of validity evidence. Construct validation provides an integrative framework that ties together all forms of validity evidence in a way continuous with empirical research into the construct, but some have suggested a less expansive view of validity as more practical. Construct validity evidence based on test consequences remains a continuing point of controversy, particularly with respect to the notion of *consequential validity* as a distinct form of validity. Finally, there remains a fundamental tension in modern validity theory between the traditional fact–value dichotomy and the fundamental role of values and evaluation in assessing the evidence in favor of specific tests, scores, interpretations, and uses.

Keith A. Markus and Chia-ying Lin

See also Content Validity; Criterion Validity; Structural Equation Modeling

Further Readings

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education and Praeger.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education and Macmillan.

Wainer, H., & Braun, H. I. (1988). *Test validity.* Mahwah, NJ: Lawrence Erlbaum.

Content Analysis

Content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the context of their use.

This entry further explores the definition and conceptions of content analysis. It then provides information on its conceptual framework and the steps involved in performing content analysis.

Definition and Conception

The phrase *content analysis,* first mentioned in a 1941 paper by Douglas Waples and Bernard Berelson, became defined in 1948 by Paul F. Lazarsfeld and Berelson. *Webster's Dictionary* has listed *content analysis* since its 1961 edition. However, the practice of analyzing media matter is almost as old as writing. It became of interest to the church, worried about the effects of the written word other than God's; to governments, trying to settle political, legal, and religious disputes; to journalists, hoping to document the changes in newspaper publishing due to its commercialization and popularization; to corporations interested in surveying their symbolic environments for opportunities and threats; and to social scientists, originally drawn into the competition between the press and newly emerging media, then radio and television, but soon discovering the importance of all kinds of mediated communication to understand social, political, economic, and psychological phenomena. Communication research advanced content analysis, but owing to the proliferation of media and the recognition that humans define themselves and each other in communication, coordinate their beliefs and actions in communication, and construct the realities they live with in communication, content analysis is now used by literally all social sciences.

As a *technique,* content analysis embraces specialized procedures. It is teachable. Its use can be divorced from the authority of the researcher. As a *research* technique, content analysis can provide new kinds of understanding social phenomena or inform decisions on pertinent actions. Content analysis is a scientific tool.

All techniques are expected to be reliable. Scientific research techniques should result in replicable findings. Replicability requires research procedures to be explicit and communicable so that researchers, working at different times and perhaps under different circumstances, can apply them and come to the same conclusions about the same phenomena.

Scientific research must also yield valid results. To establish validity, research results must survive in the face of independently available evidence of what they claim. The methodological requirements of reliability and validity are not unique to content

analysis but make particular demands on the technique that are not found as problematic in other methods of inquiry.

The reference to *text* is not intended to restrict content analysis to written material. The parenthetical phrase "or other meaningful matter" is to imply content analysis's applicability to anything humanly significant: images, works of art, maps, signs, symbols, postage stamps, songs, and music, whether mass produced, created in conversations, or private. Texts, whether composed by individual authors or produced by social institutions, are always intended to point their users to something beyond their physicality. However, content analysis does not presume that readers read a text as intended by its source; in fact, authors may be quite irrelevant, often unknown. In content analysis, available texts are analyzed to answer research questions not necessarily shared by everyone.

What distinguishes content analysis from most observational methods in the social sciences is that *the answers to its research questions are inferred from available text.* Content analysts are not interested in the physicality of texts that can be observed, measured, and objectively described. The alphabetical characters of written matter, the pixels of digital images, and the sounds one can manipulate at a control panel are mere vehicles of communication. What text means to somebody, what it represents, highlights and excludes, encourages or deters—all these phenomena do not reside inside a text but come to light in processes of someone's reading, interpreting, analyzing, concluding, and in the case of content analysis, answering pertinent research questions concerning the text's context of use.

Typical research questions that content analysts might answer are, What are the consequences for heavy and light viewers of exposure to violent television shows? What are the attitudes of a writer on issues not mentioned? Who is the author of an anonymously written work? Is a suicide note real, requiring intervention, or an empty threat? Which of two textbooks is more readable by sixth graders? What is the likely diagnosis for a psychiatric patient, known through an interview or the responses to a Rorschach test? What is the ethnic, gender, or ideological bias of a newspaper? Which economic theory underlies the reporting of business news in the national press? What is the

likelihood of cross-border hostilities as a function of how one country's national press portrays its neighbor? What are a city's problems as inferred from citizens' letters to its mayor? What do school children learn about their nation's history through textbooks? What criteria do Internet users employ to authenticate electronic documents?

Other Conceptions

Unlike content analysis, observation and measurement go directly to the phenomenon of analytic interest. Temperature and population statistics describe tangible phenomena. Experiments with human participants tend to define the range of responses in directly analyzable form, just as structured interviews delineate the interviewees' multiple choices among answers to prepared interview questions. Structured interviews and experiments with participants acknowledge subjects' responses to meanings but bypass them by standardization. Content analysts struggle with unstructured meanings.

Social scientific literature does contain conceptions of content analysis that mimic observational methods, such as those of George A. Miller, who characterizes content analysis as a method for putting large numbers of units of verbal matter into analyzable categories. A definition of this kind provides no place for methodological standards. Berelson's widely cited definition fares not much better. For him, "content analysis is a research technique for the objective, systematic and quantitative description of the manifest content of communication" (p. 18). The restriction to manifest content would rule out content analyses of psychotherapeutic matter or of diplomatic exchanges, both of which tend to rely on subtle clues to needed inferences. The requirement of quantification, associated with objectivity, has been challenged, especially because the reading of text is qualitative to start and interpretive research favors qualitative procedures without being unscientific. Taking the questionable attributes out of Berelson's definition reduces content analysis to the systematic analysis of content, which relies on a metaphor of content that locates the object of analysis inside the text—a conception that some researchers believe is not only misleading but also prevents the formulation of sound methodology.
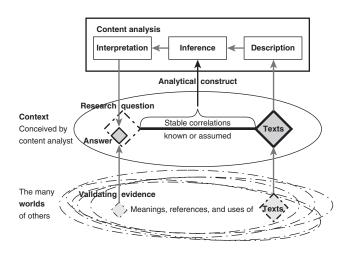
Figure 1   A Framework for Content Analysis

There are definitions, such as Charles Osgood's or Ole Holsti's, that admit inferences but restrict them to the source or destination of the analyzed messages. These definitions provide for the use of validity criteria by allowing independent evidence to be brought to bear on content analysis results, but they limit the analysis to causal inferences.

Conceptual Framework

Figure 1 depicts the methodologically relevant elements of content analysis. A content analysis usually starts with either or both (a) available text that, on careful reading, poses scientific research questions or (b) research questions that lead the researcher to search for texts that could answer them.

Research Questions

In content analysis, research questions need to go outside the physicality of text into the world of others. The main motivation for using content analysis is that the answers sought cannot be found by direct observation, be it because the phenomena of interest are historical, hence past; enshrined in the mind of important people, not available for interviews; concern policies that are deliberately hidden, as by wartime enemies; or concern the anticipated effects of available communications, hence not yet present. If these

phenomena were observable directly, content analysis of texts would be redundant. Content analysts pursue questions that could conceivably be answered by examining texts but that could also be validated by other means, at least in principle.

The latter rules out questions that have to do with a researcher's skill in processing text, albeit systematically. For example, the question of how much violence is featured on television is answerable by counting incidences of it. For content analysts, an index of violence on television needs to have empirical validity in the sense that it needs to say something about how audiences of violent shows react, how their conception of the world is shaped by being exposed to television violence, or whether it encourages or discourages engaging in violent acts. Inferences about antecedents and consequences can be validated, at least in principle. Counts can be validated only by recounting. Without designating where validating evidence could be found, statements about the physicality of text would not answer the research questions that define a content analysis.

Research questions must admit alternative answers. They are similar to a set of hypotheses to be tested, except that inferences from text determine choices among them.

Context of the Analysis

All texts can be read in multiple ways and provide diverging information to readers with diverging competencies and interests. Content analysts are not different in this regard, except for their mastery of analytical techniques. To keep the range of possible inferences manageable, content analysts need to construct a context in which their research questions can be related to available texts in ways that are transparent and available for examination by fellow scientists. This restriction is quite natural. Psychologists construct their world unlike sociologists do, and what is relevant when policy recommendation or evidence in court needs to be provided may have little to do with an analysis that aims at deciding when different parts of the Bible were written. The context of a content analysis always is the analyst's choice. There are no restrictions except for having to be explicit and arguably related to the world of others for whom the analyzed text means something, refers to

something, and is useful or effective, though not necessarily as content analysts conceptualize these things.

Description of Text

Usually, the first step in a content analysis is a description of the text. Mary Bock called content analyses that stop there "impressionistic" because they leave open what a description could mean. Three types of description may be distinguished: (1) selected word counts, (2) categorizations by common dictionaries or thesauri, and (3) recording or scaling by human coders.

Selected Word Counts

Selected word counts can easily be obtained mechanically and afford numerous comparisons by sources or situations or over time. For example, the 12 most frequent words uttered by Paris Hilton in an interview with Larry King were 285 *I,* 66 *you,* 61 *my,* 48 *like,* 45 *yes,* 44 *really,* 40 *me,* 33 *I'm,* 32 *people,* 28 *they,* 17 *life* and *time,* and 16 *jail.* That *I* is by far the most frequent word may suggest that the interviewee talked largely about herself and her own life, which incidentally included a brief visit in jail. Such a distribution of words is interesting not only because normally one does not think about words when listening to conversations but also because its skewedness is quite unusual and invites explanations. But whether Hilton is self-centered, whether her response was due to Larry King's questioning, how this interview differed from others he conducted, and what the interview actually revealed to the television audience remain speculation. Nevertheless, frequencies offer an alternative to merely listening or observing.

Many computer aids to content analysis start with words, usually omitting function words, such as articles, stemming them by removing grammatical endings, or focusing on words of particular interest. In that process, the textual environments of words are abandoned or, in the case of *keywords in context* lists, significantly reduced.

Categorizing by Common Dictionaries or Thesauri

Categorization by common dictionaries or thesauri is based on the assumptions that (a) textual meanings reside in words, not in syntax and organization; (b) meanings are shared by everyone—"manifest," in Berelson's definition—as implied in the use of published dictionaries and thesauri; and (c) certain differentiations among word meanings can be omitted in favor of the gist of semantic word classes. Tagging texts is standard in several computer aids for content analysis. The General Inquirer software, for example, assigns the words *I, me, mine,* and *myself* to the tag "self" and the tags "self," "selves," and "others" to the second-order tag "person." Where words are ambiguous, such as *play,* the General Inquirer looks for disambiguating words in the ambiguous word's environment—looking, in the case of *play,* for example, for words relating to children and toys, musical instruments, theatrical performances, or work—and thereby achieves a less ambiguous tagging.

Tagging is also used to scale favorable or unfavorable attributes or assign positive and negative signs to references.

Recording or Scaling by Human Coders

Recording or scaling by human coders is the traditional and by far the most common path taken to obtain analyzable descriptions of text. The demand for content analysis to be reliable is met by standard coding instructions, which all coders are asked to apply uniformly to all units of analysis. Units may be words, propositions, paragraphs, news items, or whole publications of printed matter; scenes, actors, episodes, or whole movies in the visual domain; or utterances, turns taken, themes discussed, or decisions made in conversations.

The use of standard coding instructions offers content analysts not only the possibility of analyzing larger volumes of text and employing many coders but also a choice between emic and etic descriptions—*emic* by relying on the very categories that a designated group of readers would use to describe the textual matter, *etic* by deriving coding categories from the theories of the context that the content analysts have adopted. The latter choice enables content analysts to describe latent contents and approach phenomena that ordinary writers and readers may not be aware of. "Good" and "bad" are categories nearly everyone understands alike, but "prosocial" and "antisocial"

attitudes, the concept of framing, or the idea of a numerical strength of word associations needs to be carefully defined, exemplified, and tested for reliability.

Inference

Abduction

Although sampling considerations are important in selecting texts for analysis, the type of inference that distinguishes content analysis from observational methods is abduction—not induction or deduction. Abduction proceeds from particulars—texts—to essentially different particulars—the answers to research questions. For example, inferring the identity of the author from textual qualities of an unsigned work; inferring levels of anxiety from speech disturbances; inferring a source's conceptualization from the proximity of words it uses; inferring Stalin's successor from public speeches by Politburo members at the occasion of Stalin's birthday; or inferring possible solutions to a conflict entailed by the metaphors used in characterizing that conflict.

Analytical Constructs

Inferences of this kind require some evidential support that should stem from the known, assumed, theorized, or experimentally confirmed stable correlations between the textuality as described and the set of answers to the research question under investigation. Usually, this evidential support needs to be operationalized into a form applicable to the descriptions of available texts and interpretable as answers to the research questions. Such operationalizations can take numerous forms. By intuition, one may equate a measure of the space devoted to a topic with the importance a source attributes to it. The relation between different speech disturbances and the diagnosis of certain psychopathologies may be established by correlation. The relation between the proximity of words and associations, having been experimentally confirmed, may be operation-alized in clustering algorithms that compute word clusters from strings of words.

While the evidential support for the intended inferences can come from anywhere, content analysts cannot bypass justifying this step. It would be methodologically inadmissible to claim to have analyzed "the" content of a certain news channel, as if no inference were made or as if content were contained in its transmissions, alike for everyone, including content analysts. It is equally inadmissible to conclude from applying a standard coding instrument and a sound statistics on reliably coded data, that the results of a content analysis say anything about the many worlds of others. They may represent nothing other than the content analyst's systematized conceptions.

Regarding the analytical construct, content analysts face two tasks, preparatory and applied. Before designing a content analysis, researchers may need to test or explore available evidence, including theories of the stable relations on grounds of which the use of analytical constructs can be justified. After processing the textual data, the inferences tendered will require similar justifications.

Interpretation

The result of an inference needs to be interpreted so as to select among the possible answers to the given research question. In identifying the author of an unsigned document, one may have to translate similarities between signed and unsigned documents into probabilities associated with conceivable authors. In predicting the use of a weapon system from enemy domestic propaganda, one may have to extrapolate the fluctuations of mentioning it into a set of dates. In ascertaining gender biases in educational material, one may have to transform the frequencies of gender references and their evaluation into weights of one gender over another.

Interpreting inferences in order to select among alternative answers to a research question can be quite rigorous. Merely testing hypotheses on the descriptive accounts of available texts stays within the impressionistic nature of these descriptions and has little to do with content analysis.

Criteria for Judging Results

There are essentially three conditions for judging the acceptability of content analysis results. In the absence of direct validating evidence for the

inferences that content analysts make, there remain reliability and plausibility.

Reliability

Reliability is the ability of the research process to be replicated elsewhere—that is, other researchers' agreeing with the findings of a content analysis or adding data to them. Traditionally, the most unreliable part of a content analysis is the recording, categorization, or scaling of text by human coders, and content analysts employing coders for this purpose are required to assess the reliability of that process quantitatively. Measures of reliability are provided by agreement coefficients with suitable reliability interpretations, such as Scott's π (pi) and Krippendorff's α (alpha). The literature contains recommendations regarding the minimum agreement required for an analytical process to be sufficiently reliable. However, that minimum should be derived from the consequences of answering the research question incorrectly. Some disagreements among coders may not make a difference, but others could direct the process to a different result.

Plausibility

Computer content analysts pride themselves in having bypassed reliability problems. However, all content analysts need to establish the plausibility of the path taken from texts to their results. This presupposes explicitness as to the analytical steps taken. The inability to examine critically the steps by which a content analysis proceeded to its conclusion introduces doubts in whether the analysis can be trusted, and implausibility can fail the effort. Content analysts cannot hide behind obscure algorithms whose inferences are unclear. Plausibility may not be quantifiable, as reliability is, but it is one criterion all content analyses must satisfy.

Validity

In content analysis, validity may be demonstrated variously. The preferred validity is *predictive,* matching the answers to the research question with subsequently obtained facts. When direct and post facto validation is not possible,

content analysts may need to rely on indirect evidence. For example, when inferring the psychopathology of a historical figure, accounts by the person's contemporaries, actions on record, or comparisons with today's norms may be used to triangulate the inferences. Similarly, when military intentions are inferred from the domestic broadcasts of wartime enemies, such intentions may be correlated with observable consequences or remain on record, allowing validation at a later time. *Correlative* validity is demonstrated when the results of a content analysis correlate with other variables. *Structural* validity refers to the degree to which the analytical construct employed does adequately model the stable relations underlying the inferences, and *functional* validity refers to the history of the analytical construct's successes. *Semantic* validity concerns the validity of the description of textual matter relative to a designated group of readers, and *sampling* validity concerns the representativeness of the sampled text. Unlike in observational research, texts need to be sampled in view of their ability to provide the answers to research questions, not necessarily to represent the typical content produced by their authors.

Klaus Krippendorff

See also Hypothesis; Interrater Reliability; Krippendorff's Alpha; Reliability; Validity of Research Conclusions

Further Readings

Berelson, B. (1952). *Content analysis in communications research*. New York: Free Press.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Krippendorff, K., & Bock, M. A. (2008). *The content analysis reader*. Thousand Oaks, CA: Sage.

CONTENT VALIDITY

Content validity refers to the extent to which the items on a test are fairly representative of the entire domain the test seeks to measure. This entry discusses origins and definitions of content validation, methods of content validation, the role of