

**WHAT IT IS**[Return to Table of Contents](#)

Content analysis is a methodology for determining the content of written, recorded, or published communications via a systematic, objective, and quantitative procedure. Thus, it is a set of procedures for collecting and organizing information in a standard format that allows analysts to draw inferences about the characteristics and meaning of recorded material.

**WHEN TO USE IT**

Content analysis can be used to make numerical comparisons among and within documents. It is especially useful for tabulating the results of open-ended survey questions and multiple interviews. It can also be used to analyze entity documentation to determine compliance with laws, rules, policies, and procedures; to clarify trends in agency activity; to assess alignment between such activity and stated goals, objectives, and strategies; or to examine differences between groups within the entity on of issues of interest.

**HOW TO PREPARE IT**

- Determine whether content analysis is appropriate. Selection of content analysis as a method will depend on project objectives and issues, availability of accurate recorded material, and the kinds of comparisons sought.
- Identify the universe of content. That is, determine what material should be included in the analysis. Consider what content material exists, the types and locations of samples you will obtain, and what time frame(s) is of interest. Such material can be in any media yielding information on which reanalysis can be performed. Note that sampling may be needed if the universe of material is too extensive to be reviewed in its entirety. (See the module on [Sampling](#) for additional information.)

Most commonly, written media are explored. Appropriate written media include plans, policies, procedures, reports, work samples, correspondence, memoranda, staff or client records, responses to interviews or questionnaires, agendas, minutes, resumes, job descriptions, etc. Possible other media include audio and video recordings, television and radio programs, movies, and photographs.

- Obtain examples of the content to be analyzed. One distinction between content analysis and other analyses is that the development of the analysis instrument is highly dependent on the content to be analyzed. Thus, be sure that the examples you obtain represent the content. Note that bad examples of the content sought may be just as important as good ones.
- Review content examples to determine if the information available meets the objectives of the analysis. Identify and explore alternate sources, as needed. You may also need to redefine or expand the universe of content.
- Select the coding units, i.e. the unit(s) of analysis which best capture data (words, phrases, ideas, etc.) In the body of work being analyzed. The five types of coding units most common in content analysis are:

- Words, the smallest possible unit. Key words are easily identified and are generally reliably classified by different content analysts. However, words can have different meanings in different contexts.
- Themes, the next larger unit. A theme may be a simple sentence or single idea. Though thematic analysis is commonly used, it can be quite difficult, particularly when the material analyzed contains complex passages. The problem is basically one of reliability. That is, if two content analysts were to divide content using the same thematic unit, they might easily disagree on where the divisions should be made.
- Character, in which the content is divided each time a person or object is mentioned. This analysis unit is quite useful for identifying relationships and spans of control between individuals within the entity.
- Item, i.e. the whole unit in which the content was originally generated. For example, if one were analyzing responses to open-ended questions, the entire response might be the item. Item analysis is useful when variations within items are small or unimportant. If the item becomes too complex, another coding unit should be considered.
- Space-and-time measures which relate to the amount of space or time displaced by the content. Examples include number of pages, column inches in periodicals, number of minutes in audio or video recordings, number of frames in movies, etc.

Note that one need not be limited to one coding unit. It may often be appropriate to use more than one unit on the same content to get at different issues of interest. Moreover, note that while choosing the coding unit precedes selecting the coding categories, selecting coding categories may require changing the coding unit(s). Finally, be aware that the unit of analysis is important later when quantitative inferences are made about the nature and scope of the content analyzed.

- Develop coding categories. Content analysis stands or falls by its categories, just as a questionnaire stands or falls by its questions. Category systems can be roughly divided into two classes -- those dealing with what is said and those dealing with how it is said. Good category systems are:
  - exhaustive, so that all relevant content can be categorized
  - mutually exclusive, so that data can be placed in one and only one category
  - independent, so that the assignment of data to one category does not affect the assignment of other data to a category

Developing coding categories, along with the actual coding of content information, is the most important part of content analysis. Among other key considerations are that:

- Issues related to the objectives of the analysis should be developed and listed.
- If a large number of issues exist, it may be helpful to roll them up into more general categories.
- One should develop codes for different reactions to the issues which meet the above requirements. This may often resemble a rating system for responses. Here it is important that content coders rate each coding unit separately on each of the different categories or variables. For example, the SAO conducts an open-ended survey of Child Protective Services field staff to determine their assessment of available automation. The following table indicates a possible coding scheme:

ISSUE	AGREE	DISAGREE	DON'T KNOW
<u>1. Adequacy of automation</u>  Adequate Needs improvement Inadequate			
<u>2. Availability of automation</u>  Availability is limited			

This table, although simplistic, provides an example of quantifying and scaling responses. The level of scaling can be more or less detailed. More detail (such as adding "strongly agree" or "strongly disagree") allows for more specific analysis but can complicate reliability.

- Code the material. Conduct a pre-test of the suitability of the categories, the coding instructions, and the reliability of coding among coders. It is quite common that changes, additions, or deletions are needed in categories, coding units, content universe, or even objectives. Additional, previously unforeseen issues may also surface. Given these problems, pre-testing the content analysis system is generally needed.

Since a primary reason for content analysis is to collect objective information, it is important that the procedure is consistently applied. This consistency is called *interrater reliability* and refers to the degree to which multiple coders yield identical results. Assessing reliability is extremely important in content analysis, given the high potential for subjective interpretation of responses by those performing the coding function. This problem increases as response scales are developed to measure intensity (e.g., strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree). Such intensity scales

require coders to make often minor distinctions in interpreting the data being analyzed.

Determining coding reliability is a two-stage process. First, one must ensure that the classification system is understood and applied to each unit separately. Then, one must compare how multiple coders classify a sample of the same material. Ideally, units of analysis would receive the same categorization by different coders. This may not be the case, however. The acceptable range of reliability is generally 80 to 100 percent. This means that two independent coders would code a group of items exactly the same at least 80 percent of the time. This is also determined during the pre-test of the system on a sample of the content to be analyzed. Low reliability should be investigated and may arise from:

- unclear definition of categories
- incompetent coders
- inadequate coder training
- inadequate instructions on the use of codes or categories
- unclear or overly subtle distinctions between categories
- substandard media (e.g., illegible text, poor quality audio, etc.)

The following table provides some sample responses to questions about the adequacy of automation issue discussed above. The reader should consider how he or she might code them.

Question: How would you assess the level of automation in your office?	
# 21	Overall, the system is pretty good. The information we have access to is very helpful. But, updates to the data base are batch processed. If the system allowed real-time updates this would give case workers more up-to-date information.
# 22	The terminals hurt my eyes. They are hard to look at for long periods of time. The system is very slow retrieving information, and much of the information we get is miscoded or incorrect. Terminal time is also limited because of frequent down time for repairs. I can't say the current system helps me much in doing my job.
# 23	There is a serious lack of printers. Case workers need to be able to print out information on our terminals, but this is time consuming because we all have to send output to one printer. Lots of days the printer is offline for repairs which complicates the situation.

- Revise the content analysis system, as needed. Proceed with coding all relevant content material obtained.
- Analyze and interpret the results. Summarize the coded data, review the data for patterns and relationships, and relate the results to data obtained from other methods. Write up the results. Add the results summary and all documentation relevant to content analysis system design and application to the project working papers.

### **A NOTE ON TYPES OF CONTENT ANALYSIS DATA**

Content analysis often yields frequencies of response by category and coding unit. Frequencies help compare elements in the content universe and are easily depicted with graphs. If intensity ratings are used, descriptive statistics, such as the mean, median, standard deviation, percentiles, and interquartile range, can also be used. It may be helpful to present frequency results and intensity results separately, especially if a subset of categories or respondents is of special interest. (See the modules on [Descriptive Statistics](#), [Frequency Distributions](#), [Graphs](#), [Interviews](#), and [Questionnaires/Surveys](#) for more information.)

### **ADVANTAGES**

Content analysis can:

- quantify largely qualitative information
- facilitate unobtrusive measurement
- cope with large volumes of source material
- help analysts learn more about issues of interest
- add qualitative richness to otherwise quantitative data
- validate evidence from other sources
- apply to virtually any project

### **DISADVANTAGES**

Content analysis can:

- be costly and time-consuming
- pose reliability and validity problems
- be challenged as too subjective