CHAPTER 7

CONTENT ANALYSIS

Content analysis involves the collection and analysis of any type of textual, visual, or audiovisual material: TV shows, movies, commercials, song lyrics, music videos, speeches, letters, e-mails, tweets, books, posters, magazine articles, Web pages, workplace policies, medical or legal records, historical documents, photographs, or any other written or visual material. Like existing statistics, content analysis involves bringing together and analyzing materials that already exist, rather than creating something new (such as a survey) to collect data. Researchers usually conduct content analysis to better understand some aspect of the norms, beliefs, or discourse of a particular culture or subculture, not to gain insight into the thoughts, feelings, motivations, or behaviors of individuals. This is usually true even when we analyze private communications like letters, diaries, or e-mails. The idea underlying content analysis is that we don’t always need to ask people questions in order to understand something about the social world in which they operate: We can learn much about a culture or subculture based on the words and images people use in private communication or for public consumption. Additionally, because people produce or consume these artifacts with different goals, they may not even consciously notice the subtle cultural messages embedded in these materials. As a method that does not involve the participation of people as respondents in the research, and because many audiovisual materials are publicly available, content analysis tends to be an inexpensive form of research whose timeline can be easily controlled by you (since you don’t have to wait for people to respond).

Content analysis can be conducted either quantitatively or qualitatively (or in some combination of the two). Though the data collection process is similar for both forms of content analysis, the data analysis process for each differs significantly. Quantitative content analysis involves counting how many times various patterns appear in the words or images being analyzed. For example, if you are conducting a quantitative analysis of alcohol advertisements, you might count how often the ads depict alcohol consumption as a daytime or nighttime activity; how often sports are included; whether the ad is coed or portrays only
Real Research: Research Methods Sociology Students Can Use

one gender; and whether the alcohol is shown in a bottle, can, glass, or in motion, such as shooting out of a bottle. You might count how many times words describing the taste are used, as well as words describing the effect of the alcohol. You might note the clothing items worn by the people in the ads, the position of their bodies (sitting, standing, lying down, in motion), who (if anyone) is touching the alcohol, and whether they are smiling when doing so. You might keep track of how old the people in the ad appear to be, and the activities in which they are shown engaging. You would then use these counts to identify the images and words that recur frequently, which you would use to tell you something about the norms and meaning of alcohol in our culture. Additionally, you could examine the relationships between different variables to understand how cultural ideas are related to one another.

Qualitative content analysis may involve looking at the exact same advertisements, but instead of counting images and words, you would try to identify particular themes: Does the advertisement evoke sexuality, and, if so, how? What images and words are used that make it feel sexual? What is that sexuality like? Is it aggressive or passive? Is it homoerotic or heteroerotic? Is it from

Real People, Real Research

Amanda Hendrix

For her undergraduate honors thesis, Amanda Hendrix conducted a content analysis of changing gender roles for female characters in comic books. Based on her literature review, which identified certain characters as having significance in comic book history, she followed two originally similar female characters from popular comics over 6 decades. She used theoretical sampling to choose 3 decades from which to sample, and to select a particular month and year in the first decade, and then analyzed the next four sequential issues. She used the same months and years from which to sample for subsequent decades. Amanda created a five-page quantitative coding sheet that focused heavily on gendered characteristics, behaviors, language, and displayed emotion. She used her literature review to help her locate existing measures, which she combined and expanded to operationalize these variables; then she revised the coding sheet after pretesting it, and made revisions before finalizing it. Amanda presented the results of her analysis at the Oceanic Popular Culture Association’s conference in Honolulu, HI.
a male or female perspective? What obvious or subtle links are made between the alcohol and sex: Is the alcohol shown as part of meeting a potential sex partner? As part of sex itself? As a tool to gain sexual cooperation? What words and images convey this message? In the case of qualitative content analysis, then, rather than looking for and counting certain predetermined words and images, you instead search for both subtle and overt themes and messages, and identify the words and images the advertising team has chosen to use in order to convey that theme or message to the reader. Qualitative content analysis may be more exploratory—it allows you to investigate a topic and see what themes arise as you look at the data, while quantitative content analysis requires that you begin with a more precise idea of exactly what sorts of words and images you are looking for. This means you must already be fairly familiar with the types of materials you will be analyzing.

It is important to point out here that content analysis is about analyzing the images and words in the materials gathered, not the intentions of the person who produced those images or words, or in the interpretation of them by the people who see or hear them. Likewise, content analysis is also not able to measure the effect that the materials have on people or on the culture. Content analysts can only describe or count what they see and hear in the materials analyzed.

METHODOLOGY

Quantitative content analysis, like other forms of quantitative research, most often uses positivist methodology. You will often begin with a list of hypotheses to test, which will help to determine what information you will need to collect from each object in the sample. You will most likely use probability sampling so that the results may be generalized, and you will numerically code the data for statistical analysis. Because no people are directly involved in the research, you don’t have to worry about objectivity in your relationship with the subjects, but you do have to maintain objectivity in collecting the data. Quantitative content analysis most often involves fairly large amounts of data, in part because it is a comparatively inexpensive research method, and therefore large samples are often feasible.

Qualitative content analysis, on the other hand, is somewhat similar to the use of semistructured qualitative interviews. The methodology used is generally an interpretivist one, where the researcher tries to understand the more subtle and complex meanings and themes woven throughout the data that underlie the more obvious content that can be counted. In conducting qualitative content analysis, you may still rely on random sampling strategies (though not
necessarily so) but will typically use a smaller sample because the analysis is more complex than mere counting. You will use consistent questions to help you gather similar data from the entire sample, but will be open to new themes that arise from the data, and will not be testing hypotheses or using predetermined variables. The goal in this case is to uncover the deeper and more complex cultural ideas and messages that may not be immediately noticed.

Often, researchers use a combination of both qualitative and quantitative content analysis. Because the methodologies are different, they may be conducted in two separate stages, or the qualitative work may be conducted on only a portion of the larger sample. The qualitative work may be used either as an exploratory stage of the research to help develop more precise variables for subsequent quantitative analysis, or it may be used after the quantitative analysis to better explain the deeper meanings and nuances that could not be thoroughly investigated with quantitative methods.

Some content analysts use a third methodology: critical methodology. This methodology takes as its main task pointing out to the reader things they have likely not realized: the ways in which words and images are being used to fool, manipulate, or indoctrinate the audience. Critical methodologists seek to reveal the ways in which power operates in mundane ways without the audience even being aware of it. Thus, critical social scientists think of themselves as activists, but also as objective. While these two terms normally do not go together, for critical methodologists, they are not contradictory. Objectivity to the critical social scientist means questioning the taken for granted so that you are seeing things as they really are, rather than as what you have been led to believe. By questioning cultural norms, beliefs, and discourse, the critical methodologist can shed light on the ways that these beliefs and norms are created and maintained in ways that support particular power arrangements. In content analysis, this is through the words and imagery that are used. The critical methodologist hopes that, much like the boy who pronounces that “the Emperor has no clothes,” once people see these power dynamics at play, they will be less likely to be duped or manipulated by them. Thus, their goal is not only revealing the truth, but in doing so, creating a social awareness whereby the population becomes more critical of these power arrangements and the status quo. Thus, from their perspective, their research is both objective and activist: In revealing what most people don’t see, they hope to create awareness that may lead to social change.

That being said, because critical methodologists seek to change people’s consciousness, there are two ways they may go about their research. The first is to try to build a solid case for the social change for which they are advocating by following many of the research protocols used by positivists. They will choose the materials that they sample in a nonbiased way and prefer large samples, so as to be able
to generalize their findings and demonstrate that the patterns they are highlighting are pervasive and not merely present in extreme cases or outliers. They will also be precise in their data collection, using standardized questions and taking standardized notes for each item in the sample so that they get the same information for each item. Furthermore, they often will work in teams to show that the patterns they identify are not just the result of personal interpretation, but that the results were replicated among different researchers within the research team.

Other critical methodologists sample by using materials that best illustrate their points, and thus avoid making generalizations to all such materials. This is similar to the style of research used by many documentary filmmakers. In the Dreamworlds movies by Sut Jhally, for example, in which Jhally analyzes the images of female sexuality in music videos, he first identified themes in hundreds of hours of music videos that he studied. Then he chose music genres and particular clips from music videos that best supported his analysis. He doesn’t ever claim that all music videos fit this analysis, but he does effectively show that many videos from different genres fall into these patterns. From his analysis and the examples he uses to support his points, the audience is likely to see the music videos in a different light, now noticing things they never had before, like camera angles, the male–female ratio in the videos, and the disparities in the ways that the men and women are depicted. While documentary filmmaking is not held to the same standards of peer review as published social research, some critical social scientists approach their research in much the same manner: They seek to change people’s perspectives by building a convincing case using carefully chosen data that are typical but that would not fit a positivist’s definition of either random or representative.

Although the criticalmethodologist hopes to create social change, content analysis is not frequently used for applied research (the same is true for positivist and interpretivist content analysis as well). Remember that applied research means using research to solve some specific problem or to help make a specific decision, usually by a particular organization. Although critical methodologists hope that those who read their research will see things in a new light, usually their goal is just that: to create awareness in their readers, but not to help provide a solution to a particular problem or decision. While someone analyzing alcohol advertisements may try to put pressure on beer companies to change their ads in some way, this is not the same as the company, for example, hiring someone to do content analysis in order to improve its own advertising strategy. The former remains basic research, even though there is some hope for change, while the latter is applied, because it uses research to solve a particular problem or make decisions for a particular organization. Regardless of methodology, content analysis is more likely to be basic research than applied.
Content analysis may be used to answer a variety of types of questions, depending on the kind of material being analyzed. When the material is part of pop culture or meant to be accessible to large groups of people (such as things posted on the Web), the questions often focus on cultural norms, how groups of people or particular behaviors are portrayed, the obvious and subtle cultural messages that are sent, and the ways that ideas are framed. If the material is private in nature (such as diaries, letters, or e-mails), questions may also focus on behavior reported by the author and their subsequent feelings and reactions to those items. If the material is historical, dating back to a specific historical time period or event, the questions may focus on the portrayal of that event, how cultural ideas of the time period were expressed, or the effects (as measured through reported behavior) of a particular event or time period. Content analysis is also appropriate for comparisons of the same or similar artifacts across time, such as advice columns in women’s magazines in the 1950s through the 2010s. Box 7.1 contains examples of research questions appropriate for content analysis.

As already mentioned, content analysis is not good for answering questions about how people interpret those media or pop-cultural portrayals, or the effects that they have. Pop-cultural artifacts do not necessarily accurately depict the lives of the kinds of people shown. Teenagers, for example, do not necessarily behave in real life the way that they are depicted in television sitcoms; nor do parents. What is interesting for the content analyst, however, is that particular types of depictions of both of these groups are used regularly—which, though not accurate, tell us something about cultural norms, values, and beliefs. This is particularly important to remember when analyzing historical artifacts, since we have less personal experience that would help remind us of the disjuncture between real life and portrayals. Additionally, even in the case of diaries and letters, it is important to note that what was written down is only a slice of someone’s life or thoughts. Letters are written with an intended recipient, and the writer may radically change, distort, or entirely ignore certain events or feelings, depending on the person for whom they wrote the letter. Likely, they would give a very different account of their prom date to their grandmother than to their best friend. Similarly, in diaries, people are most likely to write when they are struggling with an issue or problem, and to focus heavily on one issue or problem at a time. Just because a diary is filled with frustration and hope about the interest of a potential lover, it doesn’t mean that is the only thing the diarist ever thinks about, or that they aren’t facing other,
<table>
<thead>
<tr>
<th>Research Question</th>
<th>Appropriate for Content Analysis Because It Examines</th>
</tr>
</thead>
<tbody>
<tr>
<td>How do television game shows reinforce racial stereotypes?</td>
<td>Portrayal of a group of people</td>
</tr>
<tr>
<td>How is discourse used in presidential candidates’ political speeches to promote partisan hostility?</td>
<td>Subtle cultural messages</td>
</tr>
<tr>
<td>What changes have occurred in the portrayal of men’s bodies in sports magazines in the last 3 decades?</td>
<td>Portrayal of a group of people; comparison over time</td>
</tr>
<tr>
<td>What do people tweet about their anger and its causes?</td>
<td>Public expressions of personal thoughts</td>
</tr>
<tr>
<td>In what ways is the use of alcohol and drugs normalized in the Facebook postings of college students under the age of 21?</td>
<td>Cultural norms</td>
</tr>
<tr>
<td>How do newspaper articles about debates over gay marriage reinforce or challenge heteronormativity?</td>
<td>Framing of ideas</td>
</tr>
<tr>
<td>How are gender roles portrayed in the diaries of American pioneer women headed West?</td>
<td>Expression of cultural ideas; historical evidence</td>
</tr>
<tr>
<td>How do popular song lyrics depict love?</td>
<td>Framing of ideas</td>
</tr>
<tr>
<td>How has discourse in medical textbooks about what constitutes a healthy pregnancy changed over time?</td>
<td>Comparison over time</td>
</tr>
<tr>
<td>How was the nation’s reaction to the assassination of Dr. Martin Luther King Jr. portrayed in media editorials in southern states as compared to those in northern states?</td>
<td>Portrayal of historical events</td>
</tr>
</tbody>
</table>
perhaps even more serious, problems or issues. Thus, it is important to realize that what gets recorded or preserved in letters and diaries is only a glimpse into someone’s life, not an accurate depiction of its entirety.

**Check Your Understanding**

Write a research question related to race that is appropriate for content analysis. Explain what materials you would analyze in order to answer this question.

**LITERATURE REVIEW**

The literature review for content analysis is done the same way and for the same reasons as for the other methods. You will want to familiarize yourself with the research that has already been conducted so that you identify specific themes, images, or wording to look for when you analyze, and how you might measure them. You may try to find articles about a similar topic, but one that has been studied using a different group of materials, such as comparing the results from a well-known study of the portrayal of girls in children’s books to your portrayal of girls in children’s music. Much of the literature you review may not itself use content analysis; you may wish to use other research in order to get a sense of the norms and beliefs about a particular topic, as well as how those may have changed over time. In studying the portrayal of young girls, you also might look at research that addresses cultural beliefs about childhood, as well as research on adult gender roles. If you are conducting historical content analysis, you will also want to give yourself a good foundation for understanding the events and conditions in that time period, as well as of particular behaviors or beliefs that were then in transition.

**SAMPLING**

After writing the research question and conducting the literature review, the next step you will likely take is choosing your sample. Although qualitative and quantitative researchers draw their samples differently from each other when using other research methods, in content analysis the process is often identical, regardless of whether the research will be quantitative or qualitative. It involves
not only choosing the general category of materials you will use, such as children’s television cartoons, but the exact items as well—in this case, the specific titles and episodes you will use. Because content analysis is usually a relatively inexpensive research method to conduct, you are likely to draw a large sample. Sometimes, if there are a limited number of such materials, your sample will be a census, meaning the sample exactly (or almost exactly) matches the population of possible items. In the case of content analysis, that means that you include every available instance of the material in your analysis. If you were analyzing sociology textbooks for a Sociology of the Family course, you might choose, for example, every published textbook for this and similarly titled courses available from every American and British publisher, including out-of-print editions, from the past 10 years.

If your population is larger, such as children’s television cartoons, it will be impossible to do a census, and you will need to select which cartoons, which episodes, on what channels, and during what time period. Much like that for observation research, the sampling strategy may be multilayered, with different techniques used for different sampling decisions. You may use theoretical sampling to choose the time period from which to sample your cartoons. Knowledge you have about the industry, the broadcasting cycle, relevant legislation, or related cultural/historical events may influence the time period you select. If you know, for example, that two independent animation studios were bought out by a larger corporation within the last 2 years, you may seek to include cartoons from both before and after the buyouts. Similarly, if your research question seeks to make a comparison, the time comparisons you make should be informed by your literature review. Rather than randomly comparing cartoons from 1973 to those from 2003, for example, you would likely have theoretical reasons for choosing the years you did. You also may take accessibility into account: Perhaps 1973 is the first year for which you are able to access a large number of recorded cartoons. Additionally, if you were choosing to study only cartoons currently being broadcast, accessibility may affect the length of time for which you sample. If you are only able to record two shows at once on your DVR, you may end up sampling for a longer period of time (a month) than if you were able to record an unlimited number of shows at once (perhaps a week or two).

Once you have chosen your time period, you will need to pick a sampling method for selecting the specific items to be included in your final sample. Frequently, you will choose some sort of random sample, using one of the four probability sampling techniques: simple random sample, systematic random sample, stratified random sample, or cluster sample. These were discussed in Chapter 4, on surveys, and are reviewed in Box 7.2. In each case, the selections are made randomly, and each item (in this case, each aired episode) has an
equal chance of being selected. Remember that the benefit of probability sampling is that the results are generalizable to the larger population of materials (all aired television cartoons during the selected time period).

<table>
<thead>
<tr>
<th>Sampling Method</th>
<th>Description</th>
<th>Example for Content Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple random sample</td>
<td>Each item is given a number, and items are chosen using a random numbers generator.</td>
<td>Every cartoon episode scheduled to air is listed and given a number. Random numbers generator used to choose episodes.</td>
</tr>
<tr>
<td>Systematic random sample</td>
<td>Sampling frame is randomized, then each item is given a number. First number is randomly chosen, then count down the list to every nth item to select it. N represents the number in the sampling frame divided by the total number desired in the sample.</td>
<td>Every cartoon episode scheduled to air is listed in randomized order. To sample 300 cartoons out of 1,054 listed, start at a random place on the list, and then choose every 4th cartoon listed (1,054/300 = 3.51, round to 4).</td>
</tr>
<tr>
<td>Stratified random sample</td>
<td>Sampling frame is divided into groups and then simple random or systematic sampling is used to select a given number of items from each group.</td>
<td>Listed cartoons are divided between those with animals/nonhumans and those with people as the main character. Simple or systematic random sampling used to choose 150 cartoon episodes from each list, for a total of 300 cartoons.</td>
</tr>
<tr>
<td>Cluster sample</td>
<td>Groups are randomly chosen, and every item from the chosen groups is included in the sample.</td>
<td>Cartoons are divided by the time that they are broadcast. For 30-minute shows, there are 336 possible time slots in a 7-day period. Time slots are randomly chosen using simple random sampling. All episodes on all channels broadcasting at the randomly selected time periods are included in the sample.</td>
</tr>
</tbody>
</table>
Sometimes, however, especially if you are using interpretivist or critical methodology, you may choose a nonprobability sampling technique, such as theoretical sampling, in order to select the cases that are most theoretically interesting. Instead of choosing all of the aired cartoons during a 2-week period, you may instead sample all the episodes over a longer period of the 10 most popular cartoons, since children will be most likely to see these. Or you may select cartoons that specifically aim to be different than other cartoons in some sort of way, such as the ones touted as being empowering to girls or that are aimed at a multicultural, bilingual audience.

In determining your sampling frame, you will also need to decide upon the eligibility requirements for inclusion. If your population is all televised children’s cartoons aired within the next 2 weeks, you will need to specify what it means to be “televised” in order to determine which shows belong in your sampling frame. If the show is available on a home satellite system, but not on cable, will you include it? If it is aired on a channel only available to people with the most expensive cable package, will you include it? If it is animated and airing on a movie channel such as HBO, does it count as a “television cartoon”? And what is meant by “children’s” cartoon? Does *The Simpsons* count? All of these are strategic decisions with no right or wrong answer, but your decisions will affect your sample and therefore your data, so they should be considered carefully. Once again, convenience and accessibility may count: Although some children may watch cartoons available on expensive specialty channels, you may decide that you will only select cartoons aired and accessible to you with your basic cable package, both because it saves you money from upgrading your cable subscription, and because you know that many families also have only the basic package.

If you are analyzing historical items, you will be less likely to use random sampling and instead include all available items, or use a combination of theoretical and convenience sampling to choose your items. Often in historical research, however, you will have to work much harder to collect the items in your sample. If you are analyzing diaries, for example, you may need to visit the archives and special collections of several libraries around the country to access them. Since you may not be able to visit every library with relevant diaries in its collection, you may choose to visit those with some particular theoretical interest or that are conveniently located within an easier traveling distance from you. Your sample size will vary considerably depending on the availability of the objects you are studying, the type of object, and the length/size of these materials. If you are studying movies that last 1.5–2.5 hours, you will probably sample many fewer than if you are studying 30-second commercials. Likewise, if you are studying media objects that are widely and publicly
available, you will probably have a much larger sample than if you are studying private diaries. There is no particular rule of thumb here, but typically we study as many of the objects as we reasonably can, given their length and availability. It is not unusual for researchers to analyze a couple of hundred films, or several hundred advertisements, newspaper articles, or blog postings. Although research using social media such as Facebook postings and tweets is still new, one may expect to see a sample in the thousands for such research.

Check Your Understanding

Describe all the steps you will take, and the sampling method(s) you will use, for your research question about race. How large will your sample be?

CONCEPTUALIZING AND OPERATIONALIZING

Quantitative Content Analysis

As might be expected, operationalizing and conceptualizing for qualitative and quantitative content analysis differ considerably, so I will address each separately. Once you have gathered your materials, you will develop a coding sheet. For quantitative content analysis, a coding sheet is the instrument on which you record all of your counts. It is standardized, prompting you to gather the exact same information from each object and to record the notes in a standardized and systematic way. This is important for gleaning parallel information from different objects sampled and also for standardizing information across the research team. Like a survey instrument, a coding sheet takes a considerable amount of time to develop and perfect. Because it will guide all of your data collection, you will want to test and revise your coding sheet multiple times on a variety of items from your population.

The coding sheet should begin with a record of the exact object from which you are recording the information. If you are analyzing magazine advertisements,
for example, you would want to record the magazine name, date, volume number, page number, and product name. If you are analyzing websites, you need to include the website name, Web address, and the date and time it was accessed. If it is a letter, you should include the author, the recipient, the date, and the number of pages. This information is important: If there are discrepancies between researchers, or questions about the data arise later, you want to be able to easily locate the exact item in question.

From there, the coding sheet goes on to list particular information that you wish to be gathered. Like survey questions, the coding sheet for quantitative content analysis includes mostly closed-ended variable items with a series of answer choices for recording information. This increases efficiency as well as standardizes the information. Also like surveys, these variables must be both conceptualized and operationalized. Remember that to **conceptualize** is to define important variables, while to **operationalize** is to determine how to measure those variables. Let's say you are analyzing advertisements in men’s magazines to understand how traditional stereotypes of masculinity and femininity are both reinforced and challenged in these magazines. You would need to start by conceptualizing *traditional stereotypes of masculinity* and *traditional stereotypes of femininity*. How will you know these when you see them? You will also need to conceptualize *reinforcing* traditional stereotypes and *challenging* them. Suppose you define *challenging traditional gender stereotypes* as the following:

- Showing males or females engaged in behaviors or having the appearance usually associated with the other gender
- Showing people engaged in behaviors or having an appearance that is not usually associated with either gender
- Showing people engaging in behaviors or having an appearance consistent with traditional gender roles but in a highly exaggerated way that is probably meant to poke fun at or make the traditionally gendered roles seem funny, out of date, or deviant

This is a complicated definition! And you would still need to break each part of it down, defining each part further and more precisely.

To operationalize these definitions, you will need to decide how to measure them. In this example, you would first need to indicate whether there were males or females depicted in the advertisement, and how many of each. Then, based on the definition above, you would need to break “appearance” (from the first part of our definition) down into different...
measurable aspects: Perhaps you would include clothing, body position, hand gestures, facial expression, hair length and style, and the presence or absence of obvious makeup. Each of these would become a separate variable on your coding sheet, and you would provide check-off options for each. For body position, for example, you may include standing, jumping, walking/running, sitting, bending over forward, kneeling, lying on stomach, lying on back, lying on side, indeterminate, other (with space to specify), or head-only shot. In collecting the data, for each male and female in the advertisement, then, you would separately record on the coding sheet their body position using these standardized categories. You would go through the same process to operationalize your other variables as well, providing prompts of what to look for and check-off options or spaces for counts. As with surveys, you want your options to be both exhaustive and mutually exclusive in order to increase reliability in coding. Coding sheets must be precise and may be extensive. Even analyzing a single, one-page advertisement may involve coding for dozens of variables; thus, coding sheets are often multiple pages long. Box 7.3 shows an example of one page of a coding sheet for the project just described. Note that this one page is not comprehensive—the coding sheet would need to have many more pages in order to measure all the important variables from each ad, but this excerpt gives you an idea of what a coding sheet looks like.

**Qualitative Content Analysis**

Once upon a time, qualitative content analysis also relied primarily on coding sheets, though the questions were open-ended ones instead of counts and check-offs. Today, the analysis of textual objects is usually done by directly uploading the text items into a qualitative analysis software package and coding the data in much the same way you would qualitative interview transcripts. Images, video, and sound can now also be coded using such software. This software allows you to attach qualitative codes (words that represent larger themes or ideas) to particular points on a visual image. Although this innovation was enormously helpful for some forms of analysis, it can also be cumbersome, especially for audio and video files; therefore, some qualitative researchers still use coding sheets for qualitative analysis of these objects. Although using coding sheets and using software are both called “coding,” the procedures are very different. For the former, the distinction between collecting the data and analyzing them is blurred, but when
**Box 7.3  Sample Quantitative Coding Sheet**

**Gender Stereotypes in Men’s Fitness Magazines**

<table>
<thead>
<tr>
<th>Coder's initials: __________________</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magazine title: __________________</td>
</tr>
<tr>
<td>Date: __________ Volume no.: __________ Issue no.: __________ Page no.: __________</td>
</tr>
<tr>
<td>Advertisement for: Brand name: __________ Type of product: __________</td>
</tr>
<tr>
<td>Number of males: __________ Number of females: __________ Number of indeterminate gender: __________</td>
</tr>
</tbody>
</table>

**Physical position of male no. 1:**

<table>
<thead>
<tr>
<th>Standing</th>
<th>Jumping</th>
<th>Walking/running</th>
<th>Sitting</th>
<th>Bending forward</th>
<th>Kneeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lying on stomach</td>
<td>Lying on back</td>
<td>Lying on side</td>
<td>Head shot</td>
<td>Indeterminate</td>
<td>Other</td>
</tr>
</tbody>
</table>

**Facial expression of male no. 1:**

<table>
<thead>
<tr>
<th>Smiling</th>
<th>Laughing</th>
<th>Frowning</th>
<th>Crying</th>
<th>Yelling</th>
<th>Grimacing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Staring</td>
<td>Neutral</td>
<td>Not visible</td>
<td>Other</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Emotion expressed by facial expression of male no. 1:**

<table>
<thead>
<tr>
<th>Happy</th>
<th>Sad</th>
<th>Angry</th>
<th>Bored</th>
<th>Cold</th>
<th>Confident</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dreamy</td>
<td>Worried</td>
<td>Calm/relaxed</td>
<td>Curious</td>
<td>Hopeful</td>
<td>Indeterminate</td>
</tr>
</tbody>
</table>

**Clothing of male no. 1:**

<table>
<thead>
<tr>
<th>Shirt</th>
<th>Pants</th>
<th>Shoes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long-sleeve button-down</td>
<td>Jeans</td>
<td>Sneakers</td>
</tr>
<tr>
<td>Long-sleeve casual</td>
<td>Sweat/athletic pants</td>
<td>Hiking/outdoor athletic shoes</td>
</tr>
<tr>
<td>Sweater</td>
<td>Khaki-style</td>
<td>Fashion casual</td>
</tr>
<tr>
<td>Sweatshirt</td>
<td>Dress pants</td>
<td>Sandals</td>
</tr>
<tr>
<td>Long- or short-sleeve T-shirt</td>
<td>Suit</td>
<td>Dress shoes</td>
</tr>
<tr>
<td>Long- or short-sleeve athletic</td>
<td>Bathing suit</td>
<td>Cowboy boots</td>
</tr>
<tr>
<td>Short-sleeve casual</td>
<td>Underwear</td>
<td>Work boots</td>
</tr>
<tr>
<td>Tank top</td>
<td>No bottoms</td>
<td>Barefoot</td>
</tr>
<tr>
<td>Coat/jacket</td>
<td>Other</td>
<td>Other (specify)</td>
</tr>
<tr>
<td>Suit</td>
<td>Unknown</td>
<td>Indeterminate</td>
</tr>
<tr>
<td>Shirtless</td>
<td>Other (specify)</td>
<td></td>
</tr>
<tr>
<td>Other (specify)</td>
<td>Indeterminate</td>
<td></td>
</tr>
</tbody>
</table>

**Clothing of male no. 1 is (check all that apply):**

<table>
<thead>
<tr>
<th>Loose</th>
<th>Tight</th>
<th>Sexy</th>
<th>Sweaty</th>
<th>Professional</th>
<th>Sophisticated</th>
<th>Comfortable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revealing</td>
<td>Frumpy</td>
<td>Trendy</td>
<td>Hippie</td>
<td>Retro</td>
<td>Outdoorsy</td>
<td>Other</td>
</tr>
</tbody>
</table>

Copyright ©2016 by SAGE Publications, Inc. This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
using the software, the data collection and analysis phases are actually the same and there is no distinction at all. Thus, in this section we will discuss the coding of audio and visual images/video using coding sheets, but will reserve the discussion of using software to code any type of objects until our discussion of data analysis.

When coding images, audio files, or videos, coding sheets may be used to prompt and record observations. The coding sheets used for qualitative content analysis are more open-ended than those for quantitative analysis, and they generally provide prompts in the form of questions aimed at cuing the analysts about what to look for, but leaving the answers to those questions generally open-ended, with lots of room for note-taking. Let’s return to our previous example, but this time from a qualitative perspective. You are analyzing advertisements in men’s fitness magazines to understand the ways in which traditional gender roles are reinforced and challenged in pop culture targeted to men. Although you are also considering the text of the advertisements, our focus here will be on the visual images. Instead of breaking the data down into specific variables, you would ask open-ended questions that would allow you to focus more holistically on the images and the messages they convey about gender roles. This is much like the difference between a survey and a semistructured interview: In the survey (quantitative), the answer choices are provided, very precisely measuring and standardizing—but also limiting—the information that is collected; in the semistructured (qualitative) interview, the same questions are asked of each participant so that there is some standardization, but the questions are broad and open-ended so that the each participant can give a wide range of general or detailed information that will surely differ in content, form, and depth from that of other participants. In qualitative content analysis, then, the qualitative coding sheet is a guide to remind you what general information to look for, but your answers will vary considerably for each advertisement and will not ultimately be reduced to numbers. Thus, you will get different (but not better or worse) information than you would from quantitative content analysis.

With qualitative content analysis, you will still need to conceptualize the major ideas in your research question (in this case, traditional stereotypes of masculinity, traditional stereotypes of femininity, reinforcing gender stereotypes, and challenging gender stereotypes) by defining them. To operationalize them, however, you will develop the open-ended questions that will appear on your coding sheet. In this example, to operationalize challenging gender stereotypes you might include the following questions: In what ways does the physical appearance of the male(s) in the ad seem unmasculine,
Box 7.4 Sample Qualitative Coding Sheet

Gender Stereotypes in Men's Fitness Magazines

Coder's initials: _____________
Magazine title: _____________
Date: __________ Volume no.: __________ Issue no.: __________ Page no.: __________
Advertisement for: Brand name: _____________ Type of product: _____________
Number of males: ____ Number of females: ___ Number of indeterminate gender: ___
In what ways does the males' physical appearance connote masculinity?
In what ways does the males' appearance seem gender-neutral, unmasculine, or feminine?
What activities are the males engaged in? How typical or atypical is this behavior for men?
What archetypes or stereotypes of men come to mind?
How are the men interacting with one another? What type of relationship between them seems to be suggested, and how is this done?
How are the men interacting with women? What type of relationship between them seems to be suggested, and how is this done?
How does the power seem to be balanced? How obvious is this? How is it depicted?

androgynous, or feminine? What is the male(s) doing that is somewhat unusual for men to do? What emotions are evident in this ad, and how are they communicated to the reader? In what ways is the gendered behavior exaggerated, and does this seem to be reinforcing or poking fun at traditional gender roles? Note that these questions are less “objective”—they report not only what you observe in the images, but also what you interpret them to mean. Such interpretations are subjective, of course, and may vary from person to person. For this reason, it is very important to include as much information as possible on the coding sheets about why you are making this interpretation, based on what you visually observe. As many details about the image that led to this interpretation should be included as possible. This helps to give support for the interpretation and to assist in comparing interpretations across advertisements. It also will help to document patterns in
Conceptualize two concepts for your research question about race. Then do the following, remembering that you will likely need several measures for each concept:

a) Operationalize these in a quantitative way, writing them as you would for a quantitative coding sheet.

b) Operationalize these in a qualitative way, writing them as you would for a qualitative coding sheet.

ETHICS

When you are doing content analysis of publicly available materials, no informed consent is needed, nor must special precautions be taken to protect people whose name, image, or information is included. Sometimes things that we think of as private, such as letters or diaries, have already been made publicly available. There are many books published, for example, of people’s diaries and letters. These also would not need any protection because you would not be releasing new information about the individuals. If, however, the materials have not previously been made public, then you will need to take a variety of precautions, depending on the type of material, whether the individuals on which it focuses are still alive, and who owns the materials. If, for example, you were analyzing diaries written by people who are still living, you will need to get their informed consent and to protect their identities by giving them pseudonyms, changing the names of the people they write about, and deleting identifying information from your report. If you are analyzing the unpublished diaries of people who have already died, you should check with your Institutional Review Board (IRB). They may require you to get the informed consent of a family member or closest living relative.
You may also have to provide pseudonyms for the author as well as the people they wrote about, especially those who are still living, and you may need to delete certain identifying information. If the private materials are in your keeping, you should keep them in a locked drawer or file cabinet to further protect the author’s identity.

PREPARING FOR DATA COLLECTION

Preparations for data collection for both qualitative and quantitative content analysis focus on three things: gathering the sample, pretesting and revising the coding sheet, and training the research team (if there is more than one person who will be coding the data, which is common for content analysis). Gathering the sample may include purchasing or recording the audiovisual materials that you will be analyzing; downloading blogs, tweets, or Web pages; gathering magazine issues; or photocopying archived diaries, letters, or other materials. In the case of historical materials, this may involve traveling to different libraries or archives across the country or even around the world. Most often, however, you will use materials you can more easily access, and the focus of the task is to physically gather those materials. It is important to have continued access to the materials (through recordings, photocopying, scanning, printing, etc.), as very often you will want to go back and look at particular items again after coding them.

Pretesting the coding sheet involves recording the data on your coding sheet exactly as you expect to do for the research project, using a diverse subsample of your larger sample. As you code these items from your sample, you may notice that in some cases the coding sheet is not specific enough, not clear enough, or that you have cases in which the way you operationalized something is not as exhaustive as you imagined. As you come across these cases, you can revise the coding sheet to better fit the items. The wider the diversity of your items in this pretest subsample, the more likely you are to have made all the necessary changes to the coding sheet before you begin your actual data collection. This is important because once you begin recording the actual data, you cannot change your coding sheet without also having to recode all of the items you already coded prior to the change.

If there will be more than one coder, it is extremely important to train the research team so that members are as consistent as possible in the way that they code items. For qualitative content analysis, this will include documenting how each item is conceptualized. From the example in Box 7.4, what does it mean to have a “masculine physical appearance,” and how
does this differ from a “gender-neutral” one? What are archetypes and stereotypes, and what are some of the common ones that may get invoked? What is meant by power in this context, and what will you look for to recognize it?

For quantitative content analysis, training the research team begins with writing a codebook, just as in survey research. The codebook documents the variables used and the numeric codes to be entered into the statistical software for each response. In addition, however, the codebook for quantitative content analysis should include guidelines for choosing the different categories within each variable. This is not necessary for survey research, because the interpretation of the question and the categories are left up to the respondent. In content analysis, however, it is the researchers—not the participant—who must interpret the variables and related categories; therefore, documenting the definitions, differences, and guidelines for choosing these is important to increase reliability between coders, as well as among individual coders over time.

Training the team also requires thoroughly discussing each coding question or category on the coding sheet, and coming to agreement on how different cases should be coded. Then coders each code a subset of the sample (with all coders coding the same items from the sample), and then comparing the coding. Any items that were coded differently should be discussed, and you should reach an agreement about how to code similar cases in the future. This increases the intercoder reliability (the likelihood that multiple researchers have coded an item in the same way), which is often considered an important measure of the quality of the analysis, though it is more often used in quantitative than qualitative content analysis.

A list of codes that will be used should be initially generated, and then pre-tested on the subsample, with new codes being added and other codes being modified as coding progresses. The research team should operationalize the different codes, document them, and practice using them on a subset of materials, comparing their coding. Discrepancies should be discussed and resolved, with protocol developed for future similar cases.

**DATA COLLECTION**

At this point, data collection is rather simple, in that you are simply coding the full sample that you have collected, using the procedures you have worked out in your pretest. Although this may be simple, it is often time-consuming; and depending upon the types of items being sampled, the size
Chapter 7  Content Analysis

of those items, and the length of the coding sheet, it may take hundreds of hours. Research teams usually continue to meet during this process, and intercoder reliability (if multiple coders are used) is monitored. Research teams discuss items that individual coders had difficulty coding, find and discuss discrepancies, and problem-solve for new difficulties that weren’t discovered during pretesting.

DATA ANALYSIS

Quantitative Data Analysis

Quantitative data analysis usually begins with transferring the data from the coding sheets into SPSS or Excel. Each variable on the coding sheet is entered into the spreadsheet, and its categories are given numerical equivalents that are then entered for each case (each advertisement, letter, TV show, blog, tweet, etc.) that has been coded. This yields a database for which you may then calculate statistics beyond just counting the number of responses for each category (which would be reported as frequencies). This will allow you to investigate the relationship between different variables, and to test hypotheses.

After entering the data, it must be cleaned. Because the researcher is filling out the coding sheet, there should be little to no error on the coding sheet itself: no missing data, no ambiguous markings, etc. There may, however, have been mistakes in entering the data into the software, and therefore this should be checked using spot-checking (choosing coding sheets—either randomly, or systematically, such as every 10th sheet—and double-checking to make sure that the data have been correctly entered).

After cleaning the data, data analysis may begin. This will usually start with running frequencies, which will tell you the most common observations you made for each variable. Relationships between variables can be examined using crosstabs, correlations, and other bivariate statistics. In the previous example you may, for instance, explore whether when one person who challenges gender stereotypes is depicted, the others in the advertisement are also more or less likely to do so. You could investigate how conformity to traditional gender roles regarding appearance is related to conformity to these roles regarding behavior or interaction. And you might investigate whether it is males or females who are more likely to be shown challenging these gender roles, and if the ways they do so are similar or different.
In addition to calculating the statistical relationship between variables, if your objects of analysis are in written form, you may also use a word-processing system to conduct a word count on particular words. You could do a word search, for example, on the words *power* and *powerful* to see how often they are used in your ads. Of course, word searches take into account neither context nor meaning, so you have to be careful when drawing conclusions. Word counts may be used on their own as a type of frequency data. Alternatively, you can use a word processor’s search feature not only to count particular words, but also to investigate their context or their relationship to other words. You can record these data and enter them into the statistical software to use as variables. You may want to determine, for example, whether the words *power* and *powerful* are more often used to describe men or women; to see which other descriptors are most likely to be used in sentences containing these words; or to ascertain whether these words are most likely to be used to describe physical characteristics, performance, emotional states, or sexual scenarios.

**Qualitative Data Analysis**

If you are uploading your objects directly into a qualitative software analysis package (such as ATLAS.ti, NVivo, CAT, or Ethnograph), there is no distinction between data collection and data analysis. After uploading your objects, you will code them using the exact same procedures you would use for analyzing qualitative interviews. You will begin by generating a list of *a priori* codes that you expect to use in your analysis. Remember that *a priori* codes are codes you come up with in advance of conducting your analysis. You will conceptualize those by writing definitions of the codes to help you clarify their meaning and how you will use them (thus improving the consistency, or reliability, of your coding). Once you have generated a list of *a priori* codes, you will begin the process of **open coding** by attaching those codes to particular sections (usually text, but they could also be parts of images or videos) of the items you are analyzing. As you attach these codes, you will continue to develop and define new codes, adding them to your list.

Once you have coded all of your objects, you will begin looking for patterns in your data. This phase is called **axial coding**, and you may remember that there are many different types of patterns for which you can search, with frequencies, magnitudes, types, processes, and structures among them...
Chapter 7  Content Analysis

(see Chapter 2). You can find a brief review of these patterns in Box 7.5. The software packages can be very helpful in aiding you to see these patterns by sorting and focusing on different parts of the data in a wide variety of ways.

Selective coding is the third stage of analysis, and it involves systematically searching for negative cases for the patterns you think you have identified in

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>How often something occurs across objects or within a single object</td>
<td>In 100 issues of men’s magazines, men are depicted doing housework or childcare 27 times, but are depicted playing sports 568 times.</td>
</tr>
<tr>
<td>Magnitude</td>
<td>How profoundly something occurs, even if it occurs rarely</td>
<td>While references to sexual assault were usually subtle in the magazines, there were five cases in which men were explicitly and blatantly encouraged to have sex with women who were too drunk to consent.</td>
</tr>
<tr>
<td>Type</td>
<td>Subgroups of something that occurs among many of the objects</td>
<td>You find there are four types of challenges to masculine gender roles in men’s magazines: appearance-related, emotional, sexual, and financial.</td>
</tr>
<tr>
<td>Process</td>
<td>Steps or stages that are common among the objects in getting from point A to point B</td>
<td>In examining changes in portrayals of masculinity over the decades in men’s magazines, you identify that several of the magazines go through five stages over time in changing their depictions of men: exploration, reimagining, commitment, backlash, and modification.</td>
</tr>
<tr>
<td>Structure</td>
<td>The essential parts making up a phenomenon</td>
<td>Through your analysis, you identify four different fundamental aspects of what it means to challenge gender roles: critique, belief, action, and response.</td>
</tr>
</tbody>
</table>
axial coding. This stage is important because it helps to validate your analysis, ensuring that you are not ignoring data that contradict or challenge the patterns you have identified.

As with all qualitative analysis, you should also write memos throughout all three stages of analysis. These memos document procedures that you have completed or still want to complete; but even more important, they document your thoughts, ideas, hunches, insights, and questions as you progress through the analysis. This can help to strengthen your analysis by reminding you of issues you may forget about while focusing on following another lead, or by suggesting further patterns for which to search.

If you used coding sheets for your qualitative analysis, you will still follow many of these procedures. You may wish to start by uploading your coding sheets into the software package, and then coding your own coding sheets. This may sound like analysis of analysis, but the software can help you sort themes and identify patterns in the data on the coding sheets. Conversely, you may choose to skip this step, which will save coding time but will make identifying patterns more difficult. If you choose to search for patterns without the help of a software package, you will take the physical copies of your coding sheets and sort them using old-fashioned methods: putting them into piles based on themes and subthemes, color-coding sections of the coding sheet for particular codes or patterns, and reading and rereading them. Some people use large pieces of butcher paper to map out connections and patterns among different pieces of the data. In all of these cases, you are still searching for patterns across your coding sheets. Once you have identified those patterns, you will, as always, search for negative cases in order to validate the patterns you have found.

**Check Your Understanding**

*Develop a list of a priori codes for your qualitative research about race.*

**EVALUATING THE QUALITY OF DATA AND ANALYSIS**

For quantitative content analysis, reliability in coding of items is important and is usually measured through intercoder reliability. There are many different ways of calculating this, and there is no clear agreement on which of the measures is best. There are a handful of specialty programs that will calculate this for you, and SPSS includes one test that can also perform this function. All of these tests
measure the items that were coded discrepantly among the coders, and compare them to what would have happened by chance alone. Intercoder reliability is measured on a scale of 0 to 1.0, with 0.8 or higher showing a good level of reliability. Remember that reliability is related to consistency—in this case, that all coders interpreted the item as well as the question and categories on the coding sheet in the same way. On the other hand, validity, as you may recall, is the idea that you are measuring what you think you are measuring. Validity is also important in quantitative data analysis: You must demonstrate to your readers that you operationalized your variables in ways that make sense. If you are trying to operationalize submissiveness, for example, you will need to convince your audience that the appearances, body positions, and words that you looked for are actually connoting submissiveness and not something else. Crying, for example, may indicate many things, but it is probably a better measure of sadness than submissiveness.

Evaluating the quality of qualitative content analysis typically relies more on validity than on reliability, since it is hard to calculate intercoder reliability when coding is open-ended and what is being measured is less precise. The better you have conceptualized and operationalized, however, the more reliable your coding will be between objects and over time, thus improving the overall quality of your research. Validity is of utmost importance, as it is with all qualitative work. Validity is demonstrated by giving detailed examples from the data. This may be in the form of quotations from written materials, or thorough descriptions of visual materials. In some cases, you may include a few visual images from the data (if appropriate) in your presentation of the results so that the reader can see first-hand how the interpretations were made and how the theme is apparent. The validity of your analysis depends on the amount of support you have throughout the data for the patterns you’ve identified. Additionally, your analysis can only be valid if you have done selective coding, where the fewer the number of negative cases that the selective coding yields, the more valid your analysis.

**PRESENTING THE RESULTS**

Writing up the results for publication of content analysis is similar to the presentation of other data. After presenting the literature review, you will give a thorough explanation about all steps, stages, and decisions made in sampling, as well as a description of the final sample used. You will also discuss at length the items on the coding sheet (if applicable) and how concepts were operationalized, and the steps taken to minimize discrepancies either across coders, or
from item to item. This will be followed by a description of your major findings, relationships between variables, or themes found within the data. You will end with a description of what these data teach us and why they are important.

**SUMMARY POINTS**

- Content analysis is the method to use when studying textual, image, audio, or video artifacts. Its purpose is to understand both the obvious and more subtle messages in order to uncover cultural norms and beliefs.
- Content analysis may be either qualitative or quantitative or a mix thereof, and it may be conducted using positivism, interpretivism, or critical methodology.
- Content analysis is most often conducted in teams, and significant care must be taken in conceptualizing, operationalizing, and training the team in order to maintain reliability in coding the artifacts.
- Although content analysis is usually an inexpensive method to conduct and is one in which the researcher has an extraordinary amount of control (since there are no research participants on which to rely), it can also be a very time-consuming method, as coding can be tedious and must be meticulously recorded, and sample sizes can be quite large.
- Quantitative analysis for this research method is very similar to that of surveys; qualitative analysis is like that of interview research. Thus, the basic skills of analysis are transferable to other methods.

**KEY TERMS**

- *a priori* codes
- axial coding
- census
- coding sheet
- conceptualize
- critical methodology
- frequencies
- intercoder reliability
- memos
- objectivity
- open coding
- operationalize
- qualitative codes
- qualitative content analysis
- quantitative content analysis
- selective coding
- spot-checking
- theoretical sampling