
Content Analysis in Health Education Research: An Introduction to Purposes and Methods

David F. Duncan

Health education research is a difficult field for many reasons. Among the most troublesome of the problems we face in conducting health education research is the problem of "reactivity of measures." By this we mean that the way we measure things, or just the fact that we are measuring them, may cause a change in the very thing we are measuring. The way we phrase a question may bias the answer we receive. The fact that we are counting how many cigarettes a person smokes may cause that person to smoke less.

One way around this problem is to collect the data secretly. This can be done, for instance, by using such means as peepholes, hidden cameras, or one-way mirrors. Numerous other approaches have been developed for secret data collection. The impact of nutrition education on children's eating behavior has been studied by examining portions of their school lunches that students threw away. Attitudes toward controversial health issues have been estimated by studying the frequency of defacement of posters or bumper stickers. These and other such nonobstrusive measures have the advantage that the act of measurement does not have any effect on the behavior being measured.

Another approach is to collect data from "social artifacts"—products of people's social behavior. These include a great many things such as buildings, automobiles, gardens, or

musical compositions, and all of these may serve to tell us something about the people who made or bought them. Most useful are various documents such as books, newspapers, magazines, speeches, letters, poems, song lyrics, or jokes for which methods of content analysis have been developed. Content analysis methods also have been applied to such nonverbal social artifacts as paintings, photographs, or motion pictures.

Content analysis is a technique which lies at the crossroads of qualitative and quantitative methods. This set of techniques involves quantifying the frequency with which certain qualities appear in a sample of documents. It has been described by Holsti (1969) as the application of scientific methods to documentary evidence.

Long before content analysis was established as a research technique with a name of its own, its basic principles were being applied by scholars in disciplines as diverse as history, theology, and literature. Debates over whether or not Shakespeare really wrote all of the plays attributed to him and theories about the earlier documents on which the Biblical book of Genesis is based both are supported largely by evidence from what we would now call content analysis.

Modern content analysis was developed largely by Harold D. Lasswell and his associates (Lasswell, 1942; Lasswell, 1946; Lasswell, Leites and Associates, 1949) in their studies of the politically relevant content of newspapers in the late 1930s. Originally called "symbol

analysis," this technique recorded the frequency with which certain key words, such as "Germany" or "Jews" appeared in newspaper content, as well as whether their presentation was "indulgent," "deprivational," or "neutral." Favorable (indulgent) references could be coded further according to whether they stressed "strength" or "goodness," while negative (deprivational) references were coded into "weakness" and "immorality" categories.

During World War II, Lasswell's content analysis procedures were used as an intelligence gathering tool. Analysis of German newspapers obtained in neutral Switzerland permitted the Allies to assess German civilian morale and supply shortages. Even troop movements and combat casualties could be estimated to a degree using content analysis methods.

In one of the classic works on the subject, Berelson (1952) enumerated purposes for which content analysis had been conducted. Many, perhaps all, of the following purposes can be appropriate areas for research in health education:

- To study the characteristics of content, including:
 - To describe trends in communication content
 - To trace development of scholarship
 - To disclose international differences
 - To compare media or "levels" of communication
 - To audit communication content against objectives
 - To construct and apply commu-

David F. Duncan is chief epidemiologist, Illinois Association of Community Services, Chicago, IL.

- communications standards
- To aid in technical research operation
- To expose propaganda techniques
- To measure the "readability" of communications
- To discover stylistic features
- To study the producers or causes of content, including:
 - To identify intentions of the communication
 - To determine the psychological state of persons or groups
 - To detect existence of propaganda
 - To secure political or military intelligence
- To study the audience or effects of content, including:
 - To reflect the attitudes, interests, and values of groups
 - To reveal the focus of attention
 - To describe attitudinal and behavioral responses to content

Content analysis is a research technique more health educators should be familiar with. Fisher, Gandy, and Janus (1981, p. 257), argue that, "Content analysis of . . . popular media channels represents an important research activity for the critical scholar." This view is further supported by Kerlinger (1973, p. 533) who has expressed the view that, "Education has suffered from a lack of analysis of the educational information people absorb from the press and other media of public communication."

Applications in Health Education Research

There have been many applications of content analysis to health education research since Bobbit (1925) examined the health content of newspapers, magazines, and textbooks as a means of identifying objectives for health education.

Health educators often are interested in trends in media content. For instance, when the Attorney General's Commission on Pornography (1986) asserted that themes of sexual violence were increasingly frequent in pornography, they raised an issue amenable to content analysis. Unfortunately, the Commission's own

content analysis study did not report any evidence on this issue. Malamuth and Spinner (1980) and Duncan (1989) found that the Commission's assertion was false and that violent themes in pornography have not increased over recent decades. Scott and Cuvelier (1987) reached a similar conclusion from their 30 year study of the content of *Playboy* magazine.

MacDonald (1988) analyzed "latrinalia"—graffiti in public restrooms—for the frequency of sexual themes, AIDS references, and expressions of homophobia. Compared to previous latrinalia studies, MacDonald found an increase in the proportion of homophobic references since the appearance of the AIDS epidemic.

Albright, Altman, Slater and MacCoby (1988) examined the number of cigarette advertisements in each issue of eight popular magazines from 1960 and 1985. The average number of ads per issue in all eight magazines increased sharply following the 1971 ban on cigarette advertising on television. Beginning in the late 1960s, proportionately more cigarette ads appeared in those magazines targeted to the women's and youth markets, supporting the contention that cigarette companies were trying to recruit new smokers in these predominantly non-smoking markets.

In yet another trend study, Sofalvi (1985) recorded the frequency with which health-related topics appeared in selected Sunday comic strips over a period of four decades. A different application was Levy's (1987) use of content analyses of the official printed programs of the 1973 and 1983 annual meetings of the American Public Health Association (APHA) to assess women's leadership status in the APHA.

In one example of the use of content analysis for tracing developments in the field of health education, Elliott, Bearden, Byrd, Sanderson, Nichols, and Umlauf (1988) identified the authors, publications, and journals cited most frequently during a three year period in the journals *Behavioral Medicine* and *Health Psychology*. Taking these jour-

nals as exemplars of the two health education-related fields of health psychology and behavioral medicine, the citation analysis suggests different influences in research in the two fields. Health psychology seems to be shaped largely by the interface between experimental social and personality psychology and clinical psychology. Behavioral medicine, on the other hand, was shown to draw more on the literature of psychiatry. While health psychology tends to focus on the interpersonal level, behavioral medicine focuses more on the intrapsychic level.

Taylor (1985) assessed progress in sex education for the handicapped by a variety of methods. Included among those methods was a content analysis of the frequency with which the subject was addressed in human sexuality textbooks.

Petraki (1988) examined international differences by comparing the content of AIDS education materials developed in different nations. In another instance, Kalau (1980) examined the images of the elderly conveyed in children's books published in the United States and in West Germany.

Frazer (1982) examined research published in selected health education journals over a one year period. Of 448 articles published during the period, 240 (53.6 percent) reported original research. Of these, 114 (47.5 percent) were non-experimental studies, 121 (50.4 percent) were quasiexperimental, and only five (2 percent) were true experiments. Each of the 240 research articles was assessed using a measure of the quality of the research design reported. A major finding in this regard was that a majority of the reports failed to report adequately the methods used in the study. He compared the journals studied and found that each had its own content emphasis.

Although the readability of health education materials probably is tested far less often than it should be, there are numerous instances in the literature. A timely example is the study of condom instructions conducted by Richwald, Wamsley,

Coulson, and Morisky (1987). They applied three different readability formulas to 14 sets of instructions which accompany 25 brands of condoms produced by seven domestic and one foreign manufacturer. They found that eight of the 14 required a twelfth grade reading level to be understood. None of the instruction sets required less than a tenth grade reading level.

Breed and DeFoe (1981) examined alcohol-related content in comic books. Superheroes were the focus of the study because of their popularity, especially among 8-17 year olds, and their potential influence as role models for children's behavior. Superheroes were rarely depicted as drinking. "Primitive heroes," such as Conan the Barbarian, drank fairly frequently but were never depicted as being drunk. In the very few instances where any superhero was shown drunk, he or she later came to understand that drunkenness impaired crimefighting ability and personal health. Supervillians drank little more often than superheroes, but they drank "more unwisely."

There are many other examples of content analysis in health education research. Miller (1979), for instance, conducted a content analysis of the health-related articles appearing in the six leading women's magazines over a three year period. Schrink, Poole, & Regoli (1982) examined the prevalence of various sexual myths in rape jokes as a way of estimating their prevalence in the general culture. As a final example, content analysis as learning activity has been suggested by Redican, Redican and Baffi (1988) who propose that teachers have their students conduct a ten-variable content analysis of drug themes in motion pictures in order to stimulate class discussions.

Step-By-Step Procedure of Content Analysis

The procedure for content analysis can be described in terms of nine steps:

- (1) Define the units of analysis
- (2) Establish the coding categories
- (3) Pretest the coding on a subsample of the documents

(4) Assess the reliability of the coding

(5) Revise the coding categories if necessary and return to step 3

(6) Draw the sample of documents

(7) Code all the units of analysis

(8) Assess the achieved reliability of the coding

(9) Tabulate the categories and cross-tabulate where appropriate

Units of Analysis

The term units of analysis refers to those individual units about which descriptive or explanatory statements are to be made. In most health education research the units of analysis are people as individuals or in groups. For instance, if we study personality differences between smokers and non-smokers, our unit of analysis is the individual, whether a smoker or not. If we wish to study the differing divorce rates in single-income and dual-income families, our unit of analysis is the individual married couple. If our study concerned differences in child abuse rates between Black and White families, our unit of analysis would be families. If we wanted to compare drug abuse rates in inner city schools to those in suburban schools, our unit of analysis would be the individual school. In a comparison of AIDS rates in urban versus rural counties, our unit of analysis would be counties.

In content analysis the unit of analysis is a document or some element of mass communications. Units of analysis may be words, phrases, sentences, paragraphs, sections, chapters, articles, books, magazines or the entire body of work by a given author. In some cases the units of analysis may be the title of a document (e.g. Lebeque, 1985), the reference citations in it (e.g. Elliott et al., 1988), or the pictures which illustrate it (e.g. Dietz & Evans, 1982). Then each unit will be categorized for purposes of analysis.

Establishing Coding Categories

Establishing coding categories is one of the most critical steps in the planning of a content analysis

study. Categories must be designed "so that they answer the questions originally asked, be exhaustive (relative to the problem), and be mutually exclusive" (Budd, Thorp, & Donohew, 1967).

The researcher must be certain that categories established for the study truly will allow the original research questions to be answered. It is surprising how often a study design is adopted which fails to answer the original research question—a so-called Type V Error. Don't for instance, use "drinking" as a category if it is your intent to look for "alcoholism." Nor should you categorize pregnancies as "legitimate" or "illegitimate" if your research question concerns "planned" versus "unplanned" pregnancies.

It should be possible to classify each unit of analysis into one and only one category within each categorization. This should be achieved without excessive reliance upon an "other" category. In order to insure that this is so, the researcher should always pretest coding categories by coding a small subsample of the data in order to identify any ambiguities or missing categories (Budd, Thorp & Donohew, 1967).

While some studies may involve a single categorization, many will require multiple categorizations. Malamuth and Spinner (1980), for instance, categorized each of the erotic magazines in their study on the single variable of the degree of sexual violence portrayed in it. In my own studies of such magazines, on the other hand, I have categorized each magazine in terms of 11 different variables, ranging from page size to the sexual acts portrayed (Duncan, 1989). In some cases we may wish to code not only what subjects are addressed, but whether they are viewed positively or negatively and how strongly this view is expressed. Osgood (1959) presents a method for coding the direction and intensity of attitudes toward various topics.

Sampling

In the words of Kaplan and Goldsen (1965, p. 110), "Samples for

content analysis are chosen with two broad considerations in mind: (1) The sample must be representative of the universe from which it is drawn, and (2) it must be large enough to permit . . . reliable analysis." In order to achieve that representativeness, according to Beardsworth (1980), two elements are necessary: first, the universe must be explicitly defined; second, the sample must be random.

The universe of documents for a content analysis is analogous to the population in more ordinary statistical studies. The universe is the whole body of documents which the researcher wishes to generalize about. For instance, if a researcher wishes to examine the coverage of health issues in women's magazines, the researcher's first task will be to define what magazines are women's magazines. A list of the issues of such magazines which were published during the period under study serves as the sampling frame for the drawing of a sample. Issues (or smaller units) then might be drawn randomly from this sampling frame.

Since the number of magazines classifiable as women's magazines is probably quite large, the researcher may wish to limit the scope of the study to a specific set of magazines. This selection may be guided by the personal interests of the researcher or it may be systematically determined. One systematic approach to limiting the scope of the study is the popularity sample, as was used by Miller (1979) who studied the six women's magazines with the largest circulation. In a similar fashion, Frazer (1982), in his study of research studies published in health education related journals, selected the five journals he regarded as being most influential in the field of health education. Taking another approach in studying participation of women in leadership of the American Public Health Association, Levy (1987) selected for content analysis the printed programs of the association's 1973 and 1983 annual meetings as a follow-up to a 1973 study conducted by other methods.

Coding

The units of analysis are assigned to categories on each variable under study by assigning to each of them a numerical value identified with that category. It is important for the researcher to remember that these numbers represent only nominal, or at the most ordinal, level measurement. The researcher must not make the mistake of treating them as if they had arithmetic properties. If each unit of analysis has been coded according to the health domain it addresses, such that 1 = physical, 2 = mental, 3 = social, and 4 = spiritual, then a mean score calculated on this variable would be meaningless.

The process of coding each of the units of analysis into one category of each variable under study is a time consuming and exacting one. It is crucial to any worthwhile analysis that this process be carried out as accurately as possible. Coders must be carefully trained in their task, and their performance should be subjected to periodic reliability checks. An excellent brief discussion of reliability testing for content analysis can be found in Olson (1976, pp. 172-177).

- Albright, C. L., Altman, D. G., Slater, M. D. & Maccoby, N. (1988). Cigarette advertisements in magazines: Evidence for a differential focus on women's and youth magazines. *Health Education Quarterly*, 15(2), 225-233.
- Attorney General's Commission on Pornography. (1986). *Final report*. Washington, DC: Government Printing Office.
- Beardsworth, A. (1980). Analysing press content: Some technical and methodological issues. In H. Christian (Ed.), *The sociology of journalism and the press* (pp. 371-395). Hanley, England: J. H. Brookes.
- Berelson, B. (1952). *Content analysis in communication research*. Glencoe, IL: Free Press.
- Bobbitt, F. (1925). Discovering the objectives of health education: A study of health topics in newspapers, magazines and textbooks. *Elementary School Journal*, 25, 756-760.
- Breed, L. A. & DeFoe, J. R. (1981). Comic book supercharacters and their role in alcohol education. *Journal of Alcohol and Drug Education*, 27(1), 1-13.
- Budd, R. W., Thorp, R. K. & Donohew, L. (1967). *Content analysis of communications*. New York: Macmillan.
- Celline, H. B. & Duncan, D. F. (1988). Homosexual pornography: Trends in con-

- tent and form over a twenty-five year period. *Psychology*, 25(4), 1-5.
- Dietz, P. E. & Evans, B. (1982). Pornographic imagery and prevalence of paraphilia. *American Journal of Psychiatry*, 139(11), 1493-1495.
- Duncan, D. F. (1989). Trends in gay pornographic magazines: 1960-1984. *Sociology and Social Research*, 7(1), 1-4.
- Elliott, T. R., Bearden, L., Byrd, E. K., Sanderson, R. A., Nichols, R. K. & Umlauf, R. (1988). Influential research in Behavioral Medicine and Health Psychology: A study of citation behaviors. *Psychology*, 25(2), 23-31.
- Fisher, J., Gandy, O. H. & Janus, N. Z. (1981). The role of popular media in defining sickness and health. In E. G. McAnany, J. Schnitman, and N. Janus (Eds.), *Communication and social structure* (pp. 240-257). New York: Praeger.
- Frazer, G. H. (1982). An analysis of research published in health education related journals, May, 1980-May, 1981. Unpublished M. S. thesis, Southern Illinois University at Carbondale.
- Goetz, J. P. & LeComte, M. D. (1981). Ethnographic research and the problem of data reduction. *Anthropology and Education Quarterly*, 12, 51-70.
- Holsti, O. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.
- Kalau, E. I. (1980). Experience of old age as reflected in U.S. and West German children's literature. Unpublished Ph.D. dissertation, Southern Illinois University at Carbondale.
- Kaplan, A. & Goldsen, J. M. (1965). The reliability of content analysis categories. In H. D. Lasswell, N. Leites, and Associates (Eds.), *Language of politics, studies in quantitative semantics* (pp. 83-112). Cambridge: Massachusetts Institute of Technology.
- Kerlinger, F. (1973). *Foundations of behavioral research* (2nd ed.). New York: Holt, Rinehart & Winston.
- Lasswell, H. D. (1942). The politically significant content of the press: Coding procedures. *Journalism Quarterly*, 19, 12-23.
- Lasswell, H. D. (1946). Describing the contents of communications. In B. L. Smith, H. D. Lasswell & R. D. Casey (Eds.), *Propaganda, communication and public opinion*. Princeton, NJ: Princeton University Press.
- Lasswell, H. D., Leites, N., and Associates (1949). *Language of politics*. New York: George W. Stewart.
- Lebeque, B. J. (1985). Paraphilias in pornography: A study of perversions inherent in title. *Australian Journal of Sex, Marriage and the Family*, 6(1), 33-36.
- Levy, J. A. (1987). Women's leadership status in the American Public Health Association. *American Journal of Public Health*, 77, 1537-1538.
- MacDonald, B. S. (1988). Restroom graffiti as an indicator of gender differences at a large mid-western university. Unpublished M. S. thesis, Southern Illinois University at Carbondale.

