales: Understand what customers mean](#)
  - [Human Resources: Figure out what employees value](#)
  - [Brand Management: Know how your company is perceived](#)
  - [Competition and qualitative benchmarking: Compare your brand with others qualitatively](#)
  - [Product Development: Analyze expert opinions on the most important trends](#)
  - [PR: What is the public picture of your company](#)
  - [Quality Management: Boil down customer feedback to its core](#)
- [Procedure: A detailed step-by-step guide](#)
  - [Step 1: Collecting Data](#)
  - [Step 2: Building a Coding Frame](#)
  - [Step 3: Segmenting your material](#)
  - [Step 4: Try out your coding Frame](#)
  - [Step 5: Evaluate and modify your coding frame](#)
  - [Step 6: Code all your material](#)
  - [Step 7: Interpret the result](#)
- [Advantages - Ensured objectivity by strict rules](#)
- [Disadvantages - There are better methods for theory building](#)
- [Further Links](#)

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!

- [References](#)
- [Share this:](#)

## **Application & Impact: Summarize the core of what is meant**

Qualitative Content Analysis can be used in all areas where a big amount information needs to be summarized to its core what can range from analyzing technologies to analyzing how the own company is perceived in public. The main limitation of Qualitative Content Analysis is availability of data as you need ideally a wide range and a high amount of data to apply it consistently. Here are a few examples, where it might be applied:

### **Marketing & Sales: Understand what customers mean**

- What are the reasons customers buy a certain product?
- What product features do customers value?

### **Human Resources: Figure out what employees value**

- What is the most important thing that keeps our employees at the company?
- How is our company perceived according to the comments on Glassdoor?

### **Brand Management: Know how your company is perceived**

- What are the adjectives used to describe our brand?
- How does our brand differ from other brands?

### **Competition and qualitative benchmarking: Compare your brand with others qualitatively**

- How does our profile fit within the industry and how does it differ?
- What features distinguish my company from other companies in the customers' eyes?

### **Product Development: Analyze expert opinions on the most important trends**

- How is a certain technology, like AI, influencing the business and what application potentials are there?
- What are the main complaints of our customers for our current products?

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!

**PR: What is the public picture of your company**

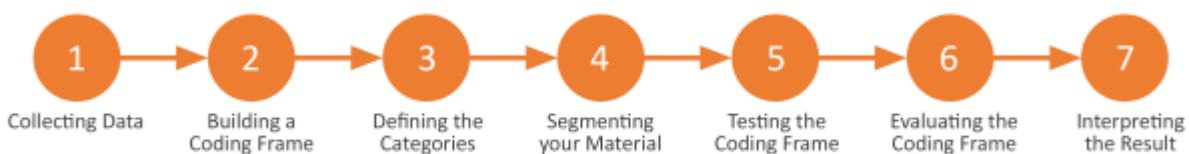
- What is said in the newspapers about my company and with what frequency?
- What adjectives and verbs are used with reference to my new product?

**Quality Management:****Boil down customer feedback to its core**

- What do our customer suggest in order to improve the product and what aspects are mentioned most often?
- What are the aspects mentioned for our product in the reviews online for instance at Amazon?

**Procedure: A detailed step-by-step guide**

Conducting a Qualitative Content Analysis involves 7 steps involving roughly making use of a coding frame, generating category definitions, segmenting the material into coding units, and distinguishing between a pilot phase and a main phase of analysis. It is very important that you have a clearly defined research question before you start. Furthermore, keep in mind that for qualitative methods, it is not as straightforward and easy to access reliability and validity as for quantitative methods. Therefore, the quality of the result is assessed by other mean as for instance by consistency and a systematic approach. They ensure that your results are valid, reliable and especially, credible.

**Qualitative Content Analysis Steps****Step 1: Collecting Data**

After you have formulated a clear research question, the first step would be to gather data that reflects ideally the full diversity of your research topic. You will also need to clarify whether you want to use one form of data, for instance only newspaper articles, or several ones, for instance videos and books. My general recommendation is that you try to focus on one form of data as it will be more consistent and easier to segment the material later on.

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!

However, depending on the focus, you might also place greater focus on diversity of your data source rather than consistency. To give you one example, if I had to analyze product reviews on Amazon for the Apple iPhone X, I would select the material that I can find online on Amazon as simple as it is and not focus on other sources, as the research question would be to identify how is the iPhone X perceived in the Amazon reviews.



Customer  
Feedback



Books &  
Stories



News &  
Articles



Webpages &  
Blogs



Videos &  
Movies



Pictures &  
Images

Qualitative Content Analysis can combine many different sources.

## Step 2: Building a Coding Frame

Building a valid and reliable coding frame is the most critical and trickiest part of the whole analysis. Coding frame is at the heart of any Qualitative Content Analysis and specifies all the different meanings that you want to capture and distinguish in your analysis. Therefore, it is especially important to be cautious here and to make right decisions.

Coding frames consist of at least one main category and at least two subcategories. They can vary in complexity and consist of any number of main categories, contain several

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!

hierarchical levels and even subcategories within subcategories. However, since the main goal is usually to understand and communicate the results from the Qualitative Content Analysis, I recommend it to keep it simple and to avoid adding further subcategories to subcategories.

The main categories represent the more abstract aspects of your material that are of interest for you. The subcategories cover concretely what is actually mentioned within that specific main category. You can think of a main category being like a variable in statistics and subcategories of the values that this variable can take on. For instance, hair color could be a main category when we want to summarize the physical aspects of people and the corresponding subcategories might be brunette, blond or black.

There are further three requirements for a coding frame to work.

1. **Mutually exclusive:** First, subcategories need to be mutually exclusive, that means that there is no overlap in the aspects of what they cover. Within a main category, only one subcategory can be assigned to one unit of analysis. Look at the following example: you want to understand what your competition is and therefore a company that you compete against is your unit of analysis. You have a main category "Company Type" which entails the subcategory "IT", "Startup", "Manufacturing" and "Medium-Sized". If you encounter an IT-Startup, you would need to assign the subcategory "Startup" and "IT", which violates the requirement of mutual exclusivity. Bad coding frame!
2. **Unidimensional:** The example from the first requirement leads us to the second one. A main category should cover only one aspect of the material. In the previous example, "Company Type" would cover two aspects, which would be the industry and the company stage. Thus, you would need to split your main category into these two categories to fulfill this criterion.
3. **Collectively exhaustive:** The main categories altogether must be cover all aspects present in the material. First, this means that for each unit of analysis, meaning company in our example, you can find a suitable subcategory within each of your main categories. Second, this means that all aspects within your text is covered by the main categories. In our example, we can consider the new coding frame of "Industry" and "Company Stage" still uncomplete, because there are other important aspects, for instance "Product Type" (Service or Product), "Location", and "Employee Number". You can easily introduce residual categories. However, introducing too many residual categories might tell you that your coding frame is not valid enough and that you should redo it. If subcategories are very similar, it might be best to collapse them.

If you meet these requirements, then you laid the first stone towards a good coding frame. A good coding frame is reliable and valid. Reliable means that even others can understand as well apply your coding frame and ideally recreate your results. Valid means that your coding frame captures all important aspects of your material so that all relevant sections of your material can be assigned to a main category or subcategory. Constructing a coding frame is not difficult and it also entails three essential steps:

**Step 1 - Selecting the material:** You select a part of your material (for instance around 50%) and ideally the most diverse parts, that you will use to build the coding frames. That way you ensure that your coding frames cover ideally all important aspects present in the whole 100% of your material. My recommendation is that you do not try to build your whole coding frame at once. It might make sense to focus on one aspect at a time. That way you will make sure that you do not miss out any important aspects and that you build a coding frame that is consistent.

**Step 2 - Creating the categories:** When you create the categories, you have three possibilities on how to start, depending on whether you decide to do it in a data-driven way, concept-driven way or a mix of both:

- **Data-driven way:** You start out without any predetermined categories and you let the coding frame fully emerge from the material. The data-driven way might be particularly interesting if you are rather interested in knowing, what aspects at all are mentioned or if there is no clear idea of what aspects are covered.
- **Concept-driven way:** You have clear ideas about the aspects that are covered and what concrete information you will find with respect to your research question, so you start out with a completely defined coding frame with predetermined categories and subcategories that you derived from a valid theory. You basically skip the part of developing the coding frame.
- **Mixed way:** The problem of the fully concept-driven way is that your initial idea of the categories may actually differ from what is actually present in the material and therefore, leave part of the material unaccounted for. This will make your coding frame invalid and maybe even less reliable. Therefore, the mixed-way is often applied, where you have an initial idea of the aspects in the material but you allow the coding frame to adapt to the material.

In case you decide to work in a data-driven or mixed way, you will have again several strategies on how to develop the categories from the material. The two most often practiced strategies that will help you derive the main categories as well as subcategories systematically are subsumption and progressive summarizing.

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!

- **Subsumption:** This method is very effective in generating the subcategories in a data-driven way after the main categories have already been set. You basically examine one passage of your material after another and iterate through the following steps. First, you read through the material until a relevant concept is encountered. Second, you check whether a subcategory that covers the concept you encountered has already been created. If this concept is new, you create a new subcategory that covers this concept. Third, you continue reading until you encounter the next relevant concept. You basically continue with these steps until you have the feeling that your coding frame reached a point of saturation, e.g. you do not encounter any new concepts.
- **Progressive summarizing:** This method is very powerful when you want to develop the whole coding frame in data-driven way including main categories and subcategories. Using these methods, you basically paraphrase relevant passages, delete from these passages anything that appears unnecessary, and you keep on summarizing similar paraphrases. Finally, after paraphrasing them several times, they will turn into categories and subcategories.

### Step 3 - Defining the categories:

In Qualitative Content Analysis, it is very important that your categories are clear even to other people and it should be clear what you mean by a given category. This is very important for the reliability of your coding frame, because when it is not clear what you mean by a given category, people will not be able to understand your coding frame and tend to assign passages to different subcategories than you. And this is a big problem not only because it will be difficult to present your results, but they will also be simply less credible if they are repeatedly misunderstood by other people.

This is why you write a definition for each main category and subcategory. For main categories, the definitions can be short, but for subcategories they should be more extensive. A definition should always include the following elements:

- **Label:** a category name that is intuitive and straight-forward
- **Description:** a short text that explains what you mean by the name and what criteria should a text passage mean in order to be assigned to this category
- **Example:** ideally a sample or a quote (in case your material is text-based) that illustrates the category and your procedure of assigning it to it
- **Decision rule:** a decision rule in case it is still not 100% clear on how to deal with the given category

Here you should to make sure that subcategories within one main category are indeed

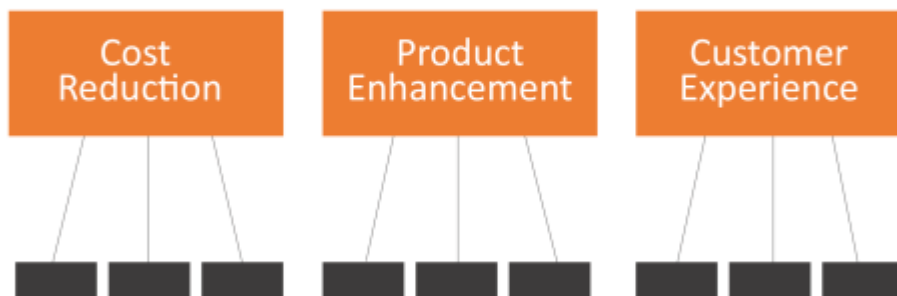
Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!



mutually exclusive. Especially for this requirement, decision rules might be very useful.

You should remember that the development of a coding frame is not a linear process. It might mean that you will need to go back to earlier steps, if you are not successful while summarizing it. From my personal experience, it is of crucial importance that you know what your research question is. If you for instance analyze how digital transformation influences the company, then you will find out that the same research question might lead to very different coding frames depending on how you interpret the question. See the following two examples:

**Scenario 1:** Coding frame focuses on the possibilities emerging through digital transformation



**Scenario 2:** Coding frame focuses on areas that are impacted by digital transformation



Two different coding scenarios based on digital transformation

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!



Here you answered the very same question with probably the very same material with entirely two different coding frames. Therefore you should have a clear idea what your goal is.

### **Step 3: Segmenting your material**

When you segment your material, you divide it into several chunks also called units of coding. Segmentation is especially important for coding, because you will need to assign a subcategory from each main category to every unit of analysis you have and it is the basis for the comparison later when two different people code. You will need to choose the units of analysis in such a way that they can be interpreted in a meaningful way with respect to the subcategories.

<overview text passage, codes, coding sheet>



### Segmenting the material into units of coding

In order to segment for instance several images, you can simply take each image as unit of analysis. If your material consists of newspapers, you can decide each newspaper article to be a segment. In order to segment the material properly, it might be necessary to define criteria specifying when a segment should start and when one should end. There are two types of such criteria.

1. **Formal criteria:** Formal criteria are based on the structure of your material. To say in simpler terms, if you have for instance a book as material, then you might decide that each chapter might be a unit of analysis, or that each paragraph should be a unit of analysis or even each page. You basically define the rule based on how it is structured.
2. **Thematic criteria:** Thematic criteria might often be more useful and includes finding places when topics change. In thematic criteria, the unit of analysis corresponds to a

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!

theme. For instance, your material might be interviews from customers that gave feedback to your product. In that case, you will segment the material whenever the general topic of the interview changes as customers talk about the appearance of product in the first unit of coding, then go over to how they used to product in the next unit of coding and so on.

Is there an advantage of formal criteria over thematic criteria and vice versa? Definitely. Thematic criteria have the advantage that one unit of coding will correspond to one particular topic and depending on the structure of the material. It might make your material even more valid when for instance formal criteria would segment the material in such a way, that one unit of analysis covers several topics. This might produce a conflict because one unit of coding would fit very well to two subcategories within a main category. Thematic criteria will avoid this information loss and make sure that your coding frame is more representative of the material. Furthermore, sometimes your research question simply favors thematic criteria. If you are interested in conflicts within a specific book, then it simply does not make sense to structure that book according to formal criteria.

On the other side, formal criteria have the advantage, that they are clear, understandable and fast. It is very easy to segment your material according to formal criteria and it is hardly ambiguous. That means, even when other people were to segment your material, they will derive the same units of coding. Likewise as for thematic coding, some research questions will already favor formal criteria such as how chapters differ from each other within a book or what are the most frequent aspects mentioned by my customers in the product reviews.

When you segment the material, you usually assign a number to each unit of coding consecutively. If you use formal criterion, you can skip the extra step of segmenting your material and it can be done in parallel with coding. If you have thematic criteria, you will need to segment the material first before you code it. Furthermore, at the end you should derive a coding sheet that you will use to code your material. The columns contain the categories, the rows contain the units of coding and in each cell you will write down the subcategory for the respective main category and unit of coding. Your coding sheet should look like this:

	Main Category 1	Main Category 2	Main Category 3	...
Unit 1				
Unit 2				
...				

If you have quantitative background, you will realize that your coding sheet will resemble a dataset consisting of only categorical variables. This actually is the way Qualitative Content Analysis can help you translate qualitative information into quantitative information, so that you can run statistical analysis on the data. After you completed all the following steps and filled out your coding sheet, you could for instance compute frequencies or correlation coefficients between different categories.

#### **Step 4: Try out your coding Frame**

After you developed your coding frame, you will need to test it out on part of the material that is ideally different from the material that you developed the coding frame on. Ideally the material selected should again cover all types of data and all aspects you anticipate to find in your data. In this step, you want to assess the quality of your coding frame before you start to apply it on the whole data. Therefore you want to check for reliability and validity of your coding frame. This you will do in the next step.

#### **Step 5: Evaluate and modify your coding frame**

##### **Assess the reliability of the coding frame**

Reliability describes to what extent your coding frame is reproducible and generalizable. In order to evaluate the reliability, you will need to double-code the material using the same coding-frame. That means that first find a second person that will help you and you apply the same coding frame on the same data at the same time independently from each other. If you will need to work alone. It is also possible to you code the material twice yourself, but make a two-week break between the two runs. If the definitions of subcategories are clear enough and if the subcategories are indeed mutually exclusive, then the units of coding should be assigned to the same subcategories by both people, you and your partner.

After you have completed the double-coding, then you assess the reliability of your coding frame. This can be done in two ways:

1. **Informally:** Your partner and you simply exchange your reasons for putting certain units of coding into different categories. That way you will find out where your coding frame is not clear enough because the description is unclear or because the subcategories overlap and where your partner has misunderstood the way you use the coding frame.
2. **Quantitatively:** You compute a coefficient of agreement for instance simply the percentage of agreement or kappa.

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!

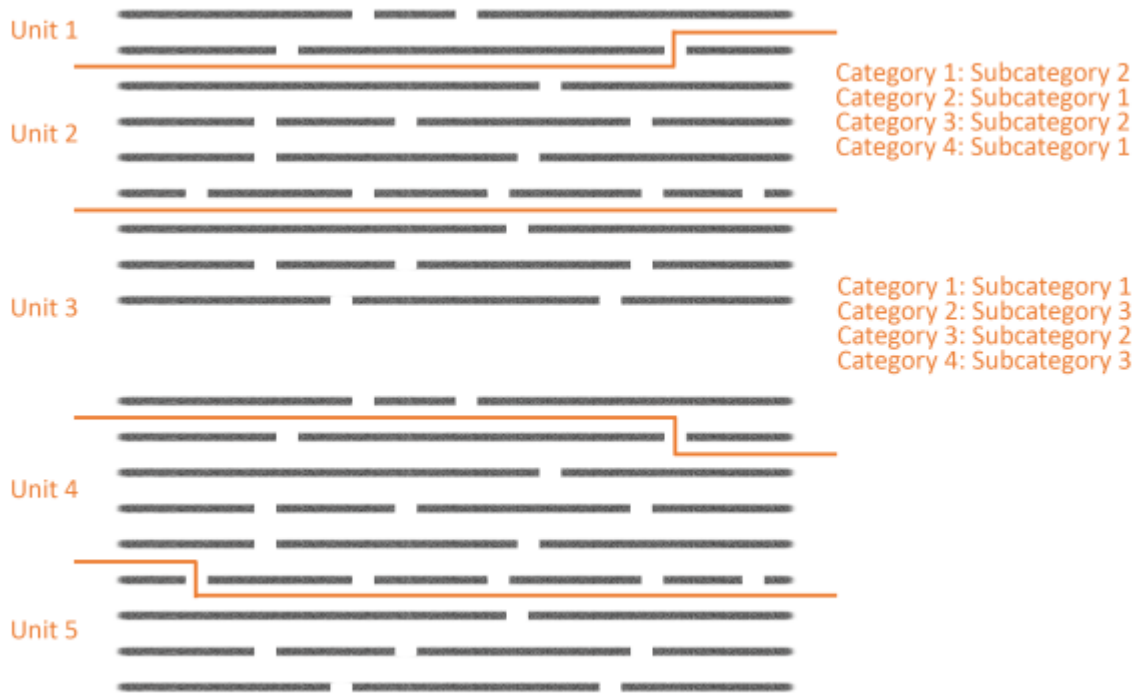
In order to achieve reliability, it is very important to keep the complexity and scope of your coding frame as small necessary. Coding frames that consist of more than 200 categories will be more likely to lead to errors by both coders as well as to disagreement. One possibility to handle larger coding frames is to not code all main categories at once, but do it consecutively.

### **Assess the validity of the coding frame**

Validity tells you the degree to which your categories cover the material and relevant concepts present in your material. For data-driven parts of your coding frame, you can assess the validity of your material first by checking whether all units of coding of your material fits into one subcategory of every main category. Second, you look whether you needed to introduce residual categories. Too many residual categories tell you that your coding frame is not valid enough. Third, you check whether you have used a subcategory much more often than others and whether certain subcategories have not been used at all. If this is the case, it might be better to split the most frequent categories into smaller more precise ones.

### **Step 6: Code all your material**

After you have adopted and revised your coding frame, it is time to code all your material. If your coding frame proved to be already sufficiently valid and reliable, you can code all your material alone without double-coding. If you needed to adapt your frame, it will be important that you double-code again in order to make sure that your newly produced coding frame is reliably and valid. However, it is not necessary to double-code everything. It might be enough to just double code around one fourth of your material. Generally, the more changes you had to make in Step 5, the more you should double-code.



Assigning values to units of coding

Finally, there is one more trick. If you have shown that your coding frame is reliable, you can divide up the unit of coding among several people and split the work. It works, because a reliable coding frame will produce the same results regardless of who is coding.

**Step 7: Interpret the result**

After you have completed Step 6, you will ideally end up with a completed coding sheet. If your goal was to analyze customer reviews for your products, then your coding sheet might look the following way:

	<b>Complaints</b>	<b>Suggestions</b>	<b>Likes</b>
<b>Customer 1</b>	Long Delivery	Open shop	Great quality
<b>Customer 2</b>	Long Delivery	Open shop	Great quality
<b>Customer 3</b>	Too expensive	Offer loyalty bonus	

At this point, depending on your goal and research questions, there might be several possibilities on how to deal with results. Generally, there are three possibilities:

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!

1. **Present the coding frame:** If the aim of the Qualitative Content Analysis was just to find out, what aspects are present in your material, then you can simply present your coding frame with the main categories and subcategories. Show what categories are in your coding frame and explain what the categories mean for your client and his specific problem.
2. **Analyze the coding sheet:** If the aim of your research was to understand patterns, you might want to look at the coding sheet and compute frequencies. Furthermore, you might look and try to understand why a particular unit of coding was assigned to certain subcategories and whether there might be any hidden patterns.
3. **Continue with further analysis:** Often Qualitative Content Analysis is used as a preparatory step for further analysis. You might use the coding sheet for more sophisticated statistical analysis. You might want to cluster the units of coding to identify similar ones or you might want to compare units of coding to find out how they differ. There is a wide range of possibilities that you could do based on the coding frame and coding sheet.

## Advantages - Ensured objectivity by strict rules

The Qualitative Content Analysis is a very flexible method that has these advantages:

- **Translation of qualitative data to quantitative data:** It is a method that can translate qualitative data into quantitative data so you can for instance test hypotheses on qualitative data
- **Availability of material:** It is generally relatively easy to find material for a Qualitative Content Analysis because there is already lots of text, pictures, and videos readily online available.
- **Relatively objective:** Qualitative Content Analysis is guided by strict rules and helps you eliminate subjectivity. It offers the ability to calculate a reliability score when you double-code your material. This reliability score can be also considered an objectivity score.
- **Data preparation:** After having conducted a Qualitative Content Analysis, you will end up having a coding frame and a coding sheet that you can use for various kinds of further analysis: regression analysis, clustering, identifying patterns, qualitative benchmarking.
- **Assessment of validity and reliability:** It is not typical for a qualitative methods to offer concrete ways to assess validity and reliability. This makes Qualitative Content Analysis special.
- **Extracting core content:** The method will help you boil down your material to the

Economalytics is the analytics blog that shows you in plain and simple which methods are available and how you can use these methods to solve your problem. Check out [www.economalytics.com](http://www.economalytics.com) for more!



most important aspect without adding information, twisting it or enriching it.

## Disadvantages - There are better methods for theory building

Despite these advantages, Qualitative Content Analysis also has some clear disadvantages:

- **Data reduction:** The main function of the method is also one of the main disadvantage. For Qualitative Content Analysis it is very important to already have a clear research question and to be experienced with a topic to some degree. Even though you have the same research question and the same material, you might still end up with many different coding frames.
- **Experience:** Even though it is a fairly straightforward method with clear and guided rules, it still requires some experience in order to build a reliable and valid coding frame.
- **Interpretation & theory building:** Qualitative Content Analysis is not suitable for interpreting material or deriving a theory, because its main focus is to describe your material but not understand it. Other methods like Coding or Discourse Analysis would be more suitable for that.
- **Singularity:** Qualitative Content Analysis does not allow you to explore several possible meanings at once and how these relate to each other. A better method for this purpose would be Semiotics.

## Further Links

In case you want to read more theoretical articles on Qualitative Content Analysis, I recommend that you look have a look at the books and articles listed in the references. Here further two articles of how Qualitative Content Analysis is applied in Marketing.

- <https://www.thebalancesmb.com/how-to-analyze-qualitative-market-research-data-2296916>
- <https://arxiv.org/ftp/arxiv/papers/1505/1505.04401.pdf>

The first one is rather theoretical and aims at giving you an overview, while the second one is more practical and shows how Qualitative Content Analysis can complement a quantitative Analysis.

If you have further questions, criticism, you need help or you have other ideas on how one could apply Qualitative Content Analysis, feel free to leave comment or to drop me a line.

## References

The SAGE Handbook of Qualitative Data Analysis - Uwe Flick.

Mayring, Philipp (2010) Qualitative Inhaltsanalyse: Grundlagen und Techniken.

Berger, Arthur A. (2000) Content Analysis, in Arthur Berger (ed.) media and Communications Research Methods. Thousand Oaks: Sage pp. 173-85.

Hsieh, Hsiu-Fang and Shannon, Sarah E. (2005) "Three approaches to qualitative content analysis", Qualitative Health Research, 15: 1277-88.

Krippendorff, Klaus (2004) Content Analysis: An Introduction to its methodology. Thousand Oaks, CA: Sage (1<sup>st</sup> edition, 1980).

Schreier, Margrit (2012) Qualitative Content Analysis In Practice. London: Sage.

## Share this:

- [Click to print \(Opens in new window\)](#)
- [Click to share on Facebook \(Opens in new window\)](#)
- [Click to share on LinkedIn \(Opens in new window\)](#)
- [Click to share on Twitter \(Opens in new window\)](#)
- [Click to share on WhatsApp \(Opens in new window\)](#)