



## Writing@CSU Writing Guide

# Content Analysis

This Writing Guide was downloaded from the Writing@CSU Web Site at Colorado State University on June 27, 2016 at 8:29 AM. You can view the guide at <http://writing.colostate.edu/guides/guide.cfm?guideid=61>. Copyright information and a citation can be found at the end of this document.

## Main Page

This guide provides an introduction to content analysis, a research methodology that examines words or phrases within a wide range of texts.

**Introduction to Content Analysis:** Read about the history and uses of content analysis.

**Conceptual Analysis:** Read an overview of conceptual analysis and its associated methodology.

**Relational Analysis:** Read an overview of relational analysis and its associated methodology.

**Commentary:** Read about issues of reliability and validity with regard to content analysis as well as the advantages and disadvantages of using content analysis as a research methodology.

**Examples:** View examples of real and hypothetical studies that use content analysis.

**Annotated Bibliography:** Complete list of resources used in this guide and beyond.

## An Introduction to Content Analysis

Content analysis is a research tool used to determine the presence of certain words or concepts within texts or sets of texts. Researchers quantify and analyze the presence, meanings and relationships of such words and concepts, then make inferences about the messages within the texts, the writer(s), the audience, and even the culture and time of which these are a part. Texts can be defined broadly as books, book chapters, essays, interviews, discussions, newspaper headlines and articles, historical documents, speeches, conversations, advertising, theater, informal conversation, or really any occurrence of communicative language. Texts in a single study may also represent a

variety of different types of occurrences, such as Palmquist's 1990 study of two composition classes, in which he analyzed student and teacher interviews, writing journals, classroom discussions and lectures, and out-of-class interaction sheets. To conduct a content analysis on any such text, the text is coded, or broken down, into manageable categories on a variety of levels--word, word sense, phrase, sentence, or theme--and then examined using one of content analysis' basic methods: conceptual analysis or relational analysis.

## **A Brief History of Content Analysis**

Historically, content analysis was a time consuming process. Analysis was done manually, or slow mainframe computers were used to analyze punch cards containing data punched in by human coders. Single studies could employ thousands of these cards. Human error and time constraints made this method impractical for large texts. However, despite its impracticality, content analysis was already an often utilized research method by the 1940's. Although initially limited to studies that examined texts for the frequency of the occurrence of identified terms (word counts), by the mid-1950's researchers were already starting to consider the need for more sophisticated methods of analysis, focusing on concepts rather than simply words, and on semantic relationships rather than just presence (de Sola Pool 1959). While both traditions still continue today, content analysis now is also utilized to explore mental models, and their linguistic, affective, cognitive, social, cultural and historical significance.

## **Uses of Content Analysis**

Perhaps due to the fact that it can be applied to examine *any* piece of writing or occurrence of recorded communication, content analysis is currently used in a dizzying array of fields, ranging from marketing and media studies, to literature and rhetoric, ethnography and cultural studies, gender and age issues, sociology and political science, psychology and cognitive science, and many other fields of inquiry. Additionally, content analysis reflects a close relationship with socio- and psycholinguistics, and is playing an integral role in the development of artificial intelligence. The following list (adapted from Berelson, 1952) offers more possibilities for the uses of content analysis:

- Reveal international differences in communication content
- Detect the existence of propaganda
- Identify the intentions, focus or communication trends of an individual, group or institution
- Describe attitudinal and behavioral responses to communications
- Determine psychological or emotional state of persons or groups

## **Types of Content Analysis**

In this guide, we discuss two general categories of content analysis: conceptual analysis

and relational analysis. Conceptual analysis can be thought of as establishing the existence and frequency of concepts most often represented by words or phrases in a text. For instance, say you have a hunch that your favorite poet often writes about hunger. With conceptual analysis you can determine how many times words such as hunger, hungry, famished, or starving appear in a volume of poems. In contrast, relational analysis goes one step further by examining the relationships among concepts in a text. Returning to the hunger example, with relational analysis, you could identify what other words or phrases hunger or famished appear next to and then determine what different meanings emerge as a result of these groupings.

## Conceptual Analysis

Traditionally, content analysis has most often been thought of in terms of conceptual analysis. In conceptual analysis, a concept is chosen for examination, and the analysis involves quantifying and tallying its presence. Also known as thematic analysis [although this term is somewhat problematic, given its varied definitions in current literature--see Palmquist, Carley, & Dale (1997) vis-a-vis Smith (1992)], the focus here is on looking at the occurrence of selected terms within a text or texts, although the terms may be implicit as well as explicit. While explicit terms obviously are easy to identify, coding for implicit terms and deciding their level of implication is complicated by the need to base judgments on a somewhat subjective system. To attempt to limit the subjectivity, then (as well as to limit problems of reliability and validity), coding such implicit terms usually involves the use of either a specialized dictionary or contextual translation rules. And sometimes, both tools are used--a trend reflected in recent versions of the Harvard and Lasswell dictionaries.

## Methods of Conceptual Analysis

Conceptual analysis begins with identifying research questions and choosing a sample or samples. Once chosen, the text must be coded into manageable content categories. The process of coding is basically one of selective reduction. By reducing the text to categories consisting of a word, set of words or phrases, the researcher can focus on, and code for, specific words or patterns that are indicative of the research question.

An example of a conceptual analysis would be to examine several Clinton speeches on health care, made during the 1992 presidential campaign, and code them for the existence of certain words. In looking at these speeches, the research question might involve examining the number of positive words used to describe Clinton's proposed plan, and the number of negative words used to describe the current status of health care in America. The researcher would be interested only in quantifying these words, not in examining how they are related, which is a function of relational analysis. In conceptual analysis, the researcher simply wants to examine presence with respect to his/her research question, i.e. is there a stronger presence of positive or negative words used with respect to proposed or current health care plans, respectively.

Once the research question has been established, the researcher must make his/her coding choices with respect to the eight category coding steps indicated by Carley (1992).

## Steps for Conducting Conceptual Analysis

The following discussion of steps that can be followed to code a text or set of texts during conceptual analysis use campaign speeches made by Bill Clinton during the 1992 presidential campaign as an example. To read about each step, click on the items in the list below:

### **Decide the level of analysis.**

First, the researcher must decide upon the level of analysis. With the health care speeches, to continue the example, the researcher must decide whether to code for a single word, such as "inexpensive," or for sets of words or phrases, such as "coverage for everyone."

### **Decide how many concepts to code for.**

The researcher must now decide how many different concepts to code for. This involves developing a pre-defined or interactive set of concepts and categories. The researcher must decide whether or not to code for every single positive or negative word that appears, or only certain ones that the researcher determines are most relevant to health care. Then, with this pre-defined number set, the researcher has to determine how much flexibility he/she allows him/herself when coding. The question of whether the researcher codes only from this pre-defined set, or allows him/herself to add relevant categories not included in the set as he/she finds them in the text, must be answered. Determining a certain number and set of concepts allows a researcher to examine a text for very specific things, keeping him/her on task. But introducing a level of coding flexibility allows new, important material to be incorporated into the coding process that could have significant bearings on one's results.

### **Decide whether to code for existence or frequency of a concept.**

After a certain number and set of concepts are chosen for coding, the researcher must answer a key question: is he/she going to code for existence or frequency? This is important, because it changes the coding process. When coding for existence, "inexpensive" would only be counted once, no matter how many times it appeared. This would be a very basic coding process and would give the researcher a very limited perspective of the text. However, the number of times "inexpensive" appears in a text might be more indicative of importance. Knowing that "inexpensive" appeared 50 times, for example, compared to 15 appearances of "coverage for everyone," might lead a researcher to interpret that Clinton is trying to sell his health care plan based more on economic benefits, not comprehensive coverage. Knowing that "inexpensive" appeared, but not that it appeared 50 times, would not allow the researcher to make this interpretation, regardless of whether it

is valid or not.

### **Decide on how you will distinguish among concepts.**

The researcher must next decide on the level of generalization, i.e. whether concepts are to be coded exactly as they appear, or if they can be recorded as the same even when they appear in different forms. For example, "expensive" might also appear as "expensiveness." The research needs to determine if the two words mean radically different things to him/her, or if they are similar enough that they can be coded as being the same thing, i.e. "expensive words." In line with this, is the need to determine the level of implication one is going to allow. This entails more than subtle differences in tense or spelling, as with "expensive" and "expensiveness." Determining the level of implication would allow the researcher to code not only for the word "expensive," but also for words that imply "expensive." This could perhaps include technical words, jargon, or political euphemism, such as "economically challenging," that the researcher decides does not merit a separate category, but is better represented under the category "expensive," due to its implicit meaning of "expensive."

### **Develop rules for coding your texts.**

After taking the generalization of concepts into consideration, a researcher will want to create translation rules that will allow him/her to streamline and organize the coding process so that he/she is coding for exactly what he/she wants to code for. Developing a set of rules helps the researcher insure that he/she is coding things consistently throughout the text, in the same way every time. If a researcher coded "economically challenging" as a separate category from "expensive" in one paragraph, then coded it under the umbrella of "expensive" when it occurred in the next paragraph, his/her data would be invalid. The interpretations drawn from that data will subsequently be invalid as well. Translation rules protect against this and give the coding process a crucial level of consistency and coherence.

### **Decide what to do with "irrelevant" information.**

The next choice a researcher must make involves irrelevant information. The researcher must decide whether irrelevant information should be ignored (as Weber, 1990, suggests), or used to reexamine and/or alter the coding scheme. In the case of this example, words like "and" and "the," as they appear by themselves, would be ignored. They add nothing to the quantification of words like "inexpensive" and "expensive" and can be disregarded without impacting the outcome of the coding.

### **Code the texts.**

Once these choices about irrelevant information are made, the next step is to code the text. This is done either by hand, i.e. reading through the text and manually writing down concept occurrences, or through the use of various computer programs. Coding with a computer is one of contemporary conceptual analysis' greatest assets. By inputting one's categories, content analysis programs can easily

automate the coding process and examine huge amounts of data, and a wider range of texts, quickly and efficiently. But automation is very dependent on the researcher's preparation and category construction. When coding is done manually, a researcher can recognize errors far more easily. A computer is only a tool and can only code based on the information it is given. This problem is most apparent when coding for implicit information, where category preparation is essential for accurate coding.

### **Analyze your results.**

Once the coding is done, the researcher examines the data and attempts to draw whatever conclusions and generalizations are possible. Of course, before these can be drawn, the researcher must decide what to do with the information in the text that is not coded. One's options include either deleting or skipping over unwanted material, or viewing all information as relevant and important and using it to reexamine, reassess and perhaps even alter one's coding scheme. Furthermore, given that the conceptual analyst is dealing only with quantitative data, the levels of interpretation and generalizability are very limited. The researcher can only extrapolate as far as the data will allow. But it is possible to see trends, for example, that are indicative of much larger ideas. Using the example from step three, if the concept "inexpensive" appears 50 times, compared to 15 appearances of "coverage for everyone," then the researcher can pretty safely extrapolate that there does appear to be a greater emphasis on the economics of the health care plan, as opposed to its universal coverage for all Americans. It must be kept in mind that conceptual analysis, while extremely useful and effective for providing this type of information when done right, is limited by its focus and the quantitative nature of its examination. To more fully explore the relationships that exist between these concepts, one must turn to relational analysis.

## **Relational Analysis**

Relational analysis, like conceptual analysis, begins with the act of identifying concepts present in a given text or set of texts. However, relational analysis seeks to go beyond presence by exploring the relationships between the concepts identified. Relational analysis has also been termed semantic analysis (Palmquist, Carley, & Dale, 1997). In other words, the focus of relational analysis is to look for semantic, or meaningful, relationships. Individual concepts, in and of themselves, are viewed as having no inherent meaning. Rather, meaning is a product of the relationships among concepts in a text. Carley (1992) asserts that concepts are "ideational kernels;" these kernels can be thought of as symbols which acquire meaning through their connections to other symbols.

## Theoretical Influences on Relational Analysis

The kind of analysis that researchers employ will vary significantly according to their theoretical approach. Key theoretical approaches that inform content analysis include linguistics and cognitive science.

**Linguistic approaches** to content analysis focus analysis of texts on the level of a linguistic unit, typically single clause units. One example of this type of research is Gottschalk (1975), who developed an automated procedure which analyzes each clause in a text and assigns it a numerical score based on several emotional/psychological scales. Another technique is to code a text grammatically into clauses and parts of speech to establish a matrix representation (Carley, 1990).

Approaches that derive from **cognitive science** include the creation of decision maps and mental models. Decision maps attempt to represent the relationship(s) between ideas, beliefs, attitudes, and information available to an author when making a decision within a text. These relationships can be represented as logical, inferential, causal, sequential, and mathematical relationships. Typically, two of these links are compared in a single study, and are analyzed as networks. For example, Heise (1987) used logical and sequential links to examine symbolic interaction. This methodology is thought of as a more generalized cognitive mapping technique, rather than the more specific mental models approach.

**Mental models** are groups or networks of interrelated concepts that are thought to reflect conscious or subconscious perceptions of reality. According to cognitive scientists, internal mental structures are created as people draw inferences and gather information about the world. Mental models are a more specific approach to mapping because beyond extraction and comparison because they can be numerically and graphically analyzed. Such models rely heavily on the use of computers to help analyze and construct mapping representations. Typically, studies based on this approach follow five general steps:

- Identifying concepts
- Defining relationship types
- Coding the text on the basis of 1 and 2
- Coding the statements
- Graphically displaying and numerically analyzing the resulting maps

To create the model, a researcher converts a text into a map of concepts and relations; the map is then analyzed on the level of concepts and statements, where a statement consists of two concepts and their relationship. Carley (1990) asserts that this makes possible the comparison of a wide variety of maps, representing multiple sources, implicit and explicit information, as well as socially shared cognitions.

## Relational Analysis: Overview of Methods

As with other sorts of inquiry, initial choices with regard to what is being studied and/or coded for often determine the possibilities of that particular study. For relational analysis,

it is important to first decide which concept type(s) will be explored in the analysis. Studies have been conducted with as few as one and as many as 500 concept categories. Obviously, too many categories may obscure your results and too few can lead to unreliable and potentially invalid conclusions. Therefore, it is important to allow the context and necessities of your research to guide your coding procedures.

The steps to relational analysis that we consider in this guide suggest some of the possible avenues available to a researcher doing content analysis. We provide an example to make the process easier to grasp. However, the choices made within the context of the example are but only a few of many possibilities. The diversity of techniques available suggests that there is quite a bit of enthusiasm for this mode of research. Once a procedure is rigorously tested, it can be applied and compared across populations over time. The process of relational analysis has achieved a high degree of computer automation but still is, like most forms of research, time consuming. Perhaps the strongest claim that can be made is that it maintains a high degree of statistical rigor without losing the richness of detail apparent in even more qualitative methods.

### **Three Subcategories of Relational Analysis**

**Affect extraction:** This approach provides an emotional evaluation of concepts explicit in a text. It is problematic because emotion may vary across time and populations. Nevertheless, when extended it can be a potent means of exploring the emotional/psychological state of the speaker and/or writer. Gottschalk (1995) provides an example of this type of analysis. By assigning concepts identified a numeric value on corresponding emotional/psychological scales that can then be statistically examined, Gottschalk claims that the emotional/psychological state of the speaker or writer can be ascertained via their verbal behavior.

**Proximity analysis:** This approach, on the other hand, is concerned with the co-occurrence of explicit concepts in the text. In this procedure, the text is defined as a string of words. A given length of words, called a *window*, is determined. The window is then scanned across a text to check for the co-occurrence of concepts. The result is the creation of a concept determined by the *concept matrix*. In other words, a matrix, or a group of interrelated, co-occurring concepts, might suggest a certain overall meaning. The technique is problematic because the window records only explicit concepts and treats meaning as proximal co-occurrence. Other techniques such as clustering, grouping, and scaling are also useful in proximity analysis.

**Cognitive mapping:** This approach is one that allows for further analysis of the results from the two previous approaches. It attempts to take the above processes one step further by representing these relationships visually for comparison. Whereas affective and proximal analysis function primarily within the preserved order of the text, cognitive mapping attempts to create a model of the overall meaning of the text. This can be represented as a graphic map that represents the relationships between concepts.



In this manner, cognitive mapping lends itself to the comparison of semantic connections across texts. This is known as map analysis which allows for comparisons to explore "how meanings and definitions shift across people and time" (Palmquist, Carley, & Dale, 1997). Maps can depict a variety of different *mental models* (such as that of the text, the writer/speaker, or the social group/period), according to the focus of the researcher. This variety is indicative of the theoretical assumptions that support mapping: mental models are representations of interrelated concepts that reflect conscious or subconscious perceptions of reality; language is the key to understanding these models; and these models can be represented as networks (Carley, 1990). Given these assumptions, it's not surprising to see how closely this technique reflects the cognitive concerns of socio-and psycholinguistics, and lends itself to the development of artificial intelligence models.

### **Steps for Conducting Relational Analysis**

The following discussion of the steps (or, perhaps more accurately, strategies) that can be followed to code a text or set of texts during relational analysis. These explanations are accompanied by examples of relational analysis possibilities for statements made by Bill Clinton during the 1998 hearings.

#### **Identify the Question.**

The question is important because it indicates where you are headed and why. Without a focused question, the concept types and options open to interpretation are limitless and therefore the analysis difficult to complete. Possibilities for the Hairy Hearings of 1998 might be:

What did Bill Clinton say in the speech? **OR** What concrete information did he present to the public?

#### **Choose a sample or samples for analysis.**

Once the question has been identified, the researcher must select sections of text/speech from the hearings in which Bill Clinton may have not told the entire truth or is obviously holding back information. For relational content analysis, the primary consideration is how much information to preserve for analysis. One must be careful not to limit the results by doing so, but the researcher must also take special care not to take on so much that the coding process becomes too heavy and extensive to supply worthwhile results.

#### **Determine the type of analysis.**

Once the sample has been chosen for analysis, it is necessary to determine what type or types of relationships you would like to examine. There are different subcategories of relational analysis that can be used to examine the relationships in texts.

In this example, we will use proximity analysis because it is concerned with the co-occurrence of explicit concepts in the text. In this instance, we are not particularly interested in affect extraction because we are trying to get to the hard facts of what

exactly was said rather than determining the emotional considerations of speaker and receivers surrounding the speech which may be unrecoverable.

Once the subcategory of analysis is chosen, the selected text must be reviewed to determine the level of analysis. The researcher must decide whether to code for a single word, such as "perhaps," or for sets of words or phrases like "I may have forgotten."

### **Reduce the text to categories and code for words or patterns.**

At the simplest level, a researcher can code merely for existence. This is not to say that simplicity of procedure leads to simplistic results. Many studies have successfully employed this strategy. For example, Palmquist (1990) did not attempt to establish the relationships among concept terms in the classrooms he studied; his study did, however, look at the change in the presence of concepts over the course of the semester, comparing a map analysis from the beginning of the semester to one constructed at the end. On the other hand, the requirement of one's specific research question may necessitate deeper levels of coding to preserve greater detail for analysis.

In relation to our extended example, the researcher might code for how often Bill Clinton used words that were ambiguous, held double meanings, or left an opening for change or "re-evaluation." The researcher might also choose to code for what words he used that have such an ambiguous nature in relation to the importance of the information directly related to those words.

### **Explore the relationships between concepts (Strength, Sign & Direction).**

Once words are coded, the text can be analyzed for the relationships among the concepts set forth. There are three concepts which play a central role in exploring the relations among concepts in content analysis.

**Strength of Relationship:** Refers to the degree to which two or more concepts are related. These relationships are easiest to analyze, compare, and graph when all relationships between concepts are considered to be equal. However, assigning strength to relationships retains a greater degree of the detail found in the original text. Identifying strength of a relationship is key when determining whether or not words like unless, perhaps, or maybe are related to a particular section of text, phrase, or idea.

**Sign of a Relationship:** Refers to whether or not the concepts are positively or negatively related. To illustrate, the concept "bear" is negatively related to the concept "stock market" in the same sense as the concept "bull" is positively related. Thus "it's a bear market" could be coded to show a negative relationship between "bear" and "market". Another approach to coding for strength entails the creation of separate categories for binary oppositions. The above example emphasizes "bull" as the negation of "bear," but could be coded as being two separate categories, one positive and one negative. There has been little research to determine the benefits and liabilities of these differing strategies. Use of Sign coding for relationships in regard to the hearings may be to find out whether or not the words under observation or in question were used adversely or in favor of the concepts (this is tricky, but important to establishing meaning).

**Direction of the Relationship:** Refers to the type of relationship categories

exhibit. Coding for this sort of information can be useful in establishing, for example, the impact of new information in a decision making process. Various types of directional relationships include, "X implies Y," "X occurs before Y" and "if X then Y," or quite simply the decision whether concept X is the "prime mover" of Y or vice versa. In the case of the 1998 hearings, the researcher might note that, "maybe implies doubt," "perhaps occurs before statements of clarification," and "if possibly exists, then there is room for Clinton to change his stance." In some cases, concepts can be said to be bi-directional, or having equal influence. This is equivalent to ignoring directionality. Both approaches are useful, but differ in focus. Coding all categories as bi-directional is most useful for exploratory studies where pre-coding may influence results, and is also most easily automated, or computer coded.

### **Code the relationships.**

One of the main differences between conceptual analysis and relational analysis is that the statements or relationships between concepts are coded. At this point, to continue our extended example, it is important to take special care with assigning value to the relationships in an effort to determine whether the ambiguous words in Bill Clinton's speech are just fillers, or hold information about the statements he is making.

### **Perform Statistical Analyses.**

This step involves conducting statistical analyses of the data you've coded during your relational analysis. This may involve exploring for differences or looking for relationships among the variables you've identified in your study.

### **Map out the Representations.**

In addition to statistical analysis, relational analysis often leads to viewing the representations of the concepts and their associations in a text (or across texts) in a graphical -- or map -- form. Relational analysis is also informed by a variety of different theoretical approaches: linguistic content analysis, decision mapping, and mental models.

## **Commentary**

The authors of this guide have created the following commentaries on content analysis.

### **Issues of Reliability & Validity**

The issues of reliability and validity are concurrent with those addressed in other research methods. The reliability of a content analysis study refers to its *stability*, or the tendency for coders to consistently re-code the same data in the same way over a period of time; *reproducibility*, or the tendency for a group of coders to classify categories membership in the same way; and *accuracy*, or the extent to which the classification of a text corresponds to a standard or norm statistically. Gottschalk (1995) points out that the

issue of reliability may be further complicated by the inescapably human nature of researchers. For this reason, he suggests that coding errors can only be minimized, and not eliminated (he shoots for 80% as an acceptable margin for reliability).

On the other hand, the validity of a content analysis study refers to the correspondence of the *categories* to the *conclusions*, and the *generalizability* of results to a theory.

The validity of *categories* in implicit concept analysis, in particular, is achieved by utilizing multiple classifiers to arrive at an agreed upon definition of the category. For example, a content analysis study might measure the occurrence of the concept category "communist" in presidential inaugural speeches. Using multiple classifiers, the concept category can be broadened to include synonyms such as "red," "Soviet threat," "pinkos," "godless infidels" and "Marxist sympathizers." "Communist" is held to be the explicit variable, while "red," etc. are the implicit variables.

The overarching problem of concept analysis research is the challenge-able nature of *conclusions* reached by its inferential procedures. The question lies in what level of implication is allowable, i.e. do the conclusions follow from the data or are they explainable due to some other phenomenon? For occurrence-specific studies, for example, can the second occurrence of a word carry equal weight as the ninety-ninth? Reasonable conclusions can be drawn from substantive amounts of quantitative data, but the question of proof may still remain unanswered.

This problem is again best illustrated when one uses computer programs to conduct word counts. The problem of distinguishing between synonyms and homonyms can completely throw off one's results, invalidating any conclusions one infers from the results. The word "mine," for example, variously denotes a personal pronoun, an explosive device, and a deep hole in the ground from which ore is extracted. One may obtain an accurate count of that word's occurrence and frequency, but not have an accurate accounting of the meaning inherent in each particular usage. For example, one may find 50 occurrences of the word "mine." But, if one is only looking specifically for "mine" as an explosive device, and 17 of the occurrences are actually personal pronouns, the resulting 50 is an inaccurate result. Any conclusions drawn as a result of that number would render that conclusion invalid.

The *generalizability* of one's conclusions, then, is very dependent on how one determines concept categories, as well as on how reliable those categories are. It is imperative that one defines categories that accurately measure the idea and/or items one is seeking to measure. Akin to this is the construction of rules. Developing rules that allow one, and others, to categorize and code the same data in the same way over a period of time, referred to as *stability*, is essential to the success of a conceptual analysis.

*Reproducibility*, not only of specific categories, but of general methods applied to establishing all sets of categories, makes a study, and its subsequent conclusions and results, more sound. A study which does this, i.e. in which the classification of a text corresponds to a standard or norm, is said to have *accuracy*.

## Advantages of Content Analysis

Content analysis offers several advantages to researchers who consider using it. In particular, content analysis:

- looks directly at communication via texts or transcripts, and hence gets at the central aspect of social interaction
- can allow for both quantitative and qualitative operations
- can provide valuable historical/cultural insights over time through analysis of texts
- allows a closeness to text which can alternate between specific categories and relationships and also statistically analyzes the coded form of the text
- can be used to interpret texts for purposes such as the development of expert systems (since knowledge and rules can both be coded in terms of explicit statements about the relationships among concepts)
- is an unobtrusive means of analyzing interactions
- provides insight into complex models of human thought and language use

## Disadvantages of Content Analysis

Content analysis suffers from several disadvantages, both theoretical and procedural. In particular, content analysis:

- can be extremely time consuming
- is subject to increased error, particularly when relational analysis is used to attain a higher level of interpretation
- is often devoid of theoretical base, or attempts too liberally to draw meaningful inferences about the relationships and impacts implied in a study
- is inherently reductive, particularly when dealing with complex texts
- tends too often to simply consist of word counts
- often disregards the context that produced the text, as well as the state of things after the text is produced
- can be difficult to automate or computerize

## Examples

The Palmquist, Carley and Dale study, a summary of "Applications of Computer-Aided Text Analysis: Analyzing Literary and Non-Literary Texts" (1997) is an example of two studies that have been conducted using both conceptual and relational analysis. The Problematic Text for Content Analysis shows the differences in results obtained by a conceptual and a relational approach to a study.

### Related Information: Example of a Problematic Text for Content Analysis

In this example, both students observed a scientist and were asked to write about the experience.

Student A: I found that scientists engage in research in order to make discoveries and generate new ideas. Such research by scientists is hard work

and often involves collaboration with other scientists which leads to discoveries which make the scientists famous. Such collaboration may be informal, such as when they share new ideas over lunch, or formal, such as when they are co-authors of a paper.

Student B: It was hard work to research famous scientists engaged in collaboration and I made many informal discoveries. My research showed that scientists engaged in collaboration with other scientists are co-authors of at least one paper containing their new ideas. Some scientists make formal discoveries and have new ideas.

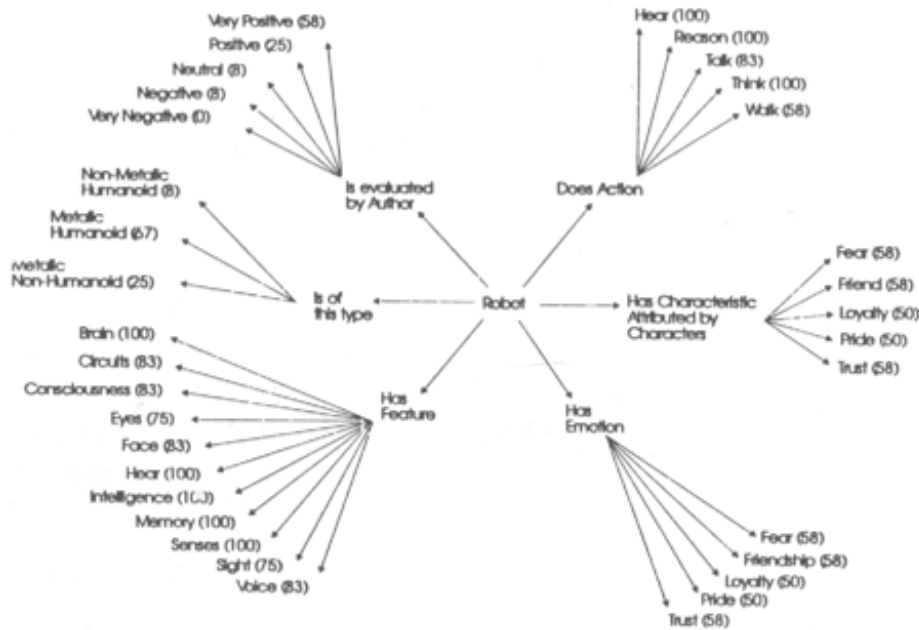
Content analysis coding for explicit concepts may not reveal any significant differences. For example, the existence of "I, scientist, research, hard work, collaboration, discoveries, new ideas, etc..." are explicit in both texts, occur the same number of times, and have the same emphasis. Relational analysis or cognitive mapping, however, reveals that while all concepts in the text are shared, only five concepts are common to both. Analyzing these statements reveals that Student A reports on what "I" found out about "scientists," and elaborated the notion of "scientists" doing "research." Student B focuses on what "I's" research was and sees scientists as "making discoveries" without emphasis on research.

### **Related Information: The Palmquist, Carley and Dale Study**

Consider these two questions: How has the depiction of robots changed over more than a century's worth of writing? And, do students and writing instructors share the same terms for describing the writing process? Although these questions seem totally unrelated, they do share a commonality: in the Palmquist, Carley & Dale study, their answers rely on computer-aided text analysis to demonstrate how different texts can be analyzed.

#### **Literary texts**

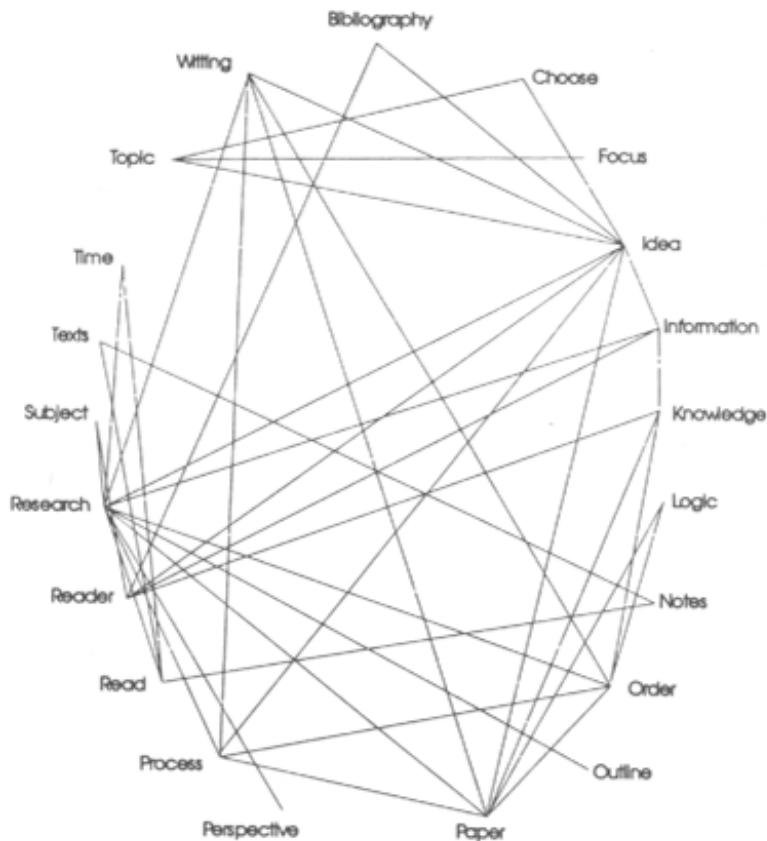
One half of the study explored the depiction of robots in 27 science fiction texts written between 1818 and 1988. After texts were divided into three historically defined groups, readers look for how the depiction of robots has changed over time. To do this, researchers had to create concept lists and relationship types, create maps using a computer software (see Fig. 1), modify those maps and then ultimately analyze them. The final product of the analysis revealed that over time authors were less likely to depict robots as metallic humanoids.



**Figure 1:** A map representing relationships among concepts.

**Non-literary texts**

The second half of the study used student journals and interviews, teacher interviews, texts books, and classroom observations as the non-literary texts from which concepts and words were taken. The purpose behind the study was to determine if, in fact, over time teacher and students would begin to share a similar vocabulary about the writing process. Again, researchers used computer software to assist in the process. This time, computers helped researchers generated a concept list based on frequently occurring words and phrases from all texts. Maps were also created and analyzed in this study (see Fig. 2).



**Figure 2:** Pairs of co-occurring words drawn from a source text

## Annotated Bibliography

### Resources On How To Conduct Content Analysis

Beard, J., & Yaprak, A. (1989). Language implications for advertising in international markets: A model for message content and message execution. A paper presented at the *8th International Conference on Language Communication for World Business and the Professions*. Ann Arbor, MI.

This report discusses the development and testing of a content analysis model for assessing advertising themes and messages aimed primarily at U.S. markets which seeks to overcome barriers in the cultural environment of international markets. Texts were categorized under 3 headings: rational, emotional, and moral. The goal here was to teach students to appreciate differences in language and culture.

Berelson, B. (1971). *Content analysis in communication research*. New York: Hafner Publishing Company.

While this book provides an extensive outline of the uses of content analysis, it is far more concerned with conveying a critical approach to current literature



on the subject. In this respect, it assumes a bit of prior knowledge, but is still accessible through the use of concrete examples.

Budd, R. W., Thorp, R.K., & Donohew, L. (1967). *Content analysis of communications*. New York: Macmillan Company.

Although published in 1967, the decision of the authors to focus on recent trends in content analysis keeps their insights relevant even to modern audiences. The book focuses on specific uses and methods of content analysis with an emphasis on its potential for researching human behavior. It is also geared toward the beginning researcher and breaks down the process of designing a content analysis study into 6 steps that are outlined in successive chapters. A useful annotated bibliography is included.

Carley, K. (1992). Coding choices for textual analysis: A comparison of content analysis and map analysis. Unpublished Working Paper.

Comparison of the coding choices necessary to conceptual analysis and relational analysis, especially focusing on cognitive maps. Discusses concept coding rules needed for sufficient reliability and validity in a Content Analysis study. In addition, several pitfalls common to texts are discussed.

Carley, K. (1990). Content analysis. In R.E. Asher (Ed.), *The Encyclopedia of Language and Linguistics*. Edinburgh: Pergamon Press.

Quick, yet detailed, overview of the different methodological kinds of Content Analysis. Carley breaks down her paper into five sections, including: Conceptual Analysis, Procedural Analysis, Relational Analysis, Emotional Analysis and Discussion. Also included is an excellent and comprehensive Content Analysis reference list.

Carley, K. (1989). *Computer analysis of qualitative data*. Pittsburgh, PA: Carnegie Mellon University.

Presents graphic, illustrated representations of computer based approaches to content analysis.

Carley, K. (1992). *MECA*. Pittsburgh, PA: Carnegie Mellon University.

A resource guide explaining the fifteen routines that compose the Map Extraction Comparison and Analysis (MECA) software program. Lists the source file, input and out files, and the purpose for each routine.

Carney, T. F. (1972). *Content analysis: A technique for systematic inference from communications*. Winnipeg, Canada: University of Manitoba Press.

This book introduces and explains in detail the concept and practice of content analysis. Carney defines it; traces its history; discusses how content analysis works and its strengths and weaknesses; and explains through examples and illustrations how one goes about doing a content analysis.

de Sola Pool, I. (1959). *Trends in content analysis*. Urbana, Ill: University of Illinois Press.

The 1959 collection of papers begins by differentiating quantitative and qualitative approaches to content analysis, and then details facets of its uses in a wide variety of disciplines: from linguistics and folklore to biography and history. Includes a discussion on the selection of relevant methods and representational models.

Duncan, D. F. (1989). Content analysis in health education research: An introduction to purposes and methods. *Health Education, 20* (7).

This article proposes using content analysis as a research technique in health education. A review of literature relating to applications of this technique and a procedure for content analysis are presented.

Gottschalk, L. A. (1995). *Content analysis of verbal behavior: New findings and clinical applications*. Hillside, NJ: Lawrence Erlbaum Associates, Inc.

This book primarily focuses on the Gottschalk-Gleser method of content analysis, and its application as a method of measuring psychological dimensions of children and adults via the content and form analysis of their verbal behavior, using the grammatical clause as the basic unit of communication for carrying semantic messages generated by speakers or writers.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology* Beverly Hills, CA: Sage Publications.

This is one of the most widely quoted resources in many of the current studies of Content Analysis. Recommended as another good, basic resource, as Krippendorff presents the major issues of Content Analysis in much the same way as Weber (1975).

Moeller, L. G. (1963). *An introduction to content analysis--including annotated bibliography*. Iowa City: University of Iowa Press.

A good reference for basic content analysis. Discusses the options of sampling, categories, direction, measurement, and the problems of reliability and validity in setting up a content analysis. Perhaps better as a historical text due to its age.

Smith, C. P. (Ed.). (1992). *Motivation and personality: Handbook of thematic content analysis*. New York: Cambridge University Press.

Billed by its authors as "the first book to be devoted primarily to content analysis systems for assessment of the characteristics of individuals, groups, or historical periods from their verbal materials." The text includes manuals for

using various systems, theory, and research regarding the background of systems, as well as practice materials, making the book both a reference and a handbook.

Solomon, M. (1993). Content analysis: a potent tool in the searcher's arsenal. *Database*, 16(2), 62-67.

Online databases can be used to analyze data, as well as to simply retrieve it. Online-media-source content analysis represents a potent but little-used tool for the business searcher. Content analysis benchmarks useful to advertisers include prominence, offspin, sponsor affiliation, verbatims, word play, positioning and notational visibility.

Weber, R. P. (1990). *Basic content analysis, second edition*. Newbury Park, CA: Sage Publications.

Good introduction to Content Analysis. The first chapter presents a quick overview of Content Analysis. The second chapter discusses content classification and interpretation, including sections on reliability, validity, and the creation of coding schemes and categories. Chapter three discusses techniques of Content Analysis, using a number of tables and graphs to illustrate the techniques. Chapter four examines issues in Content Analysis, such as measurement, indication, representation and interpretation.

### **Examples of Content Analysis**

Adams, W., & Shriebman, F. (1978). *Television network news: Issues in content research*. Washington, DC: George Washington University Press.

A fairly comprehensive application of content analysis to the field of television news reporting. The books tripartite division discusses current trends and problems with news criticism from a content analysis perspective, four different content analysis studies of news media, and makes recommendations for future research in the area. Worth a look by anyone interested in mass communication research.

Auter, P. J., & Moore, R. L. (1993). Buying from a friend: a content analysis of two teleshopping programs. *Journalism Quarterly*, 70(2), 425-437.

A preliminary study was conducted to content-analyze random samples of two teleshopping programs, using a measure of content interactivity and a locus of control message index.

Barker, S. P. (???) *Fame: A content analysis study of the American film biography*. Ohio State University. Thesis.

Barker examined thirty Oscar-nominated films dating from 1929 to 1979 using O.J. Harvey Belief System and the Kohlberg's Moral Stages to determine whether cinema heroes were positive role models for fame and success or

morally ambiguous celebrities. Content analysis was successful in determining several trends relative to the frequency and portrayal of women in film, the generally high ethical character of the protagonists, and the dogmatic, close-minded nature of film antagonists.

Bernstein, J. M. & Lacy, S. (1992). Contextual coverage of government by local television news. *Journalism Quarterly*, 69(2), 329-341.

This content analysis of 14 local television news operations in five markets looks at how local TV news shows contribute to the marketplace of ideas. Performance was measured as the allocation of stories to types of coverage that provide the context about events and issues confronting the public.

Blaikie, A. (1993). Images of age: a reflexive process. *Applied Ergonomics*, 24 (1), 51-58.

Content analysis of magazines provides a sharp instrument for reflecting the change in stereotypes of aging over past decades.

Craig, R. S. (1992). The effect of day part on gender portrayals in television commercials: a content analysis. *Sex Roles: A Journal of Research*, 26 (5-6), 197-213.

Gender portrayals in 2,209 network television commercials were content analyzed. To compare differences between three day parts, the sample was chosen from three time periods: daytime, evening prime time, and weekend afternoon sportscasts. The results indicate large and consistent differences in the way men and women are portrayed in these three day parts, with almost all comparisons reaching significance at the .05 level. Although ads in all day parts tended to portray men in stereotypical roles of authority and dominance, those on weekends tended to emphasize escape from home and family. The findings of earlier studies which did not consider day part differences may now have to be reevaluated.

Dillon, D. R. et al. (1992). Article content and authorship trends in *The Reading Teacher*, 1948-1991. *The Reading Teacher*, 45 (5), 362-368.

The authors explore changes in the focus of the journal over time.

Eberhardt, EA. (1991). *The rhetorical analysis of three journal articles: The study of form, content, and ideology*. Ft. Collins, CO: Colorado State University.

Eberhardt uses content analysis in this thesis paper to analyze three journal articles that reported on President Ronald Reagan's address in which he responded to the Tower Commission report concerning the IranContra Affair. The reports concentrated on three rhetorical elements: idea generation or content; linguistic style or choice of language; and the potential societal effect of both, which Eberhardt analyzes, along with the particular ideological orientation espoused by each magazine.

Ellis, B. G. & Dick, S. J. (1996). 'Who was 'Shadow'? The computer knows: applying

grammar-program statistics in content analyses to solve mysteries about authorship. *Journalism & Mass Communication Quarterly*, 73(4), 947-963.

This study's objective was to employ the statistics-documentation portion of a word-processing program's grammar-check feature as a final, definitive, and objective tool for content analyses - used in tandem with qualitative analyses - to determine authorship. Investigators concluded there was significant evidence from both modalities to support their theory that Henry Watterson, long-time editor of the Louisville Courier-Journal, probably was the South's famed Civil War correspondent "Shadow" and to rule out another prime suspect, John H. Linebaugh of the Memphis Daily Appeal. Until now, this Civil War mystery has never been conclusively solved, puzzling historians specializing in Confederate journalism.

Gottschalk, L. A., Stein, M. K. & Shapiro, D.H. (1997). The application of computerized content analysis in a psychiatric outpatient clinic. *Journal of Clinical Psychology*, 53(5), 427-442.

Twenty-five new psychiatric outpatients were clinically evaluated and were administered a brief psychological screening battery which included measurements of symptoms, personality, and cognitive function. Included in this assessment procedure were the Gottschalk-Gleser Content Analysis Scales on which scores were derived from five minute speech samples by means of an artificial intelligence-based computer program. The use of this computerized content analysis procedure for initial, rapid diagnostic neuropsychiatric appraisal is supported by this research.

Graham, J. L., Kamins, M. A., & Oetomo, D. S. (1993). Content analysis of German and Japanese advertising in print media from Indonesia, Spain, and the United States. *Journal of Advertising*, 22 (2), 5-16.

The authors analyze informational and emotional content in print advertisements in order to consider how home-country culture influences firms' marketing strategies and tactics in foreign markets. Research results provided evidence contrary to the original hypothesis that home-country culture would influence ads in each of the target countries.

Herzog, A. (1973). *The B.S. Factor: The theory and technique of faking it in America*. New York: Simon and Schuster.

Herzog takes a look at the rhetoric of American culture using content analysis to point out discrepancies between intention and reality in American society. The study reveals, albeit in a comedic tone, how double talk and "not quite lies" are pervasive in our culture.

Horton, N. S. (1986). *Young adult literature and censorship: A content analysis of seventy-eight young adult books*. Denton, TX: North Texas State University.

The purpose of Horton's content analysis was to analyze a representative seventy-eight current young adult books to determine the extent to which they contain items which are objectionable to would-be censors. Seventy-eight books were identified which fit the criteria of popularity and literary quality. Each book was analyzed for, and tallied for occurrence of, six categories, including profanity, sex, violence, parent conflict, drugs and condoned bad behavior.

Isaacs, J. S. (1984). *A verbal content analysis of the early memories of psychiatric patients*. Berkeley: California School of Professional Psychology.

Isaacs did a content analysis investigation on the relationship between words and phrases used in early memories and clinical diagnosis. His hypothesis was that in conveying their early memories schizophrenic patients tend to use an identifiable set of words and phrases more frequently than do nonpatients and that schizophrenic patients use these words and phrases more frequently than do patients with major affective disorders.

Jean Lee, S. K. & Hwee Hoon, T. (1993). Rhetorical vision of men and women managers in Singapore. *Human Relations*, 46 (4), 527-542.

A comparison of media portrayal of male and female managers' rhetorical vision in Singapore is made. Content analysis of newspaper articles used to make this comparison also reveals the inherent conflicts that women managers have to face. Purposive and multi-stage sampling of articles are utilized.

Kaur-Kasior, S. (1987). *The treatment of culture in greeting cards: A content analysis*. Bowling Green, OH: Bowling Green State University.

Using six historical periods dating from 1870 to 1987, this content analysis study attempted to determine what structural/cultural aspects of American society were reflected in greeting cards. The study determined that the size of cards increased over time, included more pages, and had animals and flowers as their most dominant symbols. In addition, white was the most common color used. Due to habituation and specialization, says the author, greeting cards have become institutionalized in American culture.

Koza, J. E. (1992). The missing males and other gender-related issues in music education: A critical analysis of evidence from the Music Supervisor's Journal, 1914-1924. Paper presented at the annual meeting of the American Educational Research Association. San Francisco.

The goal of this study was to identify all educational issues that would today be explicitly gender related and to analyze the explanations past music educators gave for the existence of gender-related problems. A content analysis of every gender-related reference was undertaken, finding that the current preoccupation with males in music education has a long history and

that little has changed since the early part of this century.

Laccinole, M. D. (1982). *Aging and married couples: A language content analysis of a conversational and expository speech task*. Eugene, OR: University of Oregon.

Using content analysis, this paper investigated the relationship of age to the use of the grammatical categories, and described the differences in the usage of these grammatical categories in a conversation and expository speech task by fifty married couples. The subjects Laccinole used in his analysis were Caucasian, English speaking, middle class, ranged in ages from 20 to 83 years of age, were in good health and had no history of communication disorders.

Laffal, J. (1995). A concept analysis of Jonathan Swift's 'A Tale of a Tub' and 'Gulliver's Travels.' *Computers and Humanities*, 29(5), 339-362.

In this study, comparisons of concept profiles of "Tub," "Gulliver," and Swift's own contemporary texts, as well as a composite text of 18th century writers, reveal that "Gulliver" is conceptually different from "Tub." The study also discovers that the concepts and words of these texts suggest two strands in Swift's thinking.

Lewis, S. M. (1991). Regulation from a deregulatory FCC: Avoiding discursive dissonance. Masters Thesis, Fort Collins, CO: Colorado State University.

This thesis uses content analysis to examine inconsistent statements made by the Federal Communications Commission (FCC) in its policy documents during the 1980s. Lewis analyzes positions set forth by the FCC in its policy statements and catalogues different strategies that can be used by speakers to be or to appear consistent, as well as strategies to avoid inconsistent speech or discursive dissonance.

Norton, T. L. (1987). *The changing image of childhood: A content analysis of Caldecott Award books*. Los Angeles: University of South Carolina.

Content analysis was conducted on 48 Caldecott Medal Recipient books dating from 1938 to 1985 to determine whether they reflect the idea that the social perception of childhood has altered since the early 1960's. The results revealed an increasing "loss of childhood innocence," as well as a general sentimentality for childhood pervasive in the texts. Suggests further study of children's literature to confirm the validity of such study.

O'Dell, J. W. & Weideman, D. (1993). Computer content analysis of the Schreber case. *Journal of Clinical Psychology*, 49(1), 120-125.

An example of the application of content analysis as a means of recreating a mental model of the psychology of an individual.

Pratt, C. A. & Pratt, C. B. (1995). Comparative content analysis of food and nutrition advertisements in *Ebony*, *Essence*, and *Ladies' Home Journal*. *Journal of Nutrition*

*Education*, 27(1), 11-18.

This study used content analysis to measure the frequencies and forms of food, beverage, and nutrition advertisements and their associated health-promotional message in three U.S. consumer magazines during two 3-year periods: 1980-1982 and 1990-1992. The study showed statistically significant differences among the three magazines in both frequencies and types of major promotional messages in the advertisements. Differences between the advertisements in *Ebony* and *Essence*, the readerships of which were primarily African-American, and those found in *Ladies Home Journal* were noted, as were changes in the two time periods. Interesting tie in to ethnographic research studies?

Riffe, D., Lacy, S., & Drager, M. W. (1996). Sample size in content analysis of weekly news magazines. *Journalism & Mass Communication Quarterly*, 73(3), 635-645.

This study explores a variety of approaches to deciding sample size in analyzing magazine content. Having tested random samples of size six, eight, ten, twelve, fourteen, and sixteen issues, the authors show that a monthly stratified sample of twelve issues is the most efficient method for inferring to a year's issues.

Roberts, S. K. (1987). *A content analysis of how male and female protagonists in Newbery Medal and Honor books overcome conflict: Incorporating a locus of control framework*. Fayetteville, AR: University of Arkansas.

The purpose of this content analysis was to analyze Newbery Medal and Honor books in order to determine how male and female protagonists were assigned behavioral traits in overcoming conflict as it relates to an internal or external locus of control schema. Roberts used all, instead of just a sample, of the fictional Newbery Medal and Honor books which met his study's criteria. A total of 120 male and female protagonists were categorized, from Newbery books dating from 1922 to 1986.

Schneider, J. (1993). *Square One TV content analysis: Final report*. New York: Children's Television Workshop.

This report summarizes the mathematical and pedagogical content of the 230 programs in the Square One TV library after five seasons of production, relating that content to the goals of the series which were to make mathematics more accessible, meaningful, and interesting to the children viewers.

Smith, T. E., Sells, S. P., and Clevenger, T. Ethnographic content analysis of couple and therapist perceptions in a reflecting team setting. *The Journal of Marital and Family Therapy*, 20 (3), 267-286.



An ethnographic content analysis was used to examine couple and therapist perspectives about the use and value of reflecting team practice. Postsession ethnographic interviews from both couples and therapists were examined for the frequency of themes in seven categories that emerged from a previous ethnographic study of reflecting teams. Ethnographic content analysis is briefly contrasted with conventional modes of quantitative content analysis to illustrate its usefulness and rationale for discovering emergent patterns, themes, emphases, and process using both inductive and deductive methods of inquiry.

Stahl, N. A. (1987). Developing college vocabulary: A content analysis of instructional materials. *Reading, Research and Instruction, 26* (3).

This study investigates the extent to which the content of 55 college vocabulary texts is consistent with current research and theory on vocabulary instruction. It recommends less reliance on memorization and more emphasis on deep understanding and independent vocabulary development.

Swetz, F. (1992). Fifteenth and sixteenth century arithmetic texts: What can we learn from them? *Science and Education, 1* (4).

Surveys the format and content of 15th and 16th century arithmetic textbooks, discussing the types of problems that were most popular in these early texts and briefly analyses problem contents. Notes the residual educational influence of this era's arithmetical and instructional practices.

Walsh, K., et al. (1996). Management in the public sector: a content analysis of journals. *Public Administration 74* (2), 315-325.

The popularity and implementation of managerial ideas from 1980 to 1992 are examined through the content of five journals revolving on local government, health, education and social service. Contents were analyzed according to commercialism, user involvement, performance evaluation, staffing, strategy and involvement with other organizations. Overall, local government showed utmost involvement with commercialism while health and social care articles were most concerned with user involvement.

### **For Further Reading**

Abernethy, A. M., & Franke, G. R. (1996). The information content of advertising: a meta-analysis. *Journal of Advertising, Summer 25* (2), 1-18.

Carley, K., & Palmquist, M. (1992). Extracting, representing and analyzing mental models. *Social Forces, 70* (3), 601-636.

Fan, D. (1988). *Predictions of public opinion from the mass media: Computer content analysis and mathematical modeling*. New York, NY: Greenwood Press.

Franzosi, R. (1990). Computer-assisted coding of textual data: An application to semantic

grammars. *Sociological Methods and Research*, 19(2), 225-257.

McTavish, D.G., & Pirro, E. (1990) Contextual content analysis. *Quality and Quantity*, 24, 245-265.

Palmquist, M. E. (1990). *The lexicon of the classroom: language and learning in writing class rooms*. Doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Palmquist, M. E., Carley, K.M., and Dale, T.A. (1997). Two applications of automated text analysis: Analyzing literary and non-literary texts. In C. Roberts (Ed.), *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Roberts, C.W. (1989). Other than counting words: A linguistic approach to content analysis. *Social Forces*, 68, 147-177.

### **Issues in Content Analysis**

Jolliffe, L. (1993). Yes! More content analysis! *Newspaper Research Journal*, 14(3-4), 93-97.

The author responds to an editorial essay by Barbara Luebke which criticizes excessive use of content analysis in newspaper content studies. The author points out the positive applications of content analysis when it is theory-based and utilized as a means of suggesting how or why the content exists, or what its effects on public attitudes or behaviors may be.

Kang, N., Kara, A., Laskey, H. A., & Seaton, F. B. (1993). A SAS MACRO for calculating intercoder agreement in content analysis. *Journal of Advertising*, 22(2), 17-28.

A key issue in content analysis is the level of agreement across the judgments which classify the objects or stimuli of interest. A review of articles published in the *Journal of Advertising* indicates that many authors are not fully utilizing recommended measures of intercoder agreement and thus may not be adequately establishing the reliability of their research. This paper presents a SAS MACRO which facilitates the computation of frequently recommended indices of intercoder agreement in content analysis.

Lacy, S. & Riffe, D. (1996). Sampling error and selecting intercoder reliability samples for nominal content categories. *Journalism & Mass Communication Quarterly*, 73(4), 693-704.

This study views intercoder reliability as a sampling problem. It develops a formula for generating sample sizes needed to have valid reliability estimates. It also suggests steps for reporting reliability. The resulting sample sizes will permit a known degree of confidence that the agreement in a sample of items is representative of the pattern that would occur if all content items were coded by all coders.

Riffe, D., Aust, C. F., & Lacy, S. R. (1993). The effectiveness of random, consecutive day

and constructed week sampling in newspaper content analysis. *Journalism Quarterly*, 70 (1), 133-139.

This study compares 20 sets each of samples for four different sizes using simple random, constructed week and consecutive day samples of newspaper content. Comparisons of sample efficiency, based on the percentage of sample means in each set of 20 falling within one or two standard errors of the population mean, show the superiority of constructed week sampling.

Thomas, S. (1994). Artifactual study in the analysis of culture: A defense of content analysis in a postmodern age. *Communication Research*, 21(6), 683-697.

Although both modern and postmodern scholars have criticized the method of content analysis with allegations of reductionism and other epistemological limitations, it is argued here that these criticisms are ill founded. In building and argument for the validity of content analysis, the general value of artifact or text study is first considered.

Zollars, C. (1994). The perils of periodical indexes: Some problems in constructing samples for content analysis and culture indicators research. *Communication Research*, 21(6), 698-714.

The author examines problems in using periodical indexes to construct research samples via the use of content analysis and culture indicator research. Issues of historical and idiosyncratic changes in index subject category heading and subheadings make article headings potentially misleading indicators. Index subject categories are not necessarily invalid as a result; nevertheless, the author discusses the need to test for category longevity, coherence, and consistency over time, and suggests the use of oversampling, cross-references, and other techniques as a means of correcting and/or compensating for hidden inaccuracies in classification, and as a means of constructing purposive samples for analytic comparisons.

## Citation Information

Carol Busch, Paul S. De Maret, Teresa Flynn, Rachel Kellum , Sheri Le, Brad Meyers, Matt Saunders, Robert White, and Mike Palmquist.. (1994 - 2012). Content Analysis. Writing@CSU. Colorado State University. Available at <http://writing.colostate.edu/guides/guide.cfm?guideid=61>.

## Copyright Information

[Copyright © 1994-2016 Colorado State University](#) and/or [this site's authors, developers, and contributors](#). Some material displayed on this site is used with permission.