United States General Accounting Office

GAO

March 1989

Content Analysis: A Methodology for Structuring and Analyzing Written Material

Transfer Paper 10.1.3

Preface

In this paper, we define and describe the evaluation method called "content analysis." It is a set of procedures for transforming nonstructured information into a format that allows analysis. From reading this paper, GAO analysts should gain an understanding of the basic concepts and procedures used in content analysis and also an ability to recognize the appropriate circumstances for using this evaluation method in their jobs.

Although we have focused on techniques that make quantitative analysis possible, this is not necessarily the objective of all content analyses. We have presented the techniques that are the most applicable to GAO's work. In chapter 1, we define content analysis and compare it to similar procedures already used in GAO. In chapter 2, we discuss the procedures for using content analysis. In chapter 3, we explain the advantages and disadvantages of content analysis and describe some of its potential applications in program evaluation.

The paper is designed to be self-instructional. References are provided throughout the text for readers who want more information on specific topics, and these references are keyed to the bibliography.

Research for this document began with a survey of the numerous books and articles on content analysis and its past applications. We also interviewed users of content analysis to gain information about its advantages and disadvantages, and we interviewed selected GAO staff who have participated in evaluations in which content analysis might have been appropriate. The foundation for this document is a paper written by William Carter while a student intern with GAO. The document was prepared by Teresa Spisak, formerly of the Institute for Program Evaluation (now PEMD), and was originally published in 1982 as Transfer Paper 3. It is being reissued now with only minor changes, including some updating of bibliographic materials.

Content Analysis is one of a series of papers issued by PEMD. The purpose of the series is to provide GAO evaluators with a clear and comprehensive background of the basic concepts of audit and evaluation methodology. Additionally, transfer papers explain both general and

specific applications and procedures for using the evaluation methodology. Other papers in this series include Causal Analysis, Designing Evaluations, Using Structured Interviewing Techniques, Using Questionnaires, Using Statistical Sampling, and Case Study Evaluations.

Eleanor Chelimsky
Assistant Comptroller General for Program Evaluation and Methodology

Contents

Abbreviations

GAO	U.S. General Accounting Office
HUD	U.S. Department of Housing and Urban Development
PEMD	Program Evaluation and Methodology Division

Chapter 1

What Is Content Analysis?

GAO staff often collect large quantities of written material during their jobs. Workpapers, agency documents, transcripts of meetings, previous evaluations, and the like all contain useful information that is difficult to combine and analyze because it is diverse and unstructured. Content analysis is a set of procedures for collecting and organizing this information.

One way to begin structuring written material so that it can be analyzed is to summarize and list the major issues that are contained in it. Then the frequency with which these issues occur can be counted. Both activities are usually performed at some point in GAO jobs, and both are part of content analysis.

For example, in assessing HUD's evaluation system to determine whether program offices were duplicating efforts, GAO analysts collected budget information, interviews, and evaluation reports. (GAO, 1978)[1] They began analyzing the information by identifying 31 major issues for housing and urban development. Then they reviewed 38 HUD evaluation reports from two offices, categorizing the issues addressed in each report and looking for overlaps between the offices. Simplifying and categorizing written information are part of content analysis.

In addition to requiring summaries of written material and enumerations of the frequency of statements or issues, GAO projects often require more complex analyses. Sometimes trends have to be examined over time, across different situations, or among different groups. The information that is needed to make these types of analysis may not exist in computer files. With content analysis, information from written material can be structured so that these types of analysis can be made even without computer files.

Content analysis is a set of procedures for collecting and organizing information in a standardized format that allows analysts to make inferences about the characteristics and meaning of written and other recorded material. Simple formats can be developed for summarizing information or counting the frequency of statements. More complex formats can be created for analyzing trends or detecting subtle differences in the intensity of statements.

Among the procedures of content analysis that we discuss in the next chapter are defining and sampling the written or recorded material to be

[1] Interlinear bibliographic references are cited in full in the bibliography.

analyzed, developing standardized categories, coding the material with rigorous reliability checks, analyzing and interpreting the information, and validating and reporting the results. Although in this paper we have focused on procedures that make quantitative analysis possible, this is not necessarily the objective of all forms of content analysis.

Chapter 2

What Are the Procedures in Content Analysis?

The steps to be followed in content analysis are summarized in figure 2.1. Steps 1, 2, and 6—deciding whether or not the methodology is appropriate, determining what material should be analyzed, and analyzing and interpreting the results—are integral aspects of all projects. However, steps 3, 4, and 5—choosing the units of analysis, developing coding categories, and coding the material—are unique to content analysis, and therefore we will explain these in greater detail.

Figure 2.1: Steps in Content Analysis

1. Decide to use content analysis.
2. Determine what material should be included in content analysis.
3. Select units of analysis.
4. Develop coding categories.
5. Code the material.
6. Analyze and interpret the results.

Deciding to Use Content Analysis

At step 1, analysts should consider a number of factors in deciding whether or not to use content analysis. These include a project's objectives, data availability, and the kinds of analyses required.

Objectives

Objectives are precisely worded questions that the project staff are trying to answer. (GAO, December 1988, p. 10-4) The questions should be based on a clear understanding of project needs and the available data. Precisely worded questions provide the focus for data collection, analysis, and reporting. In general, content analysis can be used to answer "What?" but not "Why?" That is, it helps analysts describe or summarize the content of written material, the attitudes or perceptions of its writer, or its effects on its audience.

The content of material can be summarized by listing or by counting the issues or statements within it, as we indicated in chapter 1. The author's attitudes and perceptions can also be described. For example, if analysts wanted to assess the effects of various programs on the lives of older people, content analysis of open-ended interview responses could be used to identify their outlook on life and their attitudes about loneliness or security. Content analysis can also be useful in describing the effects of messages on their recipients. For example, the effect of Voice of America broadcasts has been assessed by analyzing Soviet newspapers and transcripts of radio broadcasts. (Inkeles, 1952)

The Kinds of Material Available

Content analysis can be used to study any recorded material as long as the information is available to be reanalyzed for reliability checks. Although it is used most frequently to analyze written material, content analysis can be used to study any recorded communication, including television programs, movies, and photographs. It can be used to analyze congressional testimony, legislation, regulations, other public documents, workpapers, case studies, reports, answers to survey questions, news releases, newspapers, books, journal articles, and letters. A speech or a discussion, however, cannot be analyzed unless it has been transcribed or taped.

Before using content analysis, project staff should assess the written material's quality. Does the available material accurately represent what was written or said? A garbled tape recording or written material with sections missing is not a sound basis for content analysis. Findings and conclusions from content analysis can never be more accurate than the material that has been analyzed.

The Kinds of Comparison Required

Content analysis can be used for making numerical comparisons among and within documents. For example, staff who want to describe or summarize the content of written material can use content analysis to compare documents derived from a single source, such as from one federal agency, by comparing issues or statements over time, in different situations, or across differing groups. The relationship of two or more statements or issues within a single document or set of documents can also be analyzed. Alternatively, statements or issues from two or more different sources can be compared.

Determining What Material Should Be Included

Sampling is necessary if the body of material, the "universe," is too extensive to be analyzed in its entirety. Thus, at step 2, analysts who want to make valid conclusions and generalizations about a universe should select from that universe a sample that is representative of it.[1]

Selecting samples for content analysis usually involves sampling documents. For example, in a hypothetical project evaluating changes in the eligibility requirements in a food stamp program, more than 500 participants might be interviewed. By arranging the interview transcripts alphabetically and then selecting every tenth transcript for content analysis, the project staff might be able to draw a systematic sample. Other types of sampling design may also be used. (Babbie, 1973, pp. 91-102)

Selecting Units of Analysis

In content analysis, the researcher designates the units of analysis, called "recording units," and the units of context. This is step 3. Context units set limits on the portion of written material that is to be examined for categories of words or statements. Context units can be the same as the units sampled, although they are not always the same.

Since it is not always practical to use long documents as context units, chapters, sections, paragraphs, or even sentences may be better choices. This is especially true when attempts are made to identify subtle differences in content. For example, a meeting transcript can be analyzed to determine the extent to which the meeting's participants supported or opposed various issues. In this case, the analysts would choose sentences as the context unit if entire statements were relatively long and tended, as sometimes happens, to contain conflicting information. It may be typical for a given speaker to oppose an issue at the beginning of a statement but to shift to support of it at the end. To identify such shifts in position, analysts need to examine a small content unit such as the sentence.

A recording unit is the specific segment of the context unit in the written material that is placed in a category. It may be a word, a group of words (such as those that identify a theme), a sentence, a paragraph, or an entire document. It can never be larger than the context unit. In the HUD study we cited earlier, analysts used the groups of words that

[1]Readers unfamiliar with basis sampling theory and methods should refer to GAO, December 1988, pp. 11-15 to 11-19 and 11-25 to 11-36.

embodied the discussion of the issues as recording units. Their context units were the evaluation studies.

Developing Coding Categories

Categories provide the structure for grouping recording units. Step 4, formulating categories, is the heart of content analysis. Berelson, an early user of content analysis, emphasized the importance of this step when he cautioned that

> "Content analysis stands or falls by its categories. Particular studies have been productive to the extent that the categories were clearly formulated and well adapted to the problem and to the content." (Berelson, 1952, p. 147)

Figure 2.2 lists standard requirements that categories should meet. Adhering to these requirements helps keep an analysis systematic and objective, which leads to results that are amenable to statistical calculation.

Figure 2.2: Requirements for Content Categories

1. Categories should be exhaustive—so that all relevant items in the material being studied can be placed within a category.

2. Categories should be mutually exclusive—so that no item can be coded in more than one category.

3. Categories should be independent—so that a recording unit's category assignment is not affected by the category assignment of other recording units.

Category Formats

Categories can be conceptualized in numerous ways. Some common category formats are groupings, scales, and matrices.[2] Structured category formats increase coding efficiency, especially when the number of categories is large.

In our HUD example, analysts chose groups of issues as categories. They grouped 31 issues into three general categories. For example, issues such as dispersion of housing, block grants, and public housing modernization were placed in the category "Housing Assistance Issues."

Scales provide for the rank ordering of information. In the HUD example, had the analysts wanted to know the extent to which the reports they were examining supported the issues, they could have used a scale such as "supports, is uncommitted, opposes."

Matrices are useful formats when analysts seek more information about issues than simply whether they are present or absent. The group and scale categories we discussed above could be combined into a matrix format such as that shown in figure 2.3.

[2]Krippendorff discusses these and more sophisticated formats such as trees, loops, chains, cubes, and partition lattices. (Krippendorff, 1980, pp. 91-98)

Figure 2.3: Matrix Category Format

	Degree of support for issue		
Issue	**Supports**	**Opposes**	**Uncommitted**
I Housing assistance			
A. Block grants			
B. Housing dispersion			
C. Public housing modernization			

Quantification Levels

Categories can be used to measure three quantification levels—space, frequency, and intensity. To explain the differences between these quantification levels and how they relate to constructing categories, we use a hypothetical analysis of handgun control legislation for which the analyst has as major sources of information newspaper articles, public documents, and transcripts of interviews with public officials.

At the least rigorous level of quantification, the hypothetical analyst can measure the amount of space in the newspaper articles devoted to positions supporting or opposing the issue. The analyst then can use this measurement to compare the relative strength of issues supporting and opposing handgun control.

In selecting newspapers, the analyst also has to control for factors that may influence the articles' content or editorial viewpoint. The category format shown in figure 2.4 uses the newspapers' location (rural versus urban) for this purposed. For each issue of each newspaper in the sample, the analyst adds together the number of column inches from all news articles and editorials to find the total amount of space for each position. By also coding the name, location, and date of each newspaper, the analyst can examine trends across time and can compare rural and urban viewpoints.

Figure 2.4: Category Format Measuring Space

			Number of column inches		
Newspaper	**Date**	**Location**	**Supporting**	**Opposing**	**Uncommitted**
"Times"	11/12/81	Urban	4	0	2
"Examiner"	11/18/81	Rural	0	5	2

Such measurement is rapid and relatively easy, but it provides only very general information. Furthermore, analysts who use this level of quantification have to assume that the differences they find in amounts of space are valid indicators of relative emphasis or importance.

At the next level of quantification, the analyst can code the frequency of recording units by tallying the number of times each issue or statement occurs in the text. Formats for measuring frequency can be very simple, as in figure 2.5, or more complex, as in figure 2.6, depending on the information needs of the project.

Figure 2.5: Two Category Formats Measuring Frequency of Statements

Format 1

			Number of column inches		
Newspaper	**Date**	**Location**	**Supporting**	**Opposing**	**Uncommitted**
"Times"	11/21/81	Urban	2	0	1
"Examiner"	11/18/81	Rural	0	4	0

Format 2

Newspaper	**Date**	**Location**	**Statement attribution**	**Position**
"Times"	11/12/81	Urban	State politician	Supports
"Times"	11/12/81	Urban	Editorial	Supports
"Times"	11/12/81	Urban	U.S. Senator	Uncommitted
"Examiner"	11/18/81	Rural	Citizens' group	Opposes
"Examiner"	11/18/81	Rural	State politician	Opposes

Figure 2.6: Measuring Frequency of and Position Taken on Specific Proposals

Category Format

Proposals for handgun control	Supports (01)	Opposes (02)	Uncommitted/no position (03)
Banning handgun sales	(01)		
Banning importation of unassembled gun parts	(02)		
Handgun registration	(03)		
Stricter controls on handgun purchases	(04)		
Stronger penalties for using handguns to commit crimes	(05)		
More stringent enforcement of existing control	(06)		
Other	(07)		

Coding Format

		Statement	
Source	Date	Column	Row
Presidential advisory panel	8/6/81	01	02
Presidential advisory panel	8/6/81	01	07
Presidential advisory panel	8/6/81	01	04

Figure 2.5 presents two simple formats for measuring the number of statements supporting, opposing, and uncommitted to handgun control. Format 1 is similar to the format for measuring space but instead measures the number of articles that appear over a given period of time. Format 2 identifies the speaker and allows the analyst to compare positions by different individuals over time and by different locations.

Figure 2.6 shows a more elaborate means of measuring frequency, with separate formats for category and for coding. This approach could be used to analyze information from all three data sources in the hypothetical example—newspapers, public documents, and interview transcripts. In the figure, the categories describe positions on specific proposals for handgun control. The positions can be coded by assigning them four digits that indicate the positions taken (columns) on the proposals (rows).

To show how this works, we can examine the recommendations in the following statement from a New York Times article published on August 6, 1981, coded as shown in figure 2.6.

"The eight-member (Presidential advisory) panel . . . recommended legislation forbidding the importing of pistol parts, requiring citizens to report the theft or loss of a pistol, and establishing a waiting period before a pistol is purchased to permit the authorities to determine if the purchaser has a criminal record."

The recommendation for legislation forbidding the importing of pistol parts is coded as column 01 ("supports"), row 02 ("banning importation of unassembled gun parts"). The second recommendation, "requiring citizens to report the theft or loss of a pistol," is coded as "other" (01 07) since it is not in the list of specific proposals.

In general, analysts incorporate two assumptions in their research designs when they construct frequency measures. First, they assume that the frequency with which a statement occurs in the text is a valid indication of value or importance. Second, they assume that all content units can be given equal weight and therefore that each one can be compared directly with every other.

At the third level of quantification, analysts code for intensity. Frequencies are counted, but each coded statement or issue is also adjusted by a weight that measures relative intensity.[3] This measurement level allows much more sensitive data analysis.

One drawback of intensity coding, however, is that it requires coders to recognize more subtle differences in the material than they need to when coding for space or frequency. Furthermore, it is difficult to list all criteria that coders have to consider in making their decisions. For example, coders may have to consider the relative intensity of the meaning of verbs ("disagree" versus "doubt") or their tenses (past, present, future), of the meaning of adverbial modifiers ("often" versus "sometimes"), or of the meaning of statements that express what is probable (using "may") versus what is imperative (using "must").

Since it helps analysts compare subtle differences in words, this level of quantification is the most useful for analyzing direct quotations and the contents of official documents, such as public laws and regulations, in which words are understood to have been chosen carefully to convey a

[3]Three methods of calculating and assigning weights are discussed in North et al., 1963, pp. 55-103.

precise message. In the gun control example, therefore, only the interview transcripts would be analyzed at this level.

Figure 2.7 illustrates how attitude intensity can be coded. Using two hypothetical interview responses, it shows how replies can be fitted into the category form "subject, verb, common meaning term." Each reply may contain more than one statement—or recording unit—to be coded. Therefore, values ranging from +3 to -3, depending on direction and intensity, are assigned to the verb and the common meaning term in each statement. In this case, a plus is assigned to verbs and common meaning terms that appear to support gun control. Each statement's two values—the value of its verb and the value of its common meaning term—are multiplied, and then the products for all the statements in the response are summed, yielding a total score for each response.

Figure 2.7: Category Format Measuring Attitude Intensity

Response 1
"Personally, I'm for gun control, but I doubt that a general gun control bill would meet with very much success."

Subject	Verb	Value	Common meaning term	Value	Product
I	am	+3	for gun control	+3	+9
I	doubt	−2	bill would meet with very much success	+3	−6
Total					**+3**

Response 2
"I urge the government to tighten its controls on handguns sold to residents."

Subject	Verb	Value	Common meaning term	Value	Product
I	urge	+3	government to tighten its controls	+3	+9
Total					**+9**

In the example in Figure 2.7, response 1 contains two statements while response 2 contains only one. The qualifying statement in the first response lowers its intensity so that, overall, the second response is given a higher intensity rating.

Coding the Material

Material can be coded either manually or by computers, depending on the resources available and the format of the material. This is step 5 in content analysis. If the material is already computerized, the analyst should explore the possibility of obtaining a computer program to do the coding. After deciding how the material will be coded, the analyst writes the necessary instructions. Figure 2.8 spells out the minimum requirements for instructions for trained coders.

Figure 2.8: Guidelines for Contents of Coding Instructions for Trained Coders

1. Definition of recording units, including procedures for identifying them.

2. Descriptions of the variables and categories.

3. Outline of the cognitive procedures used in placing data in categories.

4. Instructions for using and administering data sheets.

Source: Adapted from K Kripendorff, Content Analysis: An Introduction to Its Methodology (Beverly Hills, Calif Sage Publications, 1980), p 174.

Pretesting

Pretesting is an important step before actual coding begins. It involves coding a small portion of the material to be analyzed or some other similar material. From the pretests, the analyst tests and revises the coding categories and instructions, and does this several times in some cases. Pretesting is necessary whether computers are used for content analysis or the analysis is done by hand. Computer analysis requires test computer runs to ensure that the program is functioning as planned.

A pretest enables the analyst to determine whether (1) the categories are clearly specified and meet the requirements in figure 2.2, (2) the coding instructions are adequate, and (3) the coders are suitable for the job. These determinations are made by assessing reliability among coders and consistency in individual coding decisions (as we discuss below). Once the analyst has been assured that the material can be coded with high reliability, the pretests are over, and the coding can begin.

Data can, of course, be coded with the help of computer programs. (Weber, 1985) This solves the reliability problem but generates others. For one, all the material to be coded must be entered on a computer tape or disk, even though this may be impractical. For another, computer programs that perform content analysis require very specific categories.

For example, using a computer usually confines analysts to words as recording units, but this means that every word being coded has to be listed in the computer's memory as in a dictionary. Preparing a dictionary, however, may be far more difficult than formulating categories. Furthermore, because a word takes on different meanings in different contexts—a subtlety which computers cannot discern but people can—the results of computer coding may lack validity.

Computers should not be completely discounted, however, because they do have advantages. They are valuable in a number of situations. Computers can save time and permit analysis of large amounts of data when the word is the optimal unit of analysis. Because computers can "remember" many more definitions than people can, they are useful when categories are numerous. They are also valuable when data will be reused. Thus, the cost of preparing a data base for a series of studies for computer analysis may be offset by the benefit of having easily manageable data in the future. (Holsti, 1969, pp. 151-54)

Checking for Reliability

A check for reliability tells analysts the extent to which a measuring procedure can produce the same results on repeated trials. (Carmines and Zeller, 1979, p. 11) In content analysis, this means determining the similarity with which two or more people categorize the same material. Analysts have to assess reliability while pretesting the coding categories and instructions and also throughout the coding process.

To check for reliability, an analyst compares the way independent coders have coded the same material.[4] For example, two coders might be given ten items to code individually. The analyst compares their coding decisions and determines the extent to which they agree.

[4]Many reliability formulas have been developed for computing the percentage agreement among coders. See Kaplan and Golden, 1949; Krippendorff, 1980; Robinson, 1957; and Spiegelman et al., 1957. Scott's formula is considered useful for two coders because it takes into account the extent of intercoder agreement that may result by chance. See Scott, 1955; see also Holsti, 1969, pp. 140-41.

What constitutes acceptable reliability is best decided case by case, although analysts generally consider nothing lower than 80 to 90 percent agreement as acceptable. Low reliability estimates do not reveal whether the fault lies with the categories or with the coders. During the pretest, therefore, it is important for the analyst to identify major sources of discrepant coding and to learn the reasons for them. If the coders are assumed to be competent, low reliability estimates indicate that they are being asked to make finer discriminations than is possible with their training and understanding of the categories.

One way to resolve this problem is to contrast data known to have been coded reliably with the data that have not. This tells the analyst whether errors are concentrated in a few categories or cut across all categories. If the latter, the analyst should seriously reconsider the entire design, including the decision to use content analysis. If only a few areas are causing problems, then revising these categories (or the instructions) may solve the problem. (Fox, 1969, pp. 670-72)

Analyzing and Interpreting the Results

The main objective of content analysis is to analyze information whose format has been transformed into one that is useful. This constitutes step 6 and involves

- summarizing the coded data,
- discovering patterns and relationships within the data,
- testing hypotheses about the patterns and relationships, and
- relating the results to data obtained from other methods or situations or from assessing the validity of the analysis.

Neither these tasks nor the analytical techniques for accomplishing them are unique to content analysis. Depending on the coding design, an analyst can use a variety of statistical methods.

Summarizing Data and Examining Their Patterns

The most common means of summarizing data is by looking at frequencies among them. Absolute frequency might be the number of times statements or issues are found in the sample; a relative frequency might be represented by a percentage of the sample size. Analysts can compare one category's frequency to the average frequency for all categories, or they can note changes in frequencies over time.

Figure 2.9: Issues Addressed by HUD's Evaluation Units

Evaluation Unit / Program Area / Type of Study	Total Number of Studies	Housing Assistance Issues																Mortgage Insurance Issues								CPD Issues									None
		1 Diversion of Housing	2 Constr vs Preservation	3 Prod Subs vs Cash Asst	4 Low vs Moderate Subsidy	5 Asst Hous Role in CDBG	6 Hous Asst vs Sect 8 Progs	7 Block Grant for Sect 8	8 Sect 8 Finance Constr	9 Change Sect 8 Financing	10 Proper Admin of Progs	11 Sect 235 Serve Moder Inc	12 Use Sect 8 for FHA	13 Fair Market Rent for Sect 8	14 Public Housing Modernization	15 Alter FHA Structure	16 Ins Mortgage or Dec Paymt	17 Insur and Counsel for Default	18 Const Rules on FHA	19 Property Mgmt and Dispos	20 Interest Rate Subsidy	21 Housing's Economic Role	22 Role of CDBG	23 Declining Cities	24 Urban Dev Action Grants	25 Applic of CDBG	26 CDBG Formula Adequacy	27 CDBG Commercial Devel	28 CD Perfor Criteria	29 CDBG and Housing	30 Purp and Effect of 701	31 Contin or Discont Sect 312	No Issues	
Impact Analysis Division																																		
Mortgage Insurance																																		
Research	15															1	2	1						1									14	
Other	3															1																	2	
Block Grant																																		
Research	7																						2			2							3	
Other	3																						1		1	1	1						1	
Total Impact Analysis	30															2	2	1					4		1	3	1						20	
Special Studies Division																																		
Housing Assistance	4	1							1		3		1	2																				
Mortgage Insurance																																		
Research	1																		1															
Other	2											1							1															
Block Grant																																		
Research	0																																	
Other	1																						1				1	1						
Total Special Studies	8	1							1		3		2	2					2				1				1	1						
Grand Total	38	1							1		3		2	2		2	2	1	2				1	4		1	3	2	1				20	

Source: U.S. General Accounting Office, HUD's Evaluation System—An Assessment, PAD-78-44 (Washington, D.C.: 1978), p 7

In the assessment of HUD's evaluation system, for example, after the GAO analysts had categorized the issues addressed in 38 evaluation reports from two offices, they summarized the number of studies discussing each issue. They used absolute frequencies, and we show their grand total in figure 2.9. Within this summary, the analysts reported that 20 of the 38 documents they reviewed were not directed toward any major housing and urban development issue and that 16 issues were not addressed at all. (GAO, 1978, p. 22)

Another way of analyzing content analysis data is to examine relations among variables by cross-tabulating the co-occurrence of variables. Figure 2.9, for example, shows the relationship between the issues addressed in various reports and the evaluation units that produced the

reports. From this information, the GAO analysts identified little duplication in the way the two offices addressed the issues.

Cross-tabulations need not be limited to two or three variables. Multivariate techniques can be used to analyze complex structures. (Reynolds, 1977) Other techniques for discovering patterns and relationships in data include contingency analysis, clustering, and factor analysis; Krippendorff discusses these and others. (Krippendorff, 1980, pp. 109-18)

Assessing Validity

Whatever the technique used, a final and important task is to assess the validity of the results by relating them to other data that are known to be reasonably valid. Validity is the extent to which an instrument measures what it is intended to measure. Reliability and adequate sampling are necessary but not sufficient conditions for validating inferences made through content analysis. In addition, analysts have to corroborate the results of content analysis with other data or by other procedures that are known to be valid indicators of the phenomena they are studying.

An example of validity assessment is provided in Ramallo's analysis of volunteers' written reports of their experiences in Crossroads Africa, a Peace Corps program. (Ramallo, 1966) He hypothesized that content analysis of reports could distinguish successful volunteers from unsuccessful ones, assuming that the unsuccessful volunteers would exhibit greater alienation from their experiences. Ramallo compared his results with supervisors' ratings for the same volunteers and found a high correlation between the two, concluding that his own analysis had produced a valid measure of success.

Other equally appropriate measures could have been used to validate Ramallo's findings. Surveying the Africans with whom the volunteers had worked is one. Measuring increases in food production or decreases in infant mortality for each volunteer's assigned village are others. The use of plentiful and generally acceptable corroborating measures reduces the risk of producing misleading evaluation findings.

Writing the Report

As in writing any GAO report, analysts should explain the scope and nature of their work to indicate to their readers what they covered and what the frame of reference is for their findings. (GAO, July 1988, chapter 12.8) Readers should be given a clear idea of what was done, why it

was done, and why the results provide a sound basis for conclusions and recommendations. Figure 2.10 outlines the record of information that analysts should maintain when they use content analysis.

Figure 2.10: Minimum Documentation for a Content Analysis Study

1. The study's objectives, which governed the choice of data, methods, and study design.

2. A justification of the choice of data, methods, and design.

3. A description of the procedures (so that the research can be replicated), including descriptions of the

- sampling plans,
- units of analysis,
- coding instructions,
- results of reliability tests,
- procedures for data handling and analysis, and
- efforts at validating parts of or the entire procedure.

4. The findings and their statistical significance.

Content analysis results should be firm enough to withstand critical scrutiny. The information represented in the items mentioned in figure 2.10 may be included in the main body of the report or in appendixes, or it may remain only in the workpapers.

In either case, it should be documented well enough to enable critical readers to estimate how much they can rely on the reported results.

Summary

Content analysis is a set of procedures for transforming nonstructured, written material into a format for analysis. In this chapter, we have described those procedures. They are summarized as follows:

- deciding to use content analysis based on a project's objectives, the material that is available, and the kinds of comparison that are required;
- determining what material should be included in content analysis, which may involve sampling;
- selecting context units and recording units;
- developing coding categories, quantification levels, and coding instructions;
- pretesting the categories and then coding the material either manually or by computer;
- checking reliability during retests and throughout the coding;
- analyzing and interpreting the coded data; and
- assessing the validity of the findings.

Why Should GAO Analysts Use Content Analysis?

In this chapter, we conclude our discussion by presenting some reasons both for using and not using content analysis. We discuss some advantages and disadvantages of content analysis and give brief hypothetical cases of potential application in GAO's work.

What Content Analysis Can Do

All researchers who want to analyze written material systematically should consider content analysis. It is a means of extracting insights from already existing data sources. Therefore, it is potentially applicable to at least part of almost every project.

It Can Provide Unobtrusive Measures

Content analysis of existing written or otherwise recorded material yields unobtrusive and nonreactive measures. One problem with some experimental methods, as with surveys, is that interactions between analysts and their subjects can cause the subjects to react to the situation rather than in their more "natural" manner, and this may introduce bias into the results. Additionally, survey questions that are considered inappropriate because they invade a respondent's privacy may have to be eliminated from analysis. Content analysis of existing documents avoids both problems.

It Can Cope With Large Volumes of Written Material

Large volumes of written material can be analyzed with the help of content analysis because explicit coding instructions, precise categories, and extensive reliability checks make it possible to use any number of trained individuals to code the material. Furthermore, it allows two or more sets of coders to work on the same kind of data in different locations, such as at headquarters and in regional offices.

It Helps Analysts Learn About the Substantive Area

Content analysis can help analysts learn more about the programs they are investigating and their issues. This benefit results from two characteristics. Content analysis is systematic in nature, and its task of devising reliable and useful categories is rigorous.

It Can Validate Other Methods

In chapter 2, we discussed how to validate content analysis findings by corroborating them with findings from other methods. Validation can also move in the opposite direction. That is, findings from content analysis can be used to test the validity of findings from other measures, such as survey data and econometric proxies. Webb and others have

described how investigators can use "multiple operations" to increase confidence in their findings. (Webb et al., 1981)

Pitfalls in Using Content Analysis

We have explained some of the many reasons for using content analysis, but analysts planning to undertake content analysis should also be aware of some pitfalls that await them. The ready availability of relevant material may tempt analysts into aimless and expensive "fishing expeditions" motivated by the hope of turning up something interesting. Quantifying documentary information may produce important and interesting data, but not resisting the temptation to count things for the sake of counting is likely to produce precise but meaningless or trivial findings.

It Can Be Costly

Content analysis is relatively costly and time consuming. Interviewing users of content analysis and reviewing the literature on the method reveal three potential contributions to prohibitive cost.

1. Formulating categories that can be reliably coded is problematic, repetitive, and time consuming. The time it takes to structure and pretest categories may range from a few days to two or three months.

2. Staff have to train coders if they intend to analyze more data than they can handle themselves. Preparing a coding manual and training and supervising the coders can add a significant length of time to a project. Content analysis can be especially expensive in regard to time expended if the categorization scheme requires subtle coding decisions.

3. Coding substantial amounts of written material takes a great deal of staff time if the recording unit is small (for example, when it is words or themes), and even more time when the context unit is large (for example, when it is lengthy reports). Since coding must be systematic, it may also be tedious and arduous. Using a computer trades the coding problem for that of computerizing the text or preparing a dictionary, which can also be time consuming and therefore expensive.

It Can Pose Reliability and Validity Problems

Reliability and validity are interdependent concepts. Generally, trade-offs have to be made between them because precisely defined categories can produce results that are highly reliable and statistically significant but that lack practical significance. The need for objective and replicable results may force analysts to forego coding what they are interested in

and to code instead what can be done mechanically, thus threatening validity. Redefining categories to increase their reliability can lead to a loss of relevance—that is, a loss of validity—and, therefore, of usefulness. Because of this dilemma, validity has to be assessed after categories have been developed.

Potential Applications in Program Evaluation

How content analysis can be used in GAO's work can be understood in terms of three factors—a project's objectives, the material to be analyzed, and the kinds of analysis required. We give brief cases of hypothetical application that focus on three program evaluation objectives, showing how content analysis could be used to study them.

Identifying Program Goals

One objective of a program evaluation might be to identify the program's goals. To do this, an analyst might gather written or tape-recorded information on the program's legislative history from its authorizing legislation and congressional committee reports, from program policy documents, and from transcripts of interviews with agency officials. With content analysis, the analyst's review of this material could be made objective and systematic. Besides providing analysts with a structured format for identifying the program's goals, this technique can facilitate determination of whether those goals are congruent with legislative intent because it allows, for example, comparison of agency documents with congressional committee reports.

Describing Program Activities

A program evaluation might have as an objective a description of the program's activities. To achieve this objective, an analyst could develop case studies, attend agency meetings, or interview program managers. Information gathered in these ways would then be documented in staff workpapers. These, in turn, can be examined by means of content analysis.

From such analysis, concise, objective summaries of the material can be produced, or more complex analyses can be designed. An example would be an analysis of trends in program activities across time. The targeting of program activities could also be investigated with content analysis. Recipients of program services could be interviewed and transcripts could be made of their responses, after which their eligibility for receiving services could be examined by comparing information obtained from the interviews with established eligibility criteria.

Determining Program Results

A program evaluation might have the ascertaining of the program's results as an objective. In this situation, analysts might gather information by studying earlier evaluation reports or by surveying program participants. In surveys, open-ended questions could be appropriate for gaining information about issues, perceptions, or attitudes that cannot otherwise be identified. Analysts who do not want to impose their own concepts on survey respondents may, therefore, be unable to formulate appropriate closed questions. Using content analysis on open-ended survey data, such analysts can examine trends in program outcomes across time and compare them to changes in program activities. Alternatively, they could examine trends across groups of program participants distinguished by geographical location, age, income, and the like.

Conclusion

We hope we have given readers of this paper a realistic sense of both the advantages and disadvantages of content analysis. The method does have limitations. Without clear objectives, content analysis can produce very precise information that is, however, meaningless. The method can be costly in that formulating categories that can be reliably coded, preparing coding instructions, and training and supervising coders can all be time consuming. Additionally, complex coding schemes, which usually yield the most interesting findings, may produce the least reliable results because they entail a substantial element of coder judgment. Content analysis, therefore, requires rigorous reliability and validity checks if its results are to withstand critical scrutiny. Moreover, the results also depend on the quality of information contained in the documents being analyzed. If these are not reliable or valid, even the most rigorous content analysis will have limited value.

Nonetheless, content analysis is potentially applicable to at least part of almost all projects. Content analysis can be used at any stage of a project, but it is particularly useful at the beginning to help analysts learn about the project's substantive area. It is an excellent method for gathering retrospective information about a program from existing data sources. It does not require the collection of new data, and this means that it saves time and money. The possibilities for application we have discussed in this chapter are not exhaustive; rather, we have intended to show the method's versatility. The number and kind of areas in which content analysis can be applied and the questions it can help answer are limited primarily by its user's ingenuity and skill in structuring reliable and valid category formats.

Bibliography

Babbie, E. R. Survey Research Methods. Belmont, Calif.: Wadsworth Publishing Co., 1973.

Berelson, B. Content Analysis in Communication Research. Glencoe, Ill.: Free Press, 1952.

Carmines, E. G., and R. A. Zeller. Reliability and Validity Assessments. Beverly Hills, Calif.: Sage Publications, 1979.

Fox, D. "Techniques for the Analysis of Quantitative Data," The Research Process in Education, ed. by D. Fox. New York: Holt, Rinehart & Winston, 1969.

Holsti, 0. R. Content Analysis for the Social Sciences and Humanities. Reading, Mass.: Addison-Wesley, 1969.

Inkeles, A. "Soviet Reactions to the Voice of America." Public Opinion Quarterly, 16 (1952), 612-17.

Kaplan, A., and J. M. Golden. "The Reliability of Content Analysis Categories," The Language of Politics: Studies in Quantitative Semantics, ed. by H. W. Lasswell et al. New York: George Stewart, 1949.

Krippendorff, K. Content Analysis: An Introduction to Its Methodology. Beverly Hills, Calif.: Sage Publications, 1980.

North, R. C., et al. Content Analysis: A Handbook with Applications for the Study of International Crisis. Evanston, Ill.: Northwestern University Press, 1963.

Ramallo, L. I. "The Integration of Subject and Object in the Content of Action: A Study of Reports Written by Successful and Unsuccessful Volunteers for Field Work in Africa," The General Inquirer: A Computer Approach to Content Analysis in the Behavioral Sciences, ed. by P. J. Stone et al. Cambridge, Mass.: MIT Press, 1966.

Reynolds, H. T. Analysis of Nominal Data. Beverly Hills, Calif.: Sage Publications, 1977.

Robinson, W. S. "The Statistical Measure of Agreement." American Sociological Review, 22 (1957), 782-86.

Scott, W. A. "Reliability of Content Analysis: The Case of Nominal Scale Coding." Public Opinion Quarterly, 19 (1955), 321-25.

Spiegelman, M., et al. "The Reliability of Agreement in Content Analysis." Journal of Social Psychology, 37 (1953), 175-87.

United States General Accounting Office. Communications Manual. Washington, D.C.: July 1988.

———. HUD's Evaluation System—An Assessment, PAD-78-44. Washington, D.C.: July 20, 1978.

———. Project Manual. Washington, D.C.: December 1988.

———. Report Manual. Washington, D.C.: October 1988.

Webb, E. J., et al. Nonreactive Measures in the Social Sciences, 2nd ed. Boston, Mass.: Houghton Miflin Co., 1981.

Weber, R.P. Basic Content Analysis. Beverly Hills, Calif.: Sage Publications, 1985.