Quantitative content analysis

Contents

1 Introduction

This supplement expands the introduction to quantitative content analysis in Chapter 6. Along with additional details of how to carry out a study using content analysis it offers an overview of the design's strengths and weaknesses, along with additional references for further study.

Content analysis refers to a family of procedures for the systematic, replicable analysis of text. In essence it involves the classification of parts of a text through the application of a structured, systematic coding scheme from which conclusions can be drawn about the message content. By clearly specifying the coding and other procedures content analysis is replicable in the sense that other researchers could reproduce the study. Content analysis can be carried out quantitatively but also qualitatively. In this technical supplement we focus on quantitative content analysis; qualitative content analysis is discussed in the technical supplement on qualitative approaches to language analysis.

Content analysis can be applied to all kinds of written text such as speeches, letters or articles whether digital or in print, as well as text in the form of pictures, video, film or other visual media. It can be used to examine both the manifest and the latent content of a text. Manifest content refers to the visible, countable components of the message. Latent content refers to the meaning that may lie behind the manifest content. Manifest content in the form of images of women in advertisements engaged in activities such as domestic work or child care, for example, might be taken as indicators of gender stereotyping, a latent concept. Both manifest

and latent content still require interpretation but the interpretations vary in depth and level of abstraction (Graneheim and Lundman 2004).

Content analysis can also be used to investigate both substantive (or content) and form (or formal) features of a text. Substantive features refer to what is being said in the message. Form features refer to how it is being said (Schreier 2012). A content analysis of advertisements, for instance, might distinguish between the way in which the benefits of a product are described (substantive features) and the way in which the advertisement makes use of headings, different font sizes, the placement of graphics and so on (form features).

2 Applications of content analysis

Content analysis provides a structured way of analysing data that are typically open-ended and relatively unstructured. Two important aims of such analysis for business and management research are:

- Description. Here the focus is on describing features of the message content. Jain et al. (2010), for example, study how the ways in which celebrities were presented in Indian television commercials varied according to the category of products they were promoting. Descriptive content analysis can be cross-sectional or longitudinal. An example of the latter is Paek et al.'s (2012) investigation of the promotion techniques used in cigarette advertising over a fifty-year period.
- Prediction. Here the main aim is to predict the outcome or effect of the messages being analysed. Through the measurement of relevant features of the message, the researcher seeks to predict audience or receiver reaction (Neuendorf 2002). A study by Naccarato and Neuendorf (1998), for example, investigates how different features of print media advertising affected recall, readership and evaluation in a business-to-business context. In that particular study, content analysis of the advertisements is combined with survey data to allow the researchers to assess the effect of particular features of the advertising.

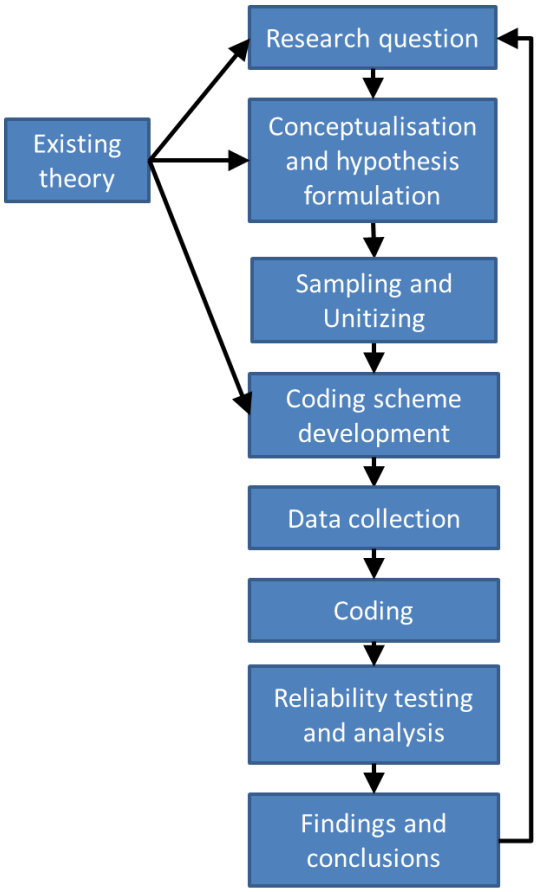
Content analysis has been applied to a wide range of social science topics including gender and race, violence, media reporting and political communication. Insch et al. (1997) suggest its use is less common in organisational research due to lack of familiarity with the method. Nevertheless it has obvious applications in the analysis of business communication, particularly marketing. The rise of the Internet creates opportunities for using content

analysis techniques to analyse online communication. Examples include Jose and Lee's (2007) study of global corporations' environmental reporting based on examination of their websites and Gebauer et al.'s (2008) analysis of online user reviews to identify requirements for mobile technology.

3 Outline of the design

Quantitative content analysis shares many of the general characteristics of quantitative research discussed in Chapter 4. Nevertheless, it does have some distinctive features, particularly with respect to the application of a coding scheme and the way in which reliability is assessed, which justify treating it as a research design in its own right and not just as a method of analysis. Figure 1 outlines the key stages in a content analysis project.

Figure 1 – Key steps in quantitative content analysis



As with other quantitative approaches we have looked at, design begins with the identification of relevant concepts and, where appropriate, the formulation of hypotheses in response to the research question. This is where you establish the rationale behind your proposed project and avoid the risk of engaging in pointless 'word crunching' rather than

purposeful analysis (Insch et al. 1997). Existing theory and prior research play an important role in helping you develop the conceptual basis for your research. This is a very important stage of the research as the concepts will form the basis of your coding scheme and thus of your final analysis.

Sampling involves identifying and selecting the material that you intend to analyse. Suppose, for instance, that your interest was in magazine advertisements. The sampling process would require decisions about what type of advertisements are of interest, in which magazines they appeared, over what time frame, and so on. You would also need to choose between including all eligible advertisements or selecting a sample. At this point the basic principles of sampling discussed in Chapter 9 apply. Collection involves obtaining the material in an analysable format, a topic discussed further in Chapter 12.

In parallel with developing a sampling plan, you will need to decide on the unit of text that you will classify during the coding process. We will refer to this as the **coding unit**, although terminology varies. Examples of coding units include words, phrases, sentences, images, paragraphs or whole documents. You should choose your coding unit taking into account your research question and the concepts that you wish to identify in your analysis. In the case of Jain et al.'s (2010) study of celebrities in TV commercials, for example, individual advertisements served as the unit of coding.

The next major step is to develop a coding scheme. This is the process of developing classification rules to assign coding units to particular categories or concepts, for example, assigning the numerical code '0' to an advertisement if the central figure is an image of a male and '1' if it contains a female. The resulting rules are detailed in a code book (or coding manual) which specifies how and what to code, an example of which is shown in Research in practice 1. The code book helps to ensure systematic and replicable coding of the data. In preparing your code book you should make sure that each category is exhaustive and mutually exclusive and that the instructions for coding are clear. Your coding scheme may draw on existing ones developed by other researchers. In addition there are a number of existing content analysis dictionaries which are available to support the analysis of written text. These dictionaries specify a range of concepts and the words or phrases that are indicators of those concepts. Alternatively, your coding scheme can be developed inductively from the data using techniques similar to those used when analysing qualitative data as discussed in Chapter 14. In addition to a code book, you will also need to prepare a coding

form (or coding schedule). This is a form which will be used to record details of the codes applied to the data during the coding process. That information can then be transferred from the coding form into a software program for further analysis.

Research in practice 1 – Example code book (adapted from Zhang et al. 2009)

Example code book from a study comparing gender role portrayals in magazine advertising

1. Gender of the central figure (male 0; female 1)
2. Age of the central figure (young adult (18–34 years of age) 1; mid adult (35–50 years of age) 2; older adult (over 50 years of age) 3; indeterminate 4
3. Occupational/non-occupational role of the central figure (occupational 0; non-occupational 1; indeterminate 2)
4. Subdivision of occupational role of the central figure (high-level executive 1; professional 2; entertainer/media 3; professional sports 4; mid-level business/semi-professional 5; non-professional/white collar 6; non-professional blue-collar 7; indeterminate 8)
5. Subdivision of non-occupational role of the central figure (family 1; recreational 2; decorative 3)
6. Credibility of the central figures (product user or presenter 1; product authority 2; decorative 3; indeterminate 4)

Once your initial coding scheme has been developed it needs to be piloted. This can be done on a randomly selected sample of the data. Piloting is essential to identify problems with the coding scheme or the coders' ability to apply it. Any such problems need to be addressed before the study proceeds (Neuendorf 2002).

When the coding scheme is finalised, coding can begin. Whilst coding can be done by a single person the use of multiple coders allows the principal researcher to see whether the coding scheme can be applied in a reliable way. In addition, if there is a large amount of data, more than one coder is likely to be needed. An alternative to human coding is to use a computer and a number of specialised computer software programs exist to support quantitative content analysis, particularly of written text. Their use has been made easier thanks to the widespread availability of textual data in digital format. More information can be found in the references in the further reading section of Chapter 6.

Final analysis involves the application of quantitative techniques as discussed in Chapter 13. Descriptive statistics, such as frequency counts, can be used to summarise findings from the sample and appropriate inferential statistics used to test any hypotheses that have been formulated. An important additional step for content analysis, however, is reliability testing.

Particularly relevant if more than one coder is used is the examination of the consistency between coders. Unless the coding scheme has been applied in a consistent way the resulting data will be unreliable. This can be tested by getting coders to code the same set of material and then measuring inter-rater reliability. One such measure is percentage agreement which is given by the following formula:

$$PA = A/n \times 100$$

where *PA* = percentage agreement, *A* = number of agreements and *n* = number of segments coded. Unfortunately, there is no clear agreement on what constitutes an acceptable level of reliability. Neuendorf (2002: 143), writing generally about inter-coder reliability, suggests that scores above 80 per cent would be acceptable 'in most situations' but with more disagreement for lower values. More sophisticated measures include Cohen's kappa and Scott's pi, details of which can be found in references in the further reading section in Chapter 6. You should include the results of reliability analysis in your final write up along with your findings and conclusions.

Research in practice 1 gives an example of a quantitative content analysis study of the portrayal of women in James Bond films.

Research in practice 2 – Example content analysis

Shaken and stirred: portrayals of women in James Bond films

How women are portrayed in the media has long been a subject of interest to researchers. Neuendorf et al.'s (2010) study uses quantitative content analysis to investigate how women are depicted in the long-running and very popular series of James Bond films. Three research questions are linked to how the portrayal of women has changed over time with respect to 1) their physical characteristics 2) their level of sexual activity 3) the amount and level of violence directed against them. Three further questions analyse whether the way that a female character is portrayed (in terms, for instance, of her physical characteristics) predicts 1) the amount of sexual activity in which she is involved 2) the amount of aggression directed at her 3) whether the character survives the film. The coding unit is female characters who meet certain criteria (e.g. they speak or are spoken to). The code book defines a range of variables to be measured for each character, such as race, hair colour, body size, physical attractiveness, along with variables measuring aspects of sexual activity and violence by and against the character. Eight coders were used to code the 20 films. Inter-coder reliability was tested using one of the twenty films and the results are reported in the study. Research findings showed some minor changes over time in terms of more female roles, more sexual activity and greater likelihood of being

recipients of physical violence but found much consistency in terms of age, body type and attractiveness. The findings on prediction showed clear links between sex and violence in the way women are portrayed in Bond films. According to the authors 'the collective body of Bond films… stands to serve as an important source of social cognitive outcomes regarding appropriate role behaviour for women – still stereotyped, with persistent allusions to violence and sex (and their linkage), and with unrealistic standards of female beauty' (Neuendorf et al. 2010: 758–9)

4 Strengths and weaknesses of content analysis

Content analysis is a flexible research approach that can be applied to a wide variety of text sources. Helped by the availability of computer software programs, content analysis can cope with large amounts of data. It can be used to investigate a topic longitudinally through the examination of contemporary texts. Content analysis can also be seen as an unobtrusive research approach in that it can be used to analyse naturally-occurring data, as discussed in Chapter 12. As a result, content analysis may be helpful in reducing the problem of social desirability bias amongst respondents when researching sensitive topics (Insch et al. 1997, Harris 2001).

Potential weaknesses of the design arise in connection with the process of sampling and coding. Document availability and the sampling process can introduce bias. Developing the coding scheme and coding always involve interpretation, even of manifest content, and thus risk similar biases to those faced by other measurement techniques (Insch et al. 1997). Abstraction of content from its context can also create problems. Taking a word or phrase in isolation of other parts of the text, for instance, may result in loss of meaning. In addition, content analysis risks overlooking what is not said in a particular text. In some situations what is omitted may be as significant as what is included.

Content analysis can struggle to provide explanations for particular findings where the source of the explanation lies outside of the text itself. Consider, for example, Todd et al.'s (1995) study of skills in information system job advertisements between 1970 and 1990. The authors report a modest increase in the demand for technical skills and little difference between industries. These 'somewhat unexpected' findings (Todd et al. 1995: 16) invite explanation and the authors offer six alternatives to explain their results but it is not possible to substantiate these from the study of the advertisements themselves. Todd et al.'s (1995) study also raises a broader question about the extent to which documentary analysis provides access to the world beyond the text. Do job advertisements, for instance, tell us something about real

changes in job skill or do they tell us only about the way in which skill is portrayed in advertising? This is not a criticism of content analysis because investigating changes in how subjects are portrayed can be valuable of itself but it should caution us against any simplistic assumption of a direct correspondence between text and the external world. Researchers interested in, for instance, the impact of advertisements on consumers can choose to combine CA with other research approaches such as surveys or experiments.

5 References

Gebauer, J., Tang, Y. and Baimai, C. (2008). 'User requirements of mobile technology: Results from a content analysis of user reviews', *Information Systems & e-Business Management*, 6(4), 361–384.

Graneheim, U. H. and Lundman, B. (2004). 'Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness', *Nurse education today*, 24(2), 105–112.

Harris, H. (2001). 'Content analysis of secondary data: A study of courage in managerial decision making', *Journal of Business Ethics*, 34(3/4), 191–208.

Insch, G. S., Moore, J. E. and Murphy, L. D. (1997). 'Content analysis in leadership research: Examples, procedures, and suggestions for future use', *The Leadership Quarterly*, 8(1), 1–25.

Jain, V., Roy, S., Daswani, A. and Sudha, M. (2010). 'How celebrities are used in Indian television commercials', *Vikalpa: The Journal for Decision Makers*, 35(4), 45–52.

Jose, A. and Shang-Mei, L. (2007). 'Environmental reporting of global corporations: A content analysis based on website disclosures', *Journal of Business Ethics*, 72(4), 307–321.

Naccarato, J. L. and Neuendorf, K. A. (1998). 'Content analysis as a predictive methodology: Recall, readership, and evaluations of business-to-business print advertising', *Journal of Advertising Research*, 38(3), 19–33.

Neuendorf, K. A. (2002). *The content analysis guidebook.* Thousand Oaks, CA: Sage.

Neuendorf, K. A., Gore, T. D., Dalessandro, A., Janstova, P. and Snyder-Suhy, S. (2010). 'Shaken and stirred: A content analysis of women's portrayals in James Bond films', *Sex Roles*, 62(11–12), 747–761.

Paek, H.-J., Reid, L. N., Jeong, H. J., Choi, H. and Krugman, D. (2012). 'Five decades of promotion techniques in cigarette advertising: A longitudinal content analysis', *Health Marketing Quarterly*, 29(1), 1–17.

Schreier, M. (2012). *Qualitative content analysis in practice.* London: Sage.

Todd, P. A., McKeen, J. D. and Gallupe, R. B. (1995). 'The evolution of IS job skills: A content analysis of IS job advertisements from 1970 to 1990', *MIS Quarterly*, 19(1), 1–27.

Zhang, L., Srisupandit, P. T. and Cartwright, D. (2009). 'A comparison of gender role portrayals in magazine advertising: The United States, China and Thailand', *Management Research News*, 32(7), 683–700.