The Content Analysis Guidebook

Kimberly A. Neuendorf
Cleveland State University

PUB 2002


Sage Publications
International Educational and Professional Publisher
Thousand Oaks ■ London ■ New Delhi

CHAPTER 1

Defining Content Analysis

Content analysis is perhaps the fastest-growing technique in quantitative research. Computer advances have made the organized study of messages quicker and easier . . . but not always better. This book explores the current options in the analysis of the content of messages.

Content analysis may be briefly defined as the systematic, objective, quantitative analysis of message characteristics. It includes the careful examination of human interactions; the analysis of character portrayals in TV commercials, films, and novels; the computer-driven investigation of word usage in news releases and political speeches; and so much more. Content analysis is applicable to many areas of inquiry, with examples ranging from the analysis of naturally occurring language (Markel, 1998) to the study of newspaper coverage of the Greenhouse Effect (Miller, Boone, & Fowler, 1992) and from a description of how the two genders are shown on TV (Greenberg, 1980) to an investigation of the approach strategies used in personal ads (Kolt, 1996). Perhaps, one of the more surprising applications is Johnson's (1987) analysis of Porky Pig's vocalics from a clinical speech therapy standpoint. He examined 37 cartoons, finding that the per-cartoon stuttering ranged from 11.6% to 51.4% of words uttered, and certain behaviors were associated with the stuttering (e.g., eye blinks, grimaces). If you are unfamiliar with the range of content analysis applications, Chapter 9 presents an overview of the major areas of study—the main "contexts" of content analysis research.

The various techniques that make up the methodology of content analysis have been growing in usage and variety. In the field of mass communication research, content analysis has been the fastest-growing technique over the past 20 years or so (Riffe & Freitag, 1997; Yale & Gilly, 1988). Perhaps, the greatest explosion in analysis capability has been the rapid advancement in computer text content analysis software, with a corresponding proliferation of online archives and databases (Evans, 1996). There has never been such ready

access to archived textual messages, and it has never been easier to perform at least basic analyses with computer-provided speed and precision. This book will explore the expansion and variety of the techniques of content analysis.

In this chapter, we will follow the development of a full definition of content analysis—how one attempts to ensure objectivity, how the scientific method provides a means of achieving systematic study, and how the various scientific criteria (e.g., reliability, validity) are met. Furthermore, standards are established, extending the expectations of students who may hold a prior notion of content analysis as necessarily "easy."

Is Content Analysis "Easy"? Is It Something That Anyone Can Do?

There seem to be certain common misconceptions about the method of content analysis: Conducting a content analysis is substantially easier than conducting other types of research, content analysis is anything a scholar or student says it is, and anyone can do it without much training or forethought. It's also widely assumed that there is little reason to use content analysis for commercial or nonacademic research. Unfortunately, these stereotypes have been reinforced by academic journals that too often fail to hold content analyses to the same standards of methodological rigor as they do other social and behavioral science methods, such as surveys, experiments, and participant observation studies. Based on more than 20 years of involvement in over 100 content analyses, I would like to dispel common myths about this method before providing a full working definition.

Myth 1: **Content analysis is easy.**

Truth: **Content analysis is as easy—or as difficult—as the researcher determines it to be. It is not necessarily easier than conducting a survey, experiment, or other type of study.**

Although content analysis must conform to the rules of good science, each researcher makes decisions as to the scope and complexity of the content-analytic study. An example of a very limited—and quite easy—content analysis is shown in the summary graph in Figure 1.1, indicating how many prime-time network TV shows have dealt with medical issues over a period of 38 years. The unit of the analysis is the individual medically oriented TV program, with three simple variables measured: (a) length of show in minutes, (b) whether the show is a drama or a comedy, and (c) the year(s) the program was aired. The raw data analyzed were listings in a readily accessible source that catalogs all TV shows on the major networks since 1948 (Brooks & Marsh,

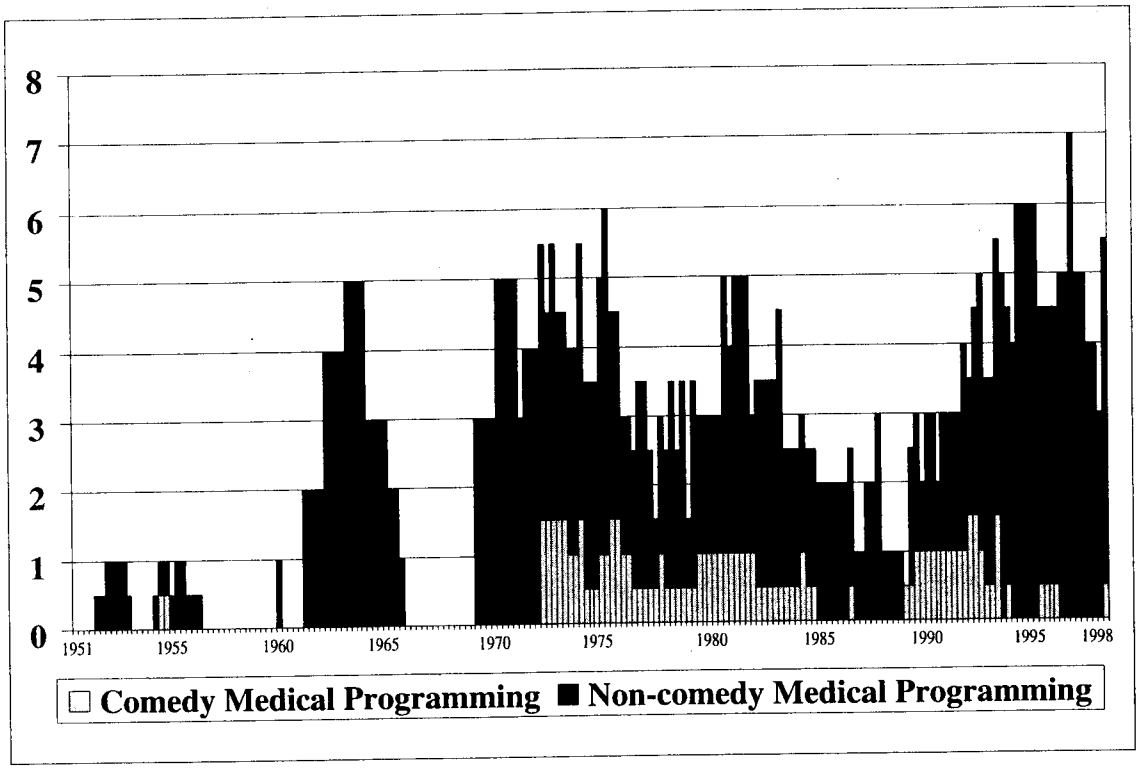

Figure 1.1. Medical Primetime Network Television Programming, 1951 to 1998 (number of hours per week)

1999). Figure 1.1 reports the findings by quarter year in a basic bar graph, indicating weekly total hours of prime-time network TV medical programming. By any assessment, this analysis would be considered easy. Correspondingly, its findings are limited in breadth and applicability. The interpretations we can make from the figure are basic: Over a 40-year period, medical shows have filled only a small portion of the prime-time period, averaging only about 4 hours per week. This has varied little over the period of study.

To make more of the findings, we must dig into the data further and examine the nature of the programs represented in the bar graph. Then, we may identify essentially two eras of TV health-related shows—the 1960s world of physician-as-God medical melodramas (e.g., *Ben Casey, Dr. Kildare*) and the 1970s-1990s era of the very human medical professional (e.g., *St. Elsewhere, ER*). Comedic medical shows have been rare, with the most successful and enduring among them being *M*A*S*H*. The 1990s included a potpourri of novel medical genres, ranging from documentary-form shows, such as *Rescue 911,* to historical dramas, such as *Dr. Quinn, Medicine Woman,* to science fiction (e.g., *Mercy Point*). Notice that these more interesting findings go beyond the content analysis itself and rely on qualitative analyses. The very simple content analysis has limited utility.

Near the tougher end of the easy-to-difficult continuum might be an ambitious master's thesis (Smith, 1999) that examined the gender role portrayals

of women in popular films from the 1930s, 1940s, and 1990s. The sampling was extremely problematic, given that no valid lists (i.e., sampling frames) of top box office hits are available for years prior to 1939. For many years after that date, all that are available are lists of the top *five* films. The researcher made the analysis even more difficult by deciding to measure 18 variables for each film and 97 variables for each primary or secondary character in each film. Some of the variables were untried in content analysis. For example, psychologist Eysenck's (1990) measures of extraversion (e.g., sociable, assertive, sensation-seeking), typically measured on individuals by self-report questionnaire, were applied to film characters, with not completely successful results. One hypothesis, that female portrayals will become less stereotypic over time, resulted in the measurement and analysis of 27 different dependent variables. With four active coders, the study took 6 months to complete; it was one of the more difficult master's theses among its contemporaries and much more difficult than many surveys and experiments.

The multifaceted results reflected the complexity and breadth of the study. The results included such wide-ranging points as (a) across the decades (1930s, 1940s, 1990s), there were several significant trends indicating a decrease in stereotypical portrayals of women in films; (b) average body shape for women varied across the decades at a near-significant level, indicating a trend toward a thinner body shape; (c) screen women who exhibited more traditional sex-role stereotyping experienced more negative life events; (d) female characters who exhibited more male sex-role traits and experienced negative life events tended to appear in films that were more successful at the box office; and (e) screen women were portrayed somewhat more traditionally in films with greater female creative control (i.e., in direction, writing, producing, or editing) (Smith, 1999).

Myth 2: **The term *content analysis* applies to *all* examinations of message content.**

Truth: **The term does not apply to every analysis of message content, only those that meet a rigorous definition. Clearly, calling an investigation a content analysis does *not* make it so.**

There are many forms of analysis—from frivolous to seminal—that may be applied to the human production of messages. Content analysis is only one type, a technique presented by this book as systematic and quantitative. Even in the scholarly literature, some confusion exists as to what may be called a content analysis. On a number of occasions, the term has been applied erroneously (e.g., Council on Interracial Books for Children, 1977; DeJong & Atkin, 1995; Goble, 1997; Hicks, 1992; Thompson, 1996), and at times, studies that warrant the term do not use it (e.g., Bales, 1950; Fairhurst, Rogers, & Sarr, 1987; Thorson, 1989).

A complete review of all the types of message analysis that compete with or complement content analysis is beyond the scope of this volume. But the reader should become aware of some of the main options for more qualitative analyses of messages (Lindlof, 1995). One good starting point is Hijmans's (1996) typology of qualitative content analyses applied to media content (according to the definitions presented in this book, we would not include these qualitative procedures as content analysis). She presents accurate descriptions of some of the main qualitative analytic methods that may be applied to messages. Based on descriptions by Hijmans (pp. 103-104) and by Gunter (2000), they are as follows:

Rhetorical Analysis

For this historically revered technique, properties of the text (both words and images) are crucial. The analyst engages in a reconstruction of manifest characteristics of text or image or both, such as the message's construction, form, metaphors, argumentation structure, and choices. The emphasis is not so much on *what* the message says as on *how* the message is presented. There is detailed reading of fragments. There is an assumption that the researcher is a competent rhetorician. This technique has a *very* long history, with its principal origins in Greek philosophy (Aristotle, 1991), and is the legitimate forebear of many of today's academic disciplines. Rhetorical analysis has been widely applied to news content, political speech, advertising, and many other forms of communication (McCroskey, 1993).

Narrative Analysis

This technique involves a description of formal narrative structure: Attention focuses on characters—their difficulties, choices, conflicts, complications, and developments. The analyst is interested not in the text as such but in characters as carriers of the story. The analysis involves reconstruction of the composition of the narrative. The assumption is that the researcher is a competent reader of narratives. One of the most complex and interesting applications of this technique is Propp's exhaustive analysis of Russian fairy tales (Propp, 1968), which establishes common character roles (e.g., hero, helper, villain, dispatcher), an identifiable linear sequence of elements in the narrative (e.g., initial situation, absentation, interdiction), and particular functions in the narrative (e.g., disguise, pursuit, transfiguration, punishment).

Discourse Analysis

This process engages in characteristics of manifest language and word use, description of topics in media texts, through consistency and connection of words to theme analysis of content and the establishment of central terms. The technique aims at typifying media representations (e.g., communicator mo-

tives, ideology). The focus is on the researcher as competent language user. Gunter (2000) identifies van Dijk's *Racism and the Press,* published in 1991, as a clear example of a large-scale discourse analysis. According to Gunter, van Dijk analyzes the "semantic macrostructures," or the overall characteristics of meanings, with regard to ethnic minorities in the news media (p. 88), concluding that minority groups are depicted as problematic.

Discourse analysis has been a popular method for analyzing public communication, with analyses ranging from the macroscopic to the very microscopic. Duncan (1996) examined the 1992 New Zealand National Kindergarten Teachers' Collective Employment Contract Negotiations and identified two discourses—"Children First" and "For the Sake of the Children." Both discourses were evident in arguments used by each side in the labor negotiations, in arguments *for* teacher pay and benefits by the teachers' representatives, and in arguments *against* such expenditures by employers and government reps. Duncan's article presents numerous direct quotes from the negotiations to support her point of view. Typical of this method, she points out that her analysis "is *one* reading of the texts, and that there will be numerous other readings possible" (p. 161).

Structuralist or Semiotic Analysis

The focus here is on deeper meanings of messages. The technique aims at deep structures, latent meanings, and the signifying process through signs, codes, and binary oppositions. Interpretations are theoretically informed, and assertions are made on central themes in culture and society. Rhetorical or narrative analysis can be preliminary to this process. The assumption is that the researcher is a competent member of the culture. (See also Eco, 1976.)

Semiotics has been a valuable technique for examining cultural artifacts. Christian Metz's (1974) classic text, *A Semiotics of the Cinema,* applies the wide range of semiotic techniques to the specific medium of narrative film. He provides a "syntagmatic" analysis of the French film, *Adieu Philippine,* indicating the *structure* of the film in shots, scenes, sequences, and the like. He also offers a detailed semiotic analysis of the self-reflexive "mirror construction" of Federico Fellini's semiautobiographical film, *8-1/2.*

Interpretative Analysis

The focus of this technique is on the formation of theory from the observation of messages and the coding of those messages. With its roots in social scientific inquiry, it involves theoretical sampling; analytical categories; cumulative, comparative analysis; and the formulation of types or conceptual categories. The methodology is clearly spelled out, but it differs from scientific inquiry in its wholly qualitative nature and its cumulative process, whereby the analyst is in a constant state of discovery and revision. The researcher is assumed to be a competent observer.

Many of the systems of analysis developed by these methods are empirical and detailed and in fact are more precise and challenging than most content analyses (e.g., Propp, 1968). With only minor adjustment, many are appropriate for use in content analysis as well (e.g., Berger, 1982, 1991).

In addition to these qualitative message analysis types reviewed by Hijmans (1996), several others deserve mention.

Conversation Analysis

Conversation analysis is a technique for analyzing naturally occurring conversations, used by social scientists in the disciplines of psychology, communication, and sociology (Sudnow, 1972). The procedure has been described as a "rigorously empirical approach which avoids premature theory construction and employs inductive methods . . . to tease out and describe the way in which ordinary speakers use and rely on conversational skills and strategies" (Kottler & Swartz, 1993, pp. 103-104). Most typically, it relies on transcribed conversations. The technique generally falls within the rubric of ethnomethodology, scholarly study in which the precise and appropriate methods used emerge from within the process of study, with the clearly subjective involvement of the investigator. Examples of its applications have included an analysis of doctor-patient interaction (Manning & Ray, 2000) and an in-depth analysis of a notorious interview of Vice President George Bush by television reporter Dan Rather as they jockeyed for position in order to control the flow of a "turbulent" interview (Nofsinger, 1988/1989).

Critical Analysis

Critical analysis, often conducted in a tradition of cultural studies, has been a widely used method for the analysis of media messages (Newcomb, 1987). The area of film studies provides a good example of a fully developed, theoretically sound literature that primarily uses the tools of critical analysis (e.g., Lyman, 1997). For example, Strong's (1996) essay about how Native Americans are "imaged" in two mid-1990s media forms—Disney Studio's *Pocahontas* and Paramount's *The Indian in the Cupboard*—is influenced heavily by her own roles as mother, musician, American raised during a period when "playing Indian" was a childhood rite of passage, and anthropologist long interested in White America's representations of Native Americans. She acknowledges these various roles and perspectives, provides precise details to back her assertions (including many lines and song lyrics from the movies), and gives summative statements that bring the details into line with cultural frameworks. For example, she concludes that "Disney has created a New Age Pocahontas to embody our millennial dreams for wholeness and harmony, while banishing our nightmares of savagery without and emptiness within" (p. 416).

Normative Analysis

Some analyses are explicitly normative or proscriptive. For example, a guide to *Stereotypes, Distortions and Omissions in U.S. History Textbooks: A Content Analysis Instrument for Detecting Racism and Sexism* (Council on Interracial Books for Children, 1977), compiled by 32 educators and consultants, provides checklists for history textbook coverage of African Americans, Asian Americans, Chicanos, Native Americans, Puerto Ricans, and women. For each group, an instrument is presented with criteria for parents and teachers to use when examining children's history texts. For instance, in the Native American checklist, the following criteria are included: "The myth of 'discovery' is blatantly Eurocentric," "War and violence were not characteristic of Native nations," "The Citizenship Act of 1924 was not a benevolent action," and "The BIA [Bureau of Indian Affairs] is a corrupt and inefficient bureaucracy controlling the affairs of one million people" (pp. 84-85). The guide is certainly well intended and a powerful tool for social change. It does not, however, fit most definitions of content analysis.

Similarly, in their article, "Evaluation Criteria and Indicators of Quality for Internet Resources," Wilkinson, Bennet, and Oliver (1997) offer a list of 125 questions to ask about a Web site. Their goal is to pinpoint characteristics that indicate accuracy of information, ease of use, and aesthetic qualities of Internet material. The work is a normative prescription for a "good" Web site. Although they call their proposal a content analysis, it does not meet the definition given in this book.

In another case of normative recommendations for message content, Legg (1996) proposes that commercial films are an important venue for the exploration of religion in American culture, and she provides tips to religious educators for using movies in teaching. She contends that "in forms like contemporary films we can see the very pertinent questions with which our culture is really wrestling" (p. 401) and urges religious educators not to limit their use of film to explicitly religious films, such as *The Ten Commandments* or *Agnes of God*. Equally useful might be explorations of manifestations of good and evil in *Batman* or a discussion of dimensions of friendship, aging, Southern ethos, prejudice, and family in *Driving Miss Daisy* (p. 403). Such detailed analyses have obvious utility; however, this process does not attempt to achieve objectivity, as does a content analysis.

Myth 3: Anyone can do content analysis; it doesn't take any special preparation.

Truth: Indeed, anyone can do it . . . but only with training and with substantial planning.

While the person who designs a content analysis must have some special knowledge and preparation, a central notion in the methodology of content analysis is that *all* people are potentially valid "human coders" (i.e., individuals who make judgments about variables as applied to each message unit). The

coding scheme must be so objective and so reliable that, once they are trained, individuals from varied backgrounds and orientations will generally agree in its application.

Clearly, however, each coder must be proficient in the language(s) of the message pool. This may require some special training for coders. To analyze natural speech, coders may actually need to learn another language or be trained in the nuances of a given dialect. Before coding television or film content, coders may have to learn about production techniques and other aspects of visual communication. To code print advertising, coders may need to learn a bit about graphic design. All this is in addition to training *with the coding scheme,* which is a necessary step for all coders.

For analyses that do not use human coders (i.e., those that use computer coding), the burden rests squarely on the researcher to establish complete and carefully researched dictionaries or other protocols. Because the step of making sure coders can understand and reliably apply a scheme is missing, the researcher needs to execute additional checks. Chapter 6 presents some notions on how this might be done.

Myth 4: Content analysis is for academic use only.

Truth: Not.

The vast majority of content analyses have been conducted by academics for scholarly purposes. However, there has been growing interest among commercial researchers and communication practitioners in particular applications of content analysis. A law firm hired a respected senior professor to conduct content analyses of news coverage of their high-profile clients, to be used as evidence in conjunction with a change-of-venue motion (i.e., excessive and negative coverage may warrant moving a court case to another city in order to obtain a fair trial; McCarty, 2001). In response to criticisms, a Southern daily newspaper hired a journalism scholar to systematically document their coverage of the local African American community (Riffe, Lacy, & Fico, 1998). The marketing research unit of a large-city newspaper has begun the process of systematically comparing its own coverage of regional issues with that provided by local television news. Organizational communication consultants sometimes include a content analysis of recorded messages (e.g., e-mail, memos) in their audit of the communication flow in the organization. And the clinical diagnostic tools of criteria-based content analysis have been used in nonacademic settings by psychologists and legal professionals.

A Six-Part Definition of Content Analysis

This book assumes that content analysis is conducted within the scientific method but with certain additional characteristics that place it in a unique position as the primary message-centered methodology.

Box 1.1 Defining Content Analysis

Some of the main players in the development of quantitative message analysis present their points of view:

Berelson (1952, p. 18): Content analysis is a research technique for the objective, systematic, and quantitative description of the manifest content of communication.

Stone, Dunphy, Smith, & Ogilvie (1966, p. 5, with credit given to Dr. Ole Holsti): Content analysis is any research technique for making inferences by systematically and objectively identifying specified characteristics within text.

Carney (1971, p. 52): The general purpose technique for posing questions to a "communication" in order to get findings which can be substantiated. . . . [T]he "communication" can be anything: A novel, some paintings, a movie, or a musical score—the technique is applicable to all alike and *not* only to analysis of literary materials.

Krippendorff (1980, p. 21): Content analysis is a research technique for making replicable and valid inferences from data to their context.

Weber (1990, p. 9): Content analysis is a research method that uses a set of procedures to make valid inferences from text.

Berger (1991, p. 25): Content analysis . . . is a research technique that is based on measuring the amount of something (violence, negative portrayals of women, or whatever) in a representative sampling of some mass-mediated popular art form.

Riffe, Lacy, & Fico (1998, p. 20): Quantitative content analysis is the systematic and replicable examination of symbols of communication, which have been assigned numeric values according to valid measurement rules, and the analysis of relationships involving those values using statistical methods, in order to describe the communication, draw inferences about its meaning, or infer from the communication to its context, both of production and consumption.

This book: Content analysis is a summarizing, quantitative analysis of messages that relies on the scientific method (including attention to objectivity-intersubjectivity, a priori design, reliability, validity, generalizability, replicability, and hypothesis testing) and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented.

Content analysis is a summarizing, quantitative analysis of messages that relies on the scientific method (including attention to objectivity-intersubjectivity, a priori design, reliability, validity, generalizability, replicability, and hypothesis testing) and is not limited as to the types of variables that may be measured or the context in which the messages are created or presented.

Box 1.1 presents some alternative definitions of content analysis for comparison's sake. More details on this book's definition are presented in the discussion that follows.

1. Content Analysis as Relying on the Scientific Method

Perhaps, the most distinctive characteristic that differentiates content analysis from other, more qualitative or interpretive message analyses is the attempt to meet the standards of the scientific method (Bird, 1998; Klee, 1997);

by most definitions, it fits the positivism paradigm of social research (Gunter, 2000)[1]. This includes attending to such criteria as the following:

Objectivity-Intersubjectivity

A major goal of any scientific investigation is to provide a description or explanation of a phenomenon in a way that avoids the biases of the investigator. Thus, objectivity is desirable. However, as the classic work, *The Social Construction of Reality* (Berger & Luckman, 1966), points out, there is no such thing as true objectivity—"knowledge" and "facts" are what are socially agreed on. According to this view, all human inquiry is inherently subjective, but still we must strive for consistency among inquiries. We do not ask, "is it true?" but rather, "do we agree it is true?" Scholars refer to this standard as *intersubjectivity* (Babbie, 1986, p. 27; Lindlof, 1995).

Another set of terms sometimes used is the comparison between *idiographic* and *nomothetic* investigations. An idiographic study seeks to fully describe a single artifact or case from a phenomenological perspective and to connect the unique aspects of the case with more general truths or principles. A nomothetic study hopes to identify generalizable findings, usually from multiple cases, and demands "specific and well-defined questions that in order to answer them it is desirable to adopt standardized criteria having known . . . characteristics" (Te'eni, 1998). Idiographic study implies conclusions that are unique, nongeneralizable, subjective, rich, and well grounded; nomothetic study implies conclusions that are broadly based, generalizable, objective, summarizing, and inflexible.

An A Priori Design

Although an a priori (i.e., "before the fact") design is actually a part of the task of meeting the requirement of objectivity-intersubjectivity, it is given its own listing here to provide emphasis. Too often, a so-called content analysis report describes a study in which variables were chosen and "measured" *after* the messages were observed. This wholly inductive approach violates the guidelines of scientific endeavor. All decisions on variables, their measurement, and coding rules must be made before the observations begin. In the case of human coding, the codebook and coding form must be constructed in advance. In the case of computer coding, the dictionary or other coding protocol must be established a priori.

However, the self-limiting nature of this "normal science" approach should be mentioned. As Kuhn's (1970) seminal work on paradigms has pointed out, deduction based on past research, theories, and bodies of evidence within the current popular paradigm does not foster innovation. Content analysis has a bit of this disadvantage, with the insistence that coding schemes be developed a priori. Still, creativity and innovation can thrive within the method. As described in Chapter 5, a lot of exploratory work can and

should be done before a final coding scheme is "set in stone." The entire process may be viewed as a combination of induction and deduction.

Reliability

Reliability has been defined as the extent to which a measuring procedure yields the same results on repeated trials (Carmines & Zeller, 1979). When human coders are used in content analysis, this translates to *intercoder reliability,* or level of agreement among two or more coders. In content analysis, reliability is paramount. Without acceptable levels of reliability, content analysis measures are meaningless. Chapter 7 addresses this important issue in detail.

Validity

Validity refers to the extent to which an empirical measure adequately reflects what humans agree on as the real meaning of a concept (Babbie, 1995, p. 127). Generally, it is addressed with the question, "Are we really measuring what we want to measure?" Although in content analysis, the researcher is the boss, making final decisions on what concepts to measure and how to measure them, there are a number of good guidelines available for improving validity (Carmines & Zeller, 1979). Chapter 6 gives a more detailed discussion.

Generalizability

The generalizability of findings is the extent to which they may be applied to other cases, usually to a larger set that is the defined *population* from which a study's sample has been drawn. After completing a poll of 300 city residents, the researchers obviously hope to generalize their findings to all residents of the city. Likewise, in a study of 800 personal ads in newspapers, Kolt (1996) generalized his findings to all personal ads in U.S. newspapers in general. He was in a good position to do so because he (a) randomly selected U.S. daily newspapers, (b) randomly selected dates for specific issues to analyze, and then (c) systematically random sampled personal ads in each issue. In Chapter 4, the options for selecting representative samples from populations will be presented.

Replicability

The replication of a study is a safeguard against overgeneralizing the findings of one particular research endeavor. Replication involves *repeating* a study with different cases or in a different context, checking to see if similar results are obtained each time (Babbie, 1995, p. 21). Whenever possible, research reports should provide enough information about the methods and protocols so that others are free to conduct replications. Throughout this

book, the assumption is made that *full* reportage of methods is optimal, for both academic and commercial research.

As Hogenraad and McKenzie (1999) caution, content analyses are sometimes at a unique disadvantage with regard to replication. Certain messages are historically situated, and repeated samplings are not possible, as with their study of political speeches leading up to the formation of the European Union. They propose an alternative—bootstrap replication—which compares and pools multiple random subsamples of the original data set.

Hypothesis Testing

The scientific method is generally considered to be hypothetico-deductive. That is, from theory, one or more hypotheses (conjectural statements or predictions about the relationship among variables) are derived. Each hypothesis is tested deductively: Measurements are made for each of the variables, and relationships among them are examined statistically to see if the predicted relationship holds true. If so, the hypothesis is supported and lends further support to the theory from which it was derived. If not, the hypothesis fails to receive support, and the theory is called into question to some extent. If existing theory is not strong enough to warrant a prediction, a sort of fallback position is to offer one or more research questions. A research question poses a query about possible relationships among variables. In the deductive scientific model, hypotheses and research questions are both posed *before* data are collected. Chapter 5 presents examples of hypotheses and research questions appropriate to content analysis.

2. The Message as the Unit of Analysis, the Unit of Data Collection, or Both

The unit in a research study is the individual "thing" that is the subject of study—what or whom is studied. Frequently, it is useful to distinguish between the *unit of data collection* (sometimes referred to as the *unit of observation;* Babbie, 1995) and the *unit of analysis,* although in particular studies, these two things are often the same. The unit of data collection is the element on which each variable is measured. The unit of analysis is the element on which data are analyzed and for which findings are reported. In most social and behavioral science investigations, the individual *person* is both the unit of data collection and the unit of analysis.

For example, when a survey of city residents is conducted to measure opinions toward the president and the mayor, let's say, the unit of data collection is the individual respondent—the person. That is, telephone interviews are conducted, and normally, each person responds alone. The variables (e.g., attitude toward the president, attitude toward the mayor, gender, age) are measured on each unit. The unit of analysis is also typically the individual per-

son. That is, in the data set, each respondent's answers will constitute one line of data, and statistical analyses will be conducted on the data set, with *n* equalling the number of people responding. When "average rating of confidence in the president" is reported as 6.8 on a 0-to-10 scale, that's the mean based on *n* respondents.

Sometimes, the unit of data collection and the unit of analysis are not the same. For example, a study of marital discord may record interactions between married partners. The unit of data collection may be the "turn" in verbal interaction: Each time an individual speaks, the tone and substance of his or her turn may be coded. However, the ultimate goal of the study may be to compare the interactions of those couples who have received intervention counseling and those who have not. Thus, the unit of *analysis* may be the dyad, pooling information about all turns and interactions for each married pair.

In content analysis, the unit of data collection *or* the unit of analysis—or both—must be a *message unit.* Quite simply, there must be communication *content* as a primary subject of the investigation for the study to be deemed a content analysis. In the marital-discord example just described, the unit of data collection is a message unit (an interaction turn), and the unit of analysis is not. It may be called a content analysis. Chapter 4 provides more examples of unitizing.

3. Content Analysis as Quantitative

The goal of any quantitative analysis is to produce *counts* of key categories, and measurements of the *amounts* of other variables (Edward L. Fink, personal communication, March 26, 1999). In either case, this is a numerical process. Although some authors maintain that a nonquantitative (i.e., "qualitative") content analysis is feasible, that is not the view presented in this book. A content analysis has as its goal a numerically based summary of a chosen message set. It is neither a gestalt impression nor a fully detailed description of a message or message set.

There is often confusion between what is considered *quantitative* and what is considered *empirical.* Empirical observations are those based on real, apprehendable phenomena. Accordingly, both quantitative and qualitative investigations may be empirical. What, then, is *not* empirical? Efforts to describe theory and conditions without making observations of events, behaviors, and other "real" subjects, such as abstract theorizing, many aspects of the discipline of philosophy, and (perhaps surprisingly) certain types of scholarship in mathematics (ironically, quite quantitative in focus). Much of the social and behavioral science literature is based on empirical work, which may be quantitative or qualitative.

It should be made clear at the outset that this book takes the viewpoint that critical and other qualitative analyses that are empirical are typically extremely useful. They are capable of providing a highly valid source of detailed or "deep" information about a text. (Note that the term *text* is a preferred

term in many critical analyses and denotes not just written text but also any other message type that is considered in its entirety. For example, the text of a film includes its dialog, its visuals, production techniques, music, characterizations, and anything else of meaning presented in the film.) The empiricism of a careful and detailed critical analysis is one of its prime strengths and may produce such a lucid interpretation of the text as to provide us with a completely new encounter with the text. Such an analysis may bring us into the world of the text (e.g., into what is called the *diegesis* of a film, "the sum of a film's denotation: the narration itself, but also the fictional space and time dimensions implied in and by the narrative, and consequently the characters, the landscapes, the events, and other narrative elements" [Metz, 1974, p. 98]). It may illuminate the intentions of the source of the text, or it may allow us to view the text through the eyes of others who may experience the text (e.g., as in providing an understanding of a child's view of the *Teletubbies,* something that may be essential to a full appreciation of such a child-centric television program).

When approaching a text—a message or message set—the researcher needs to evaluate his or her needs and the outcomes possible from both quantitative (i.e., content analysis) and nonquantitative analyses. For example, to identify and interpret pacifist markers in the film *Saving Private Ryan,* a critical analysis, perhaps with a Marxist approach, is in order. To establish the prevalence of violent acts in top-grossing films of the 1990s, a content analysis is more appropriate. The content analysis uses a broader brush and is typically more generalizable. As such, it is also typically less in-depth and less detailed.

The outlook of this book coincides nicely with the view presented by Gray and Densten (1998): "Quantitative and qualitative research may be viewed as different ways of examining the same research problem" (p. 420). This *triangulation* of methods "strengthens the researcher's claims for the validity of the conclusions drawn where mutual confirmation of results can be demonstrated" (p. 420). It is rare to find a single investigation that combines methods in this way, but such triangulated studies do exist. One study examined storytelling in Taiwanese and European American families, combining ethnographic fieldwork with content-analytic coding of audio and video recordings of naturally occurring talk in the home (Miller, Wiley, Fung, & Liang, 1997).

4. Content Analysis as Summarizing

As noted in the previous point, a content analysis summarizes rather than reports all details concerning a message set. This is consistent with the nomothetic approach to scientific investigations (i.e., seeking to generate generalizable conclusions), rather than the idiographic approach (i.e., focusing on a full and precise conclusion about a particular case).

The goal of some noncontent analysis message studies may be a type of microdocumenting, as in a syntagmatic approach to analyzing transcribed speech or written text (Propp, 1968). The computer program, NUD*IST

(Non-Numerical Unstructured Data Indexing Searching and Theorising computer software; see also *The Content Analysis Guidebook Online*), is primarily oriented to this type of detailed markup, retrieval, and description of textual content. It is based on the organization of coded text via a system of concept nodes, grouped hierarchically in a tree structure, which is displayed by the program. Buston (1997) gives a cogent description of the use of NUD*IST to organize and make sense of a set of 112 interviews with young people with chronic health problems (e.g., asthma). In addition, Buston provides reflections on the ways in which NUD*IST affects qualitative methodologies, concluding that using NUD*IST "is *not* exactly the same as working using manual methods only." On the positive side, it speeds up mundane, routine tasks. On the negative side, it may lead to " 'coding fetishism,' indexing anything and everything obsessively and unnecessarily" (p. 12).

Another program, HyperRESEARCH, a computer-assisted program for conducting qualitative assessments of multimedia, was demonstrated by Hesse-Biber, Dupuis, and Kinder (1997) to be useful for identifying and indexing (what they term *coding*) a broad mix of photographs, text samples, audio segments, and video segments. They also point out the program's utility in searching and reporting based on the codes. Again, though, the emphasis is on cataloging discrete exemplars of desired content in a manner that makes their retrieval and comparison easy. For example, after indexing is complete, the researchers might query the program to produce all examples that have been tagged "expression of self-esteem" (p. 7). These cases may be examined and cross-indexed according to other characteristics, but the responsibility for making sense of these interwoven networks of similarities rests with the researcher. This is somewhat different than the summarizing function of content analysis.

Historians have contributed a number of examples of very precise, fully explicated analyses that rely on original textual sources. Because these analyses are based on texts, we might be tempted to call them content analyses. But some of them display an obvious attempt to report all possible details across a wide variety of units of data collection rather than to summarize information for a chosen unit of data collection or analysis. One example is Kohn's (1973) book on Russia during World War I, in which he professes to attempt "an *exhaustive* inquiry into the vital statistics of Russia" (p. 3), ultimately to assess the economic and noneconomic consequences of the war on Russian society. The work is largely a reportage of numerical facts taken from a variety of textual sources. Another example, a book about the *Plantation Slaves of Trinidad,* brings the reader into the daily lives of these Caribbean slaves during the nation's slave period of 1783-1816 (John, 1988). Aggregate figures on slave mortality and childbearing are presented side by side with drawings of slave life on the Trinidad plantations.

In contrast, the content analysis summarizes characteristics of messages. In Kolt's (1996) study of personal ads in newspapers, he found 26% of them offered physical attractiveness, whereas only 8% requested physical attractive-

ness of the reader. Lin (1997) found that women in network TV commercials were nearly twice as likely to be shown as "cheesecake" (i.e., physical appearance "obviously alluring") as men were as "beefcake." These results summarize characteristics of the message pool rather than focusing on specific cases.

5. Content Analysis as Applicable to *All* Contexts

The term *content analysis* is not reserved for studies of mass media or for any other type of message content. So long as other pertinent characteristics apply (e.g., quantitative, summarizing), the study of any type of message pool may be deemed a content analysis. The messages may be mediated—that is, having some message reproduction or transmittal device interposed between source and receiver. Or they may be nonmediated—that is, experienced face to face. Although not attempting to create an exhaustive typology of communication purposes and context, the sections to follow give some examples of the range of applications of the techniques of content analysis.

Individual Messaging

Some analyses examine the creation of messages by a single individual, with the typical goal of making some inference to that source.

In psychology, there is a growing use of content analysis of naturally produced text and speech as a type of psychometric instrument (Gottschalk, 1995; Horowitz, 1998; Tully, 1998). This technique analyzes statements made by an individual to diagnose psychological disorders and tendencies, to measure psychological traits of the source, or to assess the credibility of the source (Doris, 1994). Nearly all these efforts stem from the work of Philip Stone (Stone, Dunphy, Smith, & Ogilvie, 1966) in the Harvard Department of Social Relations. His "General Inquirer" computer program was the first to apply content-analytic techniques to free-speech words (see Chapter 2). Rosenberg and others (e.g., Rosenberg & Tucker, 1976) applied the computer technique to the language of schizophrenics, with the goal of better diagnosis. In an example of a further refinement of such procedures, Broehl and McGee (1981) analyzed the writings of historical figures—three British lieutenants serving during the Indian Mutiny of 1957-1958—and on this basis developed psychological profiles for the officers. Even the Watergate tapes have been studied using content analysis to gain insights into the underlying psychological motives of the individuals involved (Weintraub & Plant, as cited in Broehl & McGee, 1981, p. 288).

Others in the field of psychology have continued to develop computer analyses that produce diagnoses from written or spoken text. For example, Gottschalk, Stein, and Shapiro (1997) compared results from standard psychometric tests, such as the MMPI (Minnesota Multiphasic Personality Inventory), with content analysis results from a computer analysis of transcripts of 5-minute speeches. Their study of 25 new psychiatric outpatients found

strong construct validity—the speech analyses were highly correlated with corresponding questionnaire outcomes. They point out the potential value in being able to use ordinary spoken or written material for an initial, rapid diagnostic appraisal that can easily remain unobtrusive (i.e., the individual does not have to submit to a lengthy questionnaire administration; p. 427). The content analysis scheme used, the 16-part Gottschalk-Gleser Content Analysis Scales, is a software program developed and validated over a period of many years.

Another application of content analysis to the individual as message generator is the common method of coding responses to open-ended questionnaire items and in-depth interviews (Gray & Densten, 1998). Although the first steps in this process usually include a qualitative review of the message pool and the development of an emergent coding scheme based on what's represented in the pool, it must be remembered that the true content analysis portion is the subsequent careful application of the a priori coding scheme to the message pool.

In the fields of linguistics, history, and literature, some attempts have been made at analyzing individual authors or other sources. Most recently, computer text content analyses have been conducted either to describe a source's style, to verify a questionable source, or to identify an unknown source (Floud, 1977; Olsen, 1993). For example, Elliott and Valenza's (1996) "Shakespeare Clinic" has developed computer tests for Shakespeare authorship, and Martindale and McKenzie (1995) used computer text content analysis to confirm James Madison's authorship of *The Federalist.*

Interpersonal and Group Messaging

This book assumes a definition of interpersonal communication that acknowledges the *intent* of the messaging to reach and be understood by a *particular individual.* This may occur face to face, or it may be mediated, as in the cases of telephoning or e-mailing. It may occur in a dyad or a small group.

To study face-to-face group processes, Bales (1950) developed a content analysis scheme that calls for the coding of each communication act. A verbal act is "usually the simple subject-predicate combination," whereas a nonverbal act is "the smallest overt segment of behavior that has 'meaning' to others in the group" (Bales, Strodtbeck, Mills, & Roseborough, 1951, p. 462). Each act is coded into one of 12 categories: (a) shows solidarity, (b) shows tension release, (c) agrees, (d) gives suggestion, (e) gives opinion, (f) gives orientation, (g) shows antagonism, (h) shows tension, (i) disagrees, (j) asks for suggestion, (k) asks for opinion, or (l) asks for orientation. Bales's scheme has been widely used and elaborated on (Bales & Cohen, 1979) and has also been adapted for analyzing human interaction in mass media content (Greenberg, 1980; Neuendorf & Abelman, 1987).

Box 1.2 Analyzing Communication in Crisis

Perpetrator and Negotiator Interpersonal Exchanges

Most standoffs between police and perpetrators are resolved nonviolently. An analysis of 137 crisis incidents handled by the New York City Police Department revealed that in 91% of the cases, neither hostages nor hostage takers were killed (Rogan & Hammer, 1995, p. 554). Nonetheless, those crisis situations that end violently—such as the 1993 Branch Davidian conflagration in Waco, Texas—focus attention on the need to better understand the negotiation process. There is interest among scholars and police professionals alike in studying the communication content of negotiations in crisis situations so that outcomes may be predicted and negative outcomes prevented.

Rogan and Hammer (1995) had such a goal for their content analysis of audio recordings of three authentic crisis negotiations obtained from the FBI training academy. They looked at message affect, a combination of message valence and language intensity, across eight phases of each negotiation process. The unit of data collection was the uninterrupted talking turn. Each turn was coded by human coders for positive-negative valence and for Donohue's (1991) five correlates of language intensity: (a) obscure words, (b) general metaphors, (c) profanity and sex, (d) death statements, and (e) expanded qualifiers. The analysis was highly systematic and achieved good reliability (i.e., agreement between independent coders).

Total "message affect" scores were calculated for perpetrator and negotiator for each of the eight time periods in each negotiation. In all three situations, the negotiator's message profile remained positive throughout, whereas the perpetrator's score became more strongly negative during stages 2 and 3. Eventually, between stages 6 and 8, the perpetrator's message affect shifted to a positive valence, approaching that of the negotiator. In the one successful negotiation studied, the perpetrator's scores remained high and positive; in the two unsuccessful incidents (one culminating in the perpetrator's suicide), the perpetrator's scores began an unrelenting slide to intense negativity at stage 6 or 7.

The researchers point out certain limitations of the study—primarily, that the analysis was limited to message affect, with no consideration of other characteristics of the communicators, no examination of substantive or relational communication content, and so on. Nevertheless, just based on message affect, the results are striking. By looking at the charted message affect scores, you can *visualize* the process of negotiation success or failure. Although not useful at this point for real-time application to ongoing crisis situations, this content analysis technique shows promise for the development of such application in the future. And researching past negotiation successes and failures provides practitioners insight into the dynamics of the process. As Rogan and Hammer (1995) note, "ultimately, such insight could enable a negotiator to more effectively control a perpetrator's level of emotional arousal, such that a negotiator could take actions to reduce a perpetrator's highly negative and intense emotionality in an effort to negate potentially violent behavior" (p. 571), perhaps the ultimate useful application of the technique of content analysis.

Box 1.3 The Variety of Content Analysis

Religious TV—Tapping Message Characteristics, Ranging From Communicator Style to Dollar Signs

In the 1980s, religious broadcasting reached a peak of popularity with the rapid growth of "televangelism" (Frankl, 1987). Concerned with a growing perception of religious broadcasting as invasive and inordinately focused on fund-raising, the organization of Roman Catholic broadcasters, UNDA-USA, commissioned a set of content analyses. During the mid-1980s, researchers at Cleveland State University conducted an extensive six-part project. All the components of the project were quantitative content analyses, and they drew on a wide array of theories and research perspectives.

A set of 81 episodes of religious programs provided the content to be analyzed. These episodes were three randomly sampled episodes for each of the top religious television or cable programs, as determined by an index of availability in a random sample of 40 U.S. towns and cities. These programs ranged from talk format shows, such as *The 700 Club,* to televangelist programs like *Jim Bakker* to narrative forms, such as the soap opera *Another Life* and the children's stop-motion animated "daily lesson" program, *Davey and Goliath.* Teams of coders were trained for five types of analysis:

1. The demography of religious television:
   With the unit of analysis the individual character (real or fictional), a dozen demographic variables were assessed (based on previous content analyses of TV characters, such as Greenberg [1980] and Gerbner, Gross, Morgan and Signorielli [1980]), including social age (child, adolescent, young adult, mature adult, elderly), occupation, and religious affiliation. An example of the results is the finding that 47% of the characters were mature adults, with 37% being young adults. Children constituted only 7% of the sample, with the elderly at only 5% (Abelman & Neuendorf, 1984a).

2. Themes and topics on religious television:
   Here, the unit of analysis was a period of *time*: the 5-minute interval. At the end of each 5-minute period, a checklist coding form was completed by the coder, with 60 measures indicating simple presence or absence of a given social, political, or religious topic within all verbalizations in the period (pulling from existing analyses of religious communication, e.g., Hadden & Swann, 1981). Also, both explicit and implied appeals for money were recorded at the end of

Organizational Messaging

Content analysis has been used less frequently for profiling messages within a defined organization than it has in other contexts. More often, messages within an organization have been scrutinized using more qualitative techniques (Stohl & Redding, 1987). An assortment of content analyses in the organizational context have used a variety of techniques.

each 5-minute period. Overall, $328.13 was explicitly requested of the viewer per hour across the sample of religious programs (Abelman & Neuendorf, 1985a, 1985b).

3. Interaction analysis of religious television content:
   Using a scheme derived and adapted from Bales (1950), Borke (1969), and Greenberg (1980), interpersonal interactions among characters on religious television were examined. The unit of analysis was each verbal utterance (act), which was coded as falling into one of 20 modes (e.g., offering information, seeking support, attacking, evading). The results suggested age and gender differences in interaction patterns; most interactions were male dominated, and the elderly were often shown as conflict-producing individuals who were the frequent targets of guidance from those who were younger (Neuendorf & Abelman, 1987).

4. Communicator style of televangelists:
   Drawing on the considerable interpersonal communication literature on communicator style, notably the work of Robert Norton (1983), this aspect of the project targeted the 14 televangelists in the program sample and used as the unit of analysis each verbal utterance within a monologue. Each utterance was coded for a variety of characteristics, including mode (similar to the interaction coding scheme), vocal intensity, pace, and facial nonverbal intensity. Based on an overall intensity index, the top three "most intense" televangelists were James Robison, Robert Schuller, and Ernest Angley (Neuendorf & Abelman, 1986).

5. Physical contact on religious television programming:
   Drawing on work in nonverbal communication (e.g., Knapp, 1978), this portion of the content analyses examined physical touch. The unit of analysis was the instance of nonaccidental physical contact. Characteristics of the initiator and recipient of the touching were tapped, as were type of touch (religious in nature, nonreligious), anatomical location of the touch, and the recipient's reaction to the touch. A sample result was that there was a clear similarity with real-life touching along gender lines: Males were the primary initiators of physical contact, and it tended to be rather formal and ritualistic (i.e., a substantial portion of the contact was religious in nature, e.g., healing; Abelman & Neuendorf, 1984b).

Organizational applications of content analysis have included the analysis of open-ended responses to employee surveys (DiSanza & Bullis, 1999), the word network analysis of voice mail (Rice & Danowski, 1991), and the application of interpersonal interaction coding to manager-subordinate control patterns (Fairhurst et al., 1987). Developing a novel coding scheme, Larey and Paulus (1999) analyzed the transcripts of brainstorming discussion

groups, of four individuals, looking for unique ideas. They found that interactive groups were less successful in generating unique ideas than were "nominal," noninteractive groups.

Mass Messaging

Mass messaging is the creation of messages that are ***intended*** for a relatively ***large, undifferentiated audience.*** These messages are most commonly mediated (e.g., via television, newspaper, radio), but they do not necessarily have to be, as in the case of a public speech.

Mass messages have been heavily studied by sociologists, social psychologists, communication scientists, marketing and advertising scholars, and others. Fully 34.8% of the mass communication articles published during 1995 in *Journalism and Mass Communication Quarterly,* one of the most prominent mass communication journals, were content analyses (Riffe & Freitag, 1997). The range of types of investigations is staggering, although some areas of study are much better represented in the content analysis literature than others; for instance, studies of journalistic coverage are common, whereas studies of films are rare.

Applied Contexts

In addition to the aforementioned means of dividing up message contexts, we might also consider such applied contexts as health communication, political communication, and the Internet, all of which transcend the distinctions of interpersonal, group, organizational, and mass communication. That is, content analyses within the health context might include analyses of doctor-patient interaction (interpersonal), the flow of e-mail among hospital employees (organizational), and images of medical professionals on TV (mass; Berlin Ray & Donohew, 1990). Yet all these studies would be better informed by a clear grasp of the norms, values, behaviors, legal constraints, and business practices within the healthcare environment. Thus, a special consideration of such applied contexts is useful. A number of these are considered in Chapter 9.

Some applications of content analysis may be highly practical. Rather than attempting to answer questions of theoretical importance, some analyses are aimed at building predictive power within a certain message arena. Box 1.2 highlights one such study. Rogan and Hammer (1995) applied a scheme to actual crisis negotiation incidents, such as hostage taking. Their findings offer insight into message patterns that may predict successful and unsuccessful resolutions to crisis incidents.

Another applied context is that of religious broadcasting. Box 1.3 describes a set of studies that took into consideration the special nature of religion on television. A variety of communication and religious perspectives informed the analyses, ranging from interpersonal communication theories to practical considerations of religious broadcasting.

6. All Message Characteristics Are Available to Content Analyze

This book takes a broad view of what types of messages and message characteristics may be analyzed. A few clarifications on terminology are in order:

Manifest Versus Latent Content

Much of the content analysis literature has concentrated on *manifest* content, the "elements that are physically present and countable" (Gray & Densten, 1998, p. 420). An alternative is to also consider the *latent* content, consisting of unobserved concept(s) that "cannot be measured directly but can be represented or measured by one or more . . . indicators" (Hair, Anderson, Tatham, & Black, 1998, p. 581). These two types of content are analogous to "surface" and "deep" structures of language and have their roots in Freud's interpretations of dreams.[2]

Although the early definition of content analysis by Berelson (1952) indicated that it is ordinarily limited to manifest content only, numerous others have boldly attempted to tap the deeper meanings of messages. For example, in the Smith (1999) study, the latent construct, "sexism," was measured by 27 manifest variables that tapped "stereotypic images of women," extracted from a variety of theoretic works (largely from feminist literature) and critical, qualitative analyses of film (e.g., Haskell, 1987).

In the case of Ghose and Dou's (1998) study of Internet Web sites, the latent variable, "interactivity" (conceptualized as related to "presence," or a sense of "being there"), was represented by 23 manifest variables that are easily measurable, such as presence or absence of a key word search, electronic couponing, online contests, and downloading of software. Although serving as the theoretic core of the study, interactivity is sufficiently abstract as to require that its more concrete elements be defined for actual measurement.

Gray and Densten (1998) promote the use of latent constructs as a way of integrating quantitative content analysis and qualitative message analysis. They used both methods to study the broad latent concept, locus of control (from Rotter's internal/external locus of control construct: An individual holding a more external locus of control feels that his or her life events are the product of circumstances beyond his or her personal control; p. 426). Their findings indicate a surprising correspondence between quantitative and qualitative methods in the discovery of new locus-of-control dimensions reflected in a variety of very specific manifest indicators.

A number of researchers have criticized any dependence on the manifest-latent dichotomy, noting the often fuzzy distinction between the two (Potter & Levine-Donnerstein, 1999; Shapiro & Markoff, 1997). It is perhaps more useful to think of a continuum from "highly manifest" to "highly latent" and to address issues of subtlety of measurement for those messages that are very latent (and therefore a challenge for objective and reliable measurement).

Another perspective one may take is that you can't measure latent content without using manifest variables. However, not all researchers would agree with this heuristic.

Content Versus Form Characteristics

Many scholars have differentiated between content and form elements of a mediated message (Berelson, 1952; Huston & Wright, 1983; Naccarato & Neuendorf, 1998). Content attributes—sometimes called *substance characteristics*—are those that may appear or exist in any medium. They are generally able to survive the translation from medium to medium. Form attributes—often called *formal features,* although there's usually nothing formal about them in the colloquial sense—are those that are relevant to the medium through which the message is sent. They are in a sense contributed by the particular medium or form of communication.

For example, the examination of self-disclosure by women to other women has been analyzed for movie characters (Capwell, 1997). The same measures of level and type of self-disclosure could be used to analyze naturally occurring discussions between women, interactions between characters on TV programs or commercials, or relationship building between characters in novels. The measures are *content* measures, applicable regardless of the medium. On the other hand, measurements of the type of camera shot (e.g., close-up vs. long shot) used when self-disclosure occurs in a film is a measure of *form,* how the content is treated in a particular medium.

Even though the distinction between content and form is an important one, the primary focus should not be on placing each variable in one category or the other. Some variables may be on the fine line between the two types, exhibiting characteristics of each. What's important is that both content and form characteristics of messages ought to be considered for *every* content analysis conducted. Form characteristics are often extremely important mediators of the content elements. Huston and Wright (1983) have summarized how formal features of TV influence the cognitive processing of TV content, notably for children. This speaks once again to the importance of the content analyst becoming well versed in the norms and syntax of the medium he or she chooses to study.

Text Analysis Versus Other Types of Content Analysis

You'll notice that some of the classic definitions of content analysis shown in Box 1.1 apply the term *only* to analyses of text (i.e., written or transcribed words). The view presented in this book is not so limiting. Content analysis may be conducted on written text, transcribed speech, verbal interactions, visual images, characterizations, nonverbal behaviors, sound events, or any other message type. In this book, the term *content analysis* encompasses all

such studies; the terms *text analysis* or *text content analysis* refer to the specific type of content analysis that focuses on written or transcribed words.

Historically, content analyses did begin with examinations of written text. And text analysis remains a vibrant part of content analysis research (Roberts, 1997b). The next chapter will trace this history and show how, over time, the applications of content analysis expanded beyond the written word.

Notes

1. According to Gunter (2000), the "overriding objective" of the positivism paradigm is to "prove or disprove hypotheses and ultimately to establish universal laws of behaviour through the use of numerically defined and quantifiable measures analogous to those used by the natural sciences" (p. 4).

2. According to Gregory (1987), "Freud's approach to the interpretation of dreams was by way of the method of free association [from which Freud's psychoanalysis procedures would evolve]. . . . As in psychoanalysis proper, the subject is required to relax and allow his mind to wander freely from elements in the dream to related ideas, recollections, or emotional reactions which they may chance to suggest" (p. 274). The dream as reported was termed by Freud the *manifest content,* and the dream's underlying thoughts and wishes Freud called the *latent content.*