In April 2000, almost 10 years after I finished my PhD at UW–Madison, Jack and I were attending a conference in Hong Kong where I presented a paper on changing meanings of media consumption in China as a result of more than two decades of economic reforms. I was arguing a cultural studies thesis with survey data. Two discussants were respectfully critical, but I was unpersuaded until Jack joined the discussion and said, in his characteristic, soft-spoken voice, "The paper needs stronger conceptualization." As he offered his thoughts in some detail, I felt the sensation of seeing a ray of light, the same feeling that I had experienced many times as his student. Yes, inadequate conceptualization was the problem in my paper, and Jack pointed out a way for me to salvage it.

That was typical of Jack. He always demands conceptual clarity in thinking. I learned that not only from his lectures but also from our almost daily conversations, which ranged from sports (on which I had practically nothing to contribute), to politics, news, and, of course, research. The idea of writing a piece on concept explication tailored to beginning research students emerged in these conversations. For me, working with Jack on this piece was a unique opportunity to learn the logic of concept explication and to observe how Jack organized and expressed ideas. He insisted on having real examples (i.e., from real data) in the empirical analysis part. He didn't want just to talk about concept explication; he wanted to show how the logic would be applied in the treatment and interpretation of data. When I brought him SPSS printouts, I could see glee in his eyes.

Almost 14 years later, I am back in Vilas Hall, revising that piece for the inclusion in this volume. The process has been akin to reliving those experiences. Years ago, Stanford University's Byron Reeves, after learning that I was working with Jack on a paper, told me: "You'll learn a lot by writing papers with Jack." It's easy to agree with Byron on this. I learned a lot when drafting this piece with Jack and again in rewriting it nearly 14 years later. How many people get a second opportunity to learn from their adviser? I must admit, I am truly lucky on this one.

Zhongdang Pan

Jack M. McLeod is Maier-Bascom Professor Emeritus at the University of Wisconsin–Madison. (BS, University of Wisconsin; MS, University of Wisconsin; MA University of Michigan; PhD, University of Michigan)

Zhongdang Pan is a professor in the Department of Communication Arts at the University of Wisconsin–Madison. (BA, Beijing Broadcasting Institute, China; MA, Stanford University; and PhD, University of Wisconsin–Madison)

---

c h a p t e r 2

CONCEPT EXPLICATION AND THEORY CONSTRUCTION

Jack M. McLeod
Zhongdang Pan

With other former teaching and project assistants of the senior author

A good part of the work called "theorizing" is taken up with the clarification of concepts—and rightly so.

—Robert Merton (1958, p. 114)

Clear thinking requires clear language. In turn, a clear language requires that its terms be explicitly defined.

—Giovanni Sartori (1984, p. 22)

Making a concept explicit is, in a broad sense, a purpose of all discourses on communication concepts.

—Steven H. Chaffee (1991, p. vii)

Research concepts are the terms that we use to direct our observations and analyses in research. They are the building blocks of our knowledge, and they "constitute, in some meaningful sense, *units of thinking*" (Sartori, 1984, p. 17). Without them, we could not build scientific knowledge that is generalizable from observations in specific situations.

Concepts play a vital role in our everyday lives. They are the terms or labels we use to organize our everyday experiences and to communicate such experiences to others. Classification and comparison through the use of concepts are essential to dealing with the world around us. Fortunately, we live in a world of concepts; we learn thousands of concepts through socialization and education during our lifetime. These are nouns, adjectives, verbs, and adverbs contained in our vocabularies. For example, the sentence "A competent journalist writes clearly" has four concepts: "journalist," "competence," "writing," and "clarity." If a person reading the sentence misunderstands any of these concepts, the meaning of the sentence may be lost or distorted.

Although concepts can be classified as nouns, adjectives, and so on, we should not equate a concept and a linguistic term. A linguistic term is the smallest unit in a composition or text. It is a building block for larger linguistic units, such as a phrase, a sentence, or a paragraph. In the previous example, "A competent journalist writes clearly," there are five linguistic terms, but only four concepts. The linguistic terms that are not also concepts, such as articles and prepositions, are important parts of a text and contribute to the expression of concepts and their connections.

A *concept* is a basic unit of meaning in our conception or thought. It can be a single word or a phrase. As an example, let's consider the theoretical statement: "Television violence viewing is positively related to aggressive behavior."

This statement has nine linguistic terms but only three concepts: "television violence viewing," "positive relationship," and "aggressive behavior." Such a distinction is recognized by analyzing the relationships between concepts and their *referents*. Every concept has its real-world referent, which could be a subject or object, such as "journalist"; or a type of behavior, such as "aggressive behavior," or "television violence viewing"; a quality, such as "competence"; a type of relationship, such as "positively related"; or any other real-world occurrences or phenomena. Table 2.1 shows the examples of concepts discussed so far, the kind of things that are denoted as their referents, and their denotative boundaries.

Concepts vary in how clearly they are linked to real-world referents as well as in how broadly their referents spread in the real world. Some concepts, because their linkages to empirical referents are so obscure, do not lend themselves to easy usage in research. One remedy is to use multiple linguistic units to denote a concept, hoping that such a phrase would mark

TABLE 2.1. Examples of Concepts and Their Referents

CONCEPT	TYPE OF REFERENT	AS DISTINCT FROM
Journalist	Subject or object	Non-journalist
Aggressive behavior	Type of behavior	Nonaggressive behavior, behavior in general
TV violence viewing	Type of behavior	Nonviolent TV viewing, TV viewing in general
Competence	Personal quality	Incompetence
Positively related	A specific pattern of association	Negative relationship, unrelated to

clear boundaries of the concept that we wish to express. For example, by using "television violence viewing," we are saying that, conceptually, what is of concern is viewing of a *particular content of TV* programming, not simply undifferentiated viewing of television.

Communication is made easier by the knowledge that others in our culture share a large portion of our vocabulary and that they are likely to have some idea of what we mean when we use a particular concept. In everyday life, the fact that a concept is embedded in a sentence or a set of sentences makes it possible for us to function fairly well without spending much time and effort in trying to define the particular concept. Rough understanding is good enough in most social situations, and the particular context of the interaction tends to clarify the meaning of a concept.

There are times, however, when we need more accurate communication. For example, we need precise directions to find a street in an unfamiliar city or to take a prescribed medicine. We may also feel the need for more careful definition when we converse with others about an abstract concept that we care about a great deal. Misunderstanding and confusion are familiar negative experiences of life. They occur in part because we don't spend sufficient time defining things either for ourselves or for others. When left without explicit definition, a given concept may float in our communication with widely varied and private meanings for different people. This poses a serious obstacle in research. As a result, we have developed ways of making concepts more precise, and accurate communication more likely.

Singleton and Straits (1999), in their widely used text on sociological research methods, defined concepts as: "abstractions communicated by words or other signs that refer to common properties among phenomena" (p. 554). This definition sheds some light on our understanding of concepts because it suggests that concepts are abstract, that they each identify some general property or dimension that is common to the phenomena or objects being examined, that they are represented (or expressed) by words or other signs, and that they are communicated. It provides a starting point for us to understand how a concept is created, defined, and used in scientific research, a process called *concept explication.* In learning how to conduct research, we need to train ourselves in handling concepts in ways that will help us achieve clarity in thinking, precision in using words, accuracy in observations, and consistency in linking the three. Concept explication is a process that involves both logical and empirical procedures, which we call *meaning analysis* and *empirical analysis*, respectively (Hage, 1972; Hempel, 1952). In this chapter, we discuss the two procedures in introductory terms, focusing explicitly on the nuts and bolts of research. Throughout our discussion, we use familiar concepts and ideas in media effects research as illustrations.

MEANING ANALYSIS

Types of Concepts

It is useful to start by recognizing different types of concepts and knowing their varying utilities in research and communication.

First, there are *singular concepts*, which denote particular objects—persons (e.g., Shaquille O'Neal, George W. Bush), places (e.g., Indiana, Peoria), things (e.g., my brown cup, your dining table), or social entities (e.g., UW–Madison, Disney). In scientific research, singular concepts represent the objects we study (or observe, interview, etc.). These objects considered collectively are potentially our units of analysis or observation.

A second type, *class concepts*, denotes collections of such objects based on some common attributes (e.g., politicians, cities, universities, individuals, television networks, and so on). Class concepts are used to categorize the units of analysis of our research and define the population from which our sample is drawn. They may be quite broad (e.g., individuals, newspapers) or more specific (e.g., eligible voters in Dane County, Wisconsin; daily newspapers).

Third, *relational concepts* make connections between concepts (e.g., equal to, greater than, associated with, caused by, part of). They are important in research because they state how our variable concepts are related (e.g., Frequency of newspaper reading *is related to* knowledge about current economic conditions; newspapers *have more* news on local government *than* local television news shows). Relational concepts thus enable the statement of theoretical propositions or empirical hypotheses.

Fourth, there are *variable concepts*—we call them simply variables for short—that distinguish among objects within a class by evaluating them as to their possession of some underlying attribute: for example, competence (of journalists), clarity (of writing samples), size (of universities), levels of political knowledge (of citizens), amount or frequency of television violence viewing (of children), and so on. Variables are crucial to scientific research in that they describe important ways in which our units of analysis differ. We discuss them in more detail later.

Fifth, we *recognize process* concepts that depict specific ways in which things change or vary. For example, *priming* is a concept describing how an external stimulus may have an impact on some conceptual nodes in one's memory (Jo & Berkowitz, 1994). *Imitation* is another such concept in that it characterizes how individuals learn to perform certain behaviors (Bandura, 1977). Similarly, *institutionalization* is a process concept depicting how certain ways of doing things get stabilized, routinized, and formalized (Giddens, 1984). And *communication*, the concept that we use to denote our field of research, is such a concept in its own right because it depicts a multifaceted process (Lasswell, 1948). These concepts are critically important in research. Although the processes they depict may not be directly tested, hypotheses about the unique and specific outcomes derived from a theorized process may be tested. For example, although we cannot test the actual process of priming, psychologists have succeeded in showing faster reaction times in responding to stimuli that are semantically related to the idea that has just been activated (Wyer & Srull, 1989).

Finally, there are *mega-concepts*. These are conceptual conglomerates each of which is entangled with a number of often unspecified concepts: for example, *nice* person (regarding a particular person), *development* (of a nation), *democracy* (in the United States and other nations). Because of their compounded nature, mega-concepts pose a potential communication problem if the receiver (listener or reader) is unsure about or misunderstands the set of concepts intended by the sender (e.g., "nice person" may represent a summary over many variables, the composition of which may vary for different people). Process concepts are similar in having a compounded nature, making them a special subcategory of mega-concepts. In addition, many of the mega-concepts are emotionally held and value-laden, as the concepts of "nice person" and "democracy" suggest, adding to the difficulties in making them amenable to calm, rational consideration.

Complex concepts such as mega-concepts and process concepts present special challenges to researchers. They must be unpacked if they are to be accessible for research. This is the case for all complex concepts, but especially so for mega-concepts and process concepts. How can a researcher break an ambiguous and/or complex concept down to less abstract and empirically measurable variables or dimensions that can be investigated and communicated? The answer is through the process of *meaning analysis*, which is examined in this chapter.

Scientific Concepts and Everyday Concepts

Scientific concepts differ from our everyday concepts largely in terms of how we develop and use them. They differ more in degree in these ways than in kind. Therefore, we could not easily pick out a list of concept labels and sort them neatly into boxes marked "scientific" and "everyday," respectively. Only by identifying the definition behind a label and its connection to observable or measurable referents in the "real world" can we make a judgment as to its scientific quality.

Perhaps the best way to judge how well a concept meets scientific standards is to think of a set of criteria, each varying in degree (say, from 1 to 10) from nonscientific to scientific. The first such criterion is *abstractness*. To have any explanatory value, a concept in scientific research must be more than just a description of an attribute of a singular object at a given point in time. It should be abstract in the sense that many objects would differ and hence could be compared on this variable concept. In another sense, a concept should be sufficiently abstract that it could be observed in many places and times using varying methods of observation. However, we are hesitant to advocate maximum abstraction because abstractness operates in tension with other criteria for concepts.

Clarity of meaning is a second criterion of scientific concepts. The "technical" language of the researcher regarding what he or she means to convey in using the concept must be sufficiently clear. Because of the limits of verbal language, perfect clarity is never possible. The techniques of concept explication that we discuss later are useful to improve clarity of meaning.

Operationalizability is a long word describing the degree to which the verbal or conceptual definition of a concept can be translated into ways of observing the concept among the units of analysis. It involves stipulating in detail how values (e.g., numbers) are to be assigned to the units of analysis on the basis of some clearly specified observational procedures. This is a process that we call *measurement*. Observation is used here in a broad sense to include such procedures as participant observation, self-reports, coding of texts or other content units in content analysis, experimental manipulation, and so on. Unfortunately, there is some tension between operationalization and abstractness. The more abstract the concept, the more difficult it is to operationalize. Easily operationalized concepts are apt to lack abstraction. Again, the techniques of explication are used. The apparent dilemma can be overcome. If used well, explication allows us to be beneficially abstract and yet at the same time have well-measured concepts.

A final standard for scientific concepts is *precision*. It involves the degree to which the meaning of a concept and its theoretical and empirical relationships to other concepts can be communicated to other scientists and to nonscientists (e.g., policymakers, general public). That is, if a concept is to be useful to others, it must be understood by others. Precision is broader than the other standards in that unless a concept is sufficiently abstract, clear in meaning, and valid in its operationalization, it cannot be precise in communication with others. Others must know not only what meaning was intended but also how and how well the operational measures fit the conceptual definition. Replication by other researchers, so crucial to the accumulation of scientific knowledge, requires that concepts have clarity. Precision is also essential to making valid inferences from research evidence.

These are high standards to meet. Most research concepts only partially meet these standards. Most are improvements over everyday concepts on these dimensions, but most could also benefit from more careful explication.

To illustrate what we have said about concepts up to now, consider the following assertion: "Television is harmful to the political process."

This could be a sanitized version of an everyday conversation at a cocktail party. But statements of this kind appear frequently in books and magazine articles, as well as in political commentaries criticizing the media. We could regard them simply as editorial or polemical arguments, but we could also treat them as serious *theoretical statements* capable of being empirically tested in research. If we do take a statement seriously, we must evaluate the concepts it uses.

What are the concepts found in the statement? "Television" is meant obviously to be a variable concept, although in what sense it is so needs to be spelled out. The second term, "harmful," is even less clear. Presumably, it has to do with the functioning of the political system, say, *effectiveness of the political process*. Before evaluating these variable concepts, however, let's think about what other types of concepts are present or implied by the statement.

First, what singular or class concepts are involved here? That is, what units of analysis are being "harmed"? Where do we look to assess the "harm"? Is the author saying that television is affecting citizens adversely, or political actors (politicians or public officials), or political institutions, or some combination of these? Furthermore, what political system(s) is (are) included in the statement? Is it referring to political processes in the United States, a democracy, or to some set of political systems (various democratic systems), or to all political systems (e.g., totalitarian, authoritarian, and democratic systems)? We cannot go further in designing research until the units of analysis are specified (e.g., adult U.S. citizens, American presidential candidates, etc.).

Second, what *relational concepts* are involved? The phrase "is harmful" in the statement appears to set up a causal relationship or at least a *negative association* between television and the political process. Of course, we need more details about the conditions under which and ways in which such influences of television are manifested.

Let's return to the *variable concepts*. "Television," the source of the alleged negative effect, is really an example of a *mega-concept* until it is explicated fully. The reason is that the nature of the comparison is unclear (e.g., is it television vs. other media, or television content vs. the format of this medium?), and we do not know how it varies or whether a single variable or a complex of variables underlies it.

The concept "television" certainly meets our first standard of scientific concepts, *abstractness*. Unfortunately, it is overly abstract, a distinct characteristic of mega-concepts. It fares badly on our second standard, *clarity of meaning*, in not telling us what is meant by the term. For example, is it that the time people spend watching television diverts them from participating in the system? If so, then "television" would be defined in terms of time-budgeting or displacement of other activities. Or is it that viewing television entertainment content that portrays politicians in a negative light leads to political apathy among citizens? Or that watching television news content conveys mistrust of all politicians? Or is the effect not directly on the television viewers, but alternatively enacted through politicians' focusing on images rather than on issues when facing television cameras (e.g., President George W. Bush landing a fighter jet on an aircraft carrier before delivering his speech to declare that major battles in Iraq were over), or through television reporters asking "horseracing" rather than substantive questions (Patterson, 1993)? All of these five alternative interpretations of "television" appear to be operationalizable; we can imagine reasonable ways of measuring each of them. But we need to decide on which dimension of "television" fits our "theory" before going on to evaluate the quality of measurement. (We return to this point later. Here, we want to impress in your mind the point that a concept must be explicated to fit a theory or to serve a purpose.) Obviously, many definition problems must be solved before we can consider the precision of the concept.

"Effectiveness of the political process," the second variable, also suffers from a lack of *clarity of meaning*. What criteria (dimensions or indicators) of political process effectiveness should we use? Voter turnout, campaigning strategies, political knowledge, and frequency of political discussion are possible and seemingly operationalizable criteria for assessment. But before moving to measurement, we need a theory of political process to guide us.

We should regard our examples, "television" and "effectiveness of the political process" as being prescientific; before they can be considered as acceptable scientific variable concepts they must be more fully explicated.

Practices of Defining Concepts

Sharing of meaning between sender (e.g., writer, speaker) and receiver (e.g., reader, listener) is the basis of all communication. The social purpose of defining a concept is to increase the amount of shared meaning between the person stating the definition and the others. Concepts are the vehicles for conveying complex ideas and for making sense. Researchers also have an additional responsibility to define concepts used so that other researchers can *replicate* and extend the research findings linked to those concepts. To go back to our earlier example, other research investigators must be able to understand precisely what "television" and "effectiveness of the political process" mean in order to assess whatever findings served as evidence for the assertion of their relationship.

Defining a term is a frequently occurring communicative activity in our everyday life. It happens whenever we try to answer "what do you mean" kinds of questions. From our everyday practices to scientific research, there is a wide range of techniques that people use to define a concept. We have summarized seven such techniques and their values and limitations in research in Table 2.2.

While going through the seven techniques, we can stop and think about the uses of each in our everyday life. By doing so, we can see that they are all common practices. Each of them embeds in it certain logic that scientific concept explication relies upon. But each is inadequate as a standalone approach to concepts when we assess each concept in terms of the four criteria of scientific concepts stated in the previous section. We can discuss the seven techniques briefly by grouping them into three broad categories based on their focus in operation.

The first category consists of the first four approaches in Table 2.2: examples, exclusion, comparing and contrasting subsets of units, and procedures. They have in common the focus on some empirical manifestations of a concept. A concept so defined bears a proximity to the empirical world and is therefore operationalizable, at least potentially. With the definition by procedure, it is even possible to achieve a high degree of empirical precision in measuring a concept. However, these approaches do not provide sufficient abstractness to a concept for researchers to follow when using the same concept under different conditions. The different conditions could mean different research settings, different study populations, different measurement procedures, or any combination of the three. In brief, the common problem is that none of these approaches lays down the basis for specifying durable (i.e., not accidental, momentary, or setting-specific) and consensual relationships between a concept and its empirical referents. As a result, *clarity in meaning* of a concept may also be lost; it certainly opens to questions when a concept is transported from one research setting to

TABLE 2.2. Some Common Approaches to Defining Concepts

APPROACHES	PROCEDURES	FOCI	EXAMPLES	USES	LIMITATIONS
1. By examples	Listing exemplars or objects that comprise the empirical domain of a concept	Empirical constituents of a concept	"What do you mean by 'mass media'?" "Well, I mean things like TV, newspapers, magazines, and so on."	Making an abstract idea concrete and familiar and enabling communication at a lower level of abstraction	Inefficient and imprecise because it does not specify the essential meaning of a concept
2. By exclusion	Listing what are not parts of the empirical domain of a concept	Empirical boundaries between concepts	"What is 'media use'?" "It is not the same as face-to-face conversation, making a speech to a group, or gardening."	Making it possible to recognize distinctions among empirical phenomena and objects and to classify them	Same as above. It may cause confusion as empirical bases for distinction shift under changed conditions.
3. By comparing and contrasting subsets of units	Listing how subsets of a population differ	Empirical distinction among subsets or subgroups of units	"What is 'opinion leader-ship'?" "It refers to those who are better informed and more frequently give advice to others."	Making us recognize variation among units of a population and develop categories for classification.	It may disguise the general variable concepts used for such comparisons and contrasts.
4. By procedures	Labeling a measure or describing the procedure of measurement	Procedure of actual measurement	"What is 'IQ'?" "It is one's score obtained in the standardized IQ test."	Making it possible for us to refine and improve empirical measures of a concept and to observe empirical phenomena in areas where theory is lacking	It may lead to reification and/or atheoretical tendency in research.

APPROACHES	PROCEDURES	FOCI	EXAMPLES	USES	LIMITATIONS
5. By drawing analogies	Linking a concept to familiar ideas or vivid images	Similarities between concepts or ideas	"What is 'frame'?" "It is a mental framework similar to 'window to the world' or 'picture frame'."	Making it possible for us to know new phenomena, to see connections among different phenomena, and to describe the unfamiliar with familiar terms	It may be imprecise and limiting, as the familiar terms or mental images impede the required mental elasticity in concept explication.
6. By identifying function	Inferring the effects of a concept	Causal functions of the concept	"What is 'fear appeal'?" "It is an attribute of a message that arouses aversive emotional reactions among audiences."	Making it easier for us to position a concept in some causal propositions	Inefficient because changed conditions may call for different casual funtions. It introduces uncertainties as it mixes defining concepts and specifying causal func-tions among them.
7. By identifying antecedents	Recognizing the causes of a concept	Causal functions of some other concepts	"What is 'cognitive dissonance'?" "It is an internal tension resulted from having conflicting cognitions."	Same as above	Same as above

another. Researchers may use the same label without the shared meanings of the concept that the label is designed to represent.

To see how this may be the case, we can think about the literature on "TV attention." Without careful explication, this literature would appear to be a collection of distinct exemplars or empirical procedures (e.g., pupil fixation on TV screen, brainwave fluctuations, performance of a secondary task, and self-reported levels of "attention to" a type of media content or closeness in "following" a particular story). These empirical procedures or exemplars seem to have little in common (e.g., Anderson & Burns, 1991; Chaffee & Schleuder, 1986; Lang, 1990; Reeves et al., 1985) unless they are linked to and prescribed by a careful explication of the concept that is purported to manifest in these ways under different settings.

The second category has one approach, definition by analogies, the fifth approach in Table 2.2. This approach involves linking a concept to one or more familiar and/or vivid concepts or mental images. Although drawing such a linkage does help illuminate the meaning of the concept under concern, such illumination may be only partial without the specification of the essential qualities of the phenomena or process that the concept represents. Even more troubling is the tendency that the familiar or vivid concepts or mental images have in restricting our thinking by highlighting similarities while neglecting differences. Often, theoretical usefulness of a concept lies in its differences from existing ones. Moreover, analogy is a very imprecise approach to a concept. It does not lead to clarity in meaning or the most fruitful operationalization of the concept.

This might be a lesson that we are beginning to learn from the literature of "framing" research. Entman (1993) called for unifying the "fractured paradigm." However, that continues to be an elusive goal. We can see vast differences among various conceptions (e.g., Cappella & Jamieson, 1997; Gamson, 1996; Iyengar, 1992; McCombs & Ghanem, 2001; Pan & Kosicki, 1993, 2001; Scheufele, 1999; Snow, Rochford, Worden, & Benford, 1986). Analogous thinking is observable in all these conceptions—a frame is thought to be an organizing scheme similar to "a window on the world" or "a photographic frame" that brackets the reality in a particular way (Tuchman, 1978). Although helpful, such analogous thinking has not advanced us very far toward recognizing how framing involves (a) configuring symbolic resources in accordance with cultural conventions of interpretation, (b) specifying the key premise for mental processing, and (c) prescribing practices of interactions to steer the establishment of shared meanings toward some specific directions (Goffman, 1974; Reese, 2001).

The third category consists of the last two approaches in Table 2.2 that involve specifying causal functions in defining a concept. These are common practices both in our everyday life and in scientific research. For example, in our everyday life, we may characterize a scene from a movie as

"terrifying," based not so much on the analysis of the key elements of representation and their configuration in this scene but on our emotional reactions to the scene. The same logic is often employed in defining the "fear appeal" of a persuasive message in terms of message recipients' reactions rather than some intrinsic properties of the message (see Petty & Cacioppo, 1981). Such approaches become particularly useful when the introduction of a construct would enable us to build a coherent system of explanation called *theory*. In such a case, a concept is assumed to be *latent*, or not directly observable. We see such a practice at work when reading about how Isaac Newton formulated his idea of the universal gravity (Mason, 1962).

We can see the same practice at work when the concept of "attitude" was invoked to understand the rationale and functions of individual behavior (Pratkanis, Breckler, & Greenwald, 1989). A potential danger with these approaches is not paying enough attention to specifying the properties of the phenomena that the concept is designed to represent. When that happens, we may commit the *fallacy of reification*, or the logical error of attributing concrete observations to an empty conceptual entity or equating a functional relationship in some specific situations with generalizable conceptual abstraction. In either case, the clarity in meaning, precision, and abstractness of a concept would be lost. Such a fallacy can be avoided only if one has developed the functional relationships of a concept with its causal and effect variables based on the clearly specified meanings of the concept.

From this brief discussion of various approaches to defining a concept, we can see that an adequate handling of a concept must strive for specifying its essential meanings and principles for establishing stable linkages to its empirical referents (Hempel, 1952). This would be the eighth approach to defining concepts, which we call *concept explication*. It requires a researcher to (a) analyze the meaning of a concept label (a term or phrase), (b) describe the essential qualities of the concept, (c) when necessary (i.e., when a concept is relatively complex and abstract), identify its key dimensions, and (d) prescribe ways in which the concept and its dimensions may be connected to the real world. It leads to the formation of both *conceptual* and *operational definitions* of the concept. Both deductive and inductive reasoning are needed. It follows deductive reasoning in that such a definition points to ways to identify and locate the empirical referents of the concept. It involves inductive reasoning because it requires a researcher to extract commonalities from a wide range of objects or phenomena that the concept is designed to represent.

In essence, concept explication strives to build a "durable" or "universally recognized" relationship between a mental image or idea and its real-world referents. With that, researchers operating in varied research settings may invent different measurement techniques of a concept based on a shared understanding of what the concept means and how it is manifested in the empirical world.

This approach to concepts incorporates logical strengths of the other approaches. It uses the techniques from the other seven approaches listed in Table 2.2. All of them, despite their deficiencies, have their roles to play in our efforts toward explicating a concept. They could help us to start our long journey of concept explication, to illustrate the properties of a concept and its boundaries, to make scholarly writings more accessible, and to enable us to cultivate new areas or domains of research. Concept explication in scientific research is built on these logical strengths and delineates general principles for identifying the dimensions and indicators of any given concept under varying situations. It develops a common conceptual basis and vocabulary to develop generalizable scientific knowledge. We now turn to how to do such an explication.

The Logic of Concept Explication

Concept explication is the process by which an abstract concept is linked systematically to observed variations in the "real world" represented by the concept. This is accomplished by carefully developing two types of definitions: conceptual definitions, verbal descriptions of the essential properties that the researcher intends to be included within the concept's meaning; and operational definitions, procedures by which the concept is to be observed (as in participant observation), measured (as in sample surveys), or manipulated (as in experiments). For example, consider the concept "television violence viewing." Its *conceptual definition* might be something like "the amount of exposure to depictions of TV characters using physical force or verbal insults against other characters." Its *operational definition* might be the sum of a person's self-reported frequency (every week, most weeks, once or twice a month, never) of watching the entertainment television shows found to contain high levels of violent, antisocial content. Later, we discuss problems with this operational definition. For now, we focus on the difference between the two types of definitions.

Two sets of procedures are involved in concept explication: *meaning analysis,* in which we employ logical procedures to analyze a concept so that we can clearly link its conceptual and operational definitions; and *empirical analysis,* in which we evaluate the explication of the concept on the basis of empirical evidence. For now, we concentrate on meaning analysis. In the second part of this chapter, we discuss empirical analysis.

Working out the conceptual and operational definitions of a concept is the goal of meaning analysis. For that, as Chaffee (1991) advised, we need to take the following steps, although they are not necessarily taken in the order listed:

1. *Identify the concept.* When developing a concept to describe the objects of our study, the first question we need to ask is whether this is a variable concept (i.e., does it vary among the units of analysis?). If it is not a variable, how could it be made into one (e.g., transforming the imprecise label "television" into more precise "exposure to television violence")? Other early questions: What is the unit of analysis (e.g., individual, a group, a community, a 30-minute TV show, a newspaper article, etc.)? What is the purpose for using the variable? How does the concept fit into or help us to specify the research problem? A concept must be developed for a purpose, which is to develop a system of propositions capable of explaining a finite amount of phenomena. Identifying a concept involves much more than assigning a name or label to a mental image of ours (Hempel, 1952). It is a logical analysis of what we intend to investigate and explain.
2. *Search the literature.* Use various library resources to locate how previous researchers have conceptualized the topic of your study and used the concept that you have in mind. You need to review (i.e., to organize and structure) the literature. Some questions to be asked now are: What are the different conceptual meanings that have been assigned to the concept? What have been the research purposes of each? What operational definitions have been used? What are the different labels under which the concept has been studied? In view of the purpose of the study at hand, which of the various usages of the concept is most promising?
3. *Examine empirical properties.* Examine the properties of operational definitions of the selected concept, including their central tendencies, range of variation, shape of their distribution, and so on, either across the units of analysis studied, if the concept under concern depicts cross-sectional variations in a population (e.g., levels of education, age, hours per day watching TV), or across time, if the concept is purported to depict temporal variations (e.g., increase in knowledge, decay of physiological arousal), or both (e.g., differential rates of learning among a sample over a period of time). A variable that does not have sufficient variance is incapable of revealing the predicted co-variations with other variables, a point we return to in the discussion on empirical analysis. Also to be noted are the concept's antecedents and effects shown in previous studies. This examination should provide a better feel of how the variable is to be operationalized and how fruitful such an operational def-

inition (e.g., where and how systematic variations take place) would be in your intended research.

4. *Develop a tentative conceptual definition.* Based on the review and analysis presented here, try to develop a conceptual definition of the concept—one that would fit your theoretical framework. In doing so, you need to address a series of questions. For your research purposes, what does the concept label mean conceptually? What are its essential elements? What lower order (less abstract) *dimensions* does this general concept subsume? What are instances or examples that could be developed into indicators of the concept? Would these indicators be expected to correlate highly or moderately? If not, would it suggest further complexity of the concept and the presence of more differentiable dimensions? (For more discussion of *dimensions,* see later.) Much trial and error and revision of your definition are involved here. From the conceptual definition, we should be able to understand the boundaries of the concept (what is excluded from it) and the conditions for observation. Concepts are defined by less abstract and more familiar terms, and ultimately by *primitive terms,* words whose meanings are widely shared and incapable of further definition (e.g., individual, frequency, exposure) except by using synonyms (e.g., person, how often, viewing).
5. *Define the concept operationally.* Each dimension of the concept should be capable of direct observation or creation in the "real world" of experience. If possible, each dimension would have more than one empirical *indicator* (items or instances). Ideally, each would be capable of observation with more than one method (e.g., observer rating, self-report measurement, experimental manipulation). The operational definition should include: the conditions of observation, details of instrument construction (e.g., question wording, rating scale, equipment setup), and manipulation procedures, as well as analysis procedures and statistical operations implied.
6. *Data gathering.* The concept and its operational definition should then be included in actual data collection. The data will be used in the second phase of the explication process, empirical analysis.

It should be clear by now that meaning analysis runs through the whole process, from identifying a research topic to the onset of the actual data gathering. As the first of the two-part process of explication, it involves primarily (but not exclusively) moving downward in abstraction

from conceptual definitions to operational definitions, from an abstract conceptual level to a more concrete empirical level. In essence, meaning analysis is part of the process of building sound linkages between conceptual and empirical worlds. Building and verifying such linkages is at the core of scientific research.

Often, a complex and/or abstract concept poses difficulties in linking it with an operational definition. As a result, we need to specify its less abstract constituting parts, which we call *dimensions.* These are subconcepts that capture some particular aspects of the complex concept, as required by the logic of a theory of which the concept is part. For example, for many research purposes, the concept of "television viewing" is too complex to be meaningfully related to most effect variables. Therefore, we may wish to break it down to two or more dimensions (e.g., "exposure to TV news," "exposure to TV entertainment shows"). Different models of dimensionalization of TV viewing may be appropriate for different theoretical purposes. By specifying such dimensions, which must be done in some specific theoretical framework, we also enrich our understanding of TV viewing by studying its multiple aspects and potentially diverse effects, as they are specified in different theories.

The building of complex research concepts through specifying their dimensions is one reason that scientific concepts are also called *constructs.* The other reason is that the researcher intentionally builds concepts with specially defined meanings rather than depending on everyday concepts. Developing such specialized language, or being able to express exactly what a scientific theory is designed to say, is a necessary condition of science (Sartori, 1984). For our purposes, research *concepts* and *constructs* are considered equivalent terms.

Dimensions are less abstract than the concept to which they contribute, but they are more abstract than the concrete *indicators* that measure them. For example, the indicators for "exposure to TV news" might be days per week watching the early evening national news, days per week watching the late evening local news, days per week watching the morning news, and so on. Similarly, "exposure to TV entertainment" may be measured by frequency of watching soap operas, crime and detective shows, and other entertainment shows on TV.

Figure 2.1 illustrates the three levels (concept, dimensions, and indicators) of the construct "television viewing."

By dividing "television viewing" into two dimensions, news and entertainment, the researcher is asserting several things. The first is that the distinction of content types matters in testing our research hypothesis. (Note that a different assertion and thus conceptual and operational definitions of TV viewing—the amount of TV viewing regardless of content watched—is the key in cultivation research!) The second assertion is that

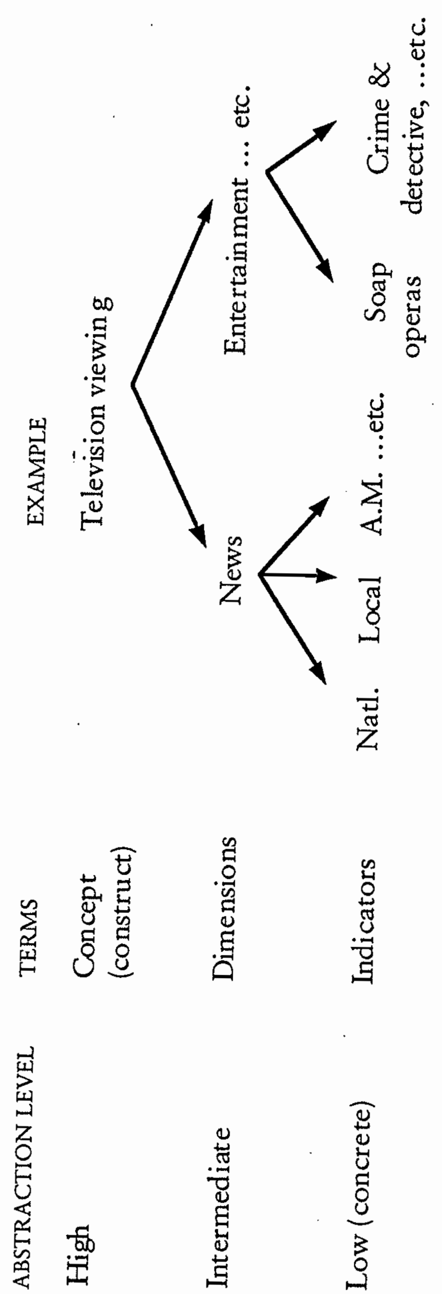

FIGURE 2.1. An illustration of the dimensions and indicators of TV viewing. Natl. = national evening news, Local = local news, A.M. = network morning news shows.

the indicators of news viewing tend to agree (i.e., those watching national news frequently will also tend to watch local news), the indicators of entertainment viewing tend to agree (i.e., those watching soap operas frequently also tend to watch more crime and detective shows), and that there will be less agreement across the dimensions than within each dimension (i.e., those who watch news frequently are less likely to watch soap operas with a similarly high frequency). These assertions become empirical hypotheses to be tested in the second stage of concept explication, called empirical analysis.

This example is meant to illustrate the relationships of concepts, dimensions, and indicators. Obviously, other dimensions and indicators might be specified depending on one's research problem. For example, TV entertainment viewing might be divided into watching situation comedies, quiz shows, and adventure-drama shows, if one is interested in learning different levels of viewer involvement while watching TV.

There are theoretical as well as practical reasons for dimensionalizing a concept. We have already said that making an abstract concept less abstract allows its meaning to be more precisely communicated and its measures more precisely determined. This is a practical reason. We also wish to stress theoretical advances that may result from specifying dimensions of a concept. That is, different dimensions of a concept may be related to different causes and effects. By specifying dimensions of a concept, we can test a theory more precisely. For example, it has been shown reliably that different demographic variables help us to predict frequencies of TV news viewing and TV entertainment viewing. Although TV news viewing is found to be related to some gains in knowledge on public affairs and politics and even some kinds of participation in the political process, TV entertainment viewing is found to be related negatively to these effect variables. It is more likely to be related to overestimation of crime rates and to fear of becoming victims of crimes. Such empirical findings are predicated on dimensionalizing TV viewing; they demonstrate the need to theorize effects of TV viewing in more differentiated terms than treating television as an undifferentiated singular entity.

Concept Explication and Theory Construction

Understanding how research concepts are defined is as important to a research consumer (e.g., student, media practitioner, decision maker) as it is to a researcher. A common problem in our everyday communication and even in scientific research is that we fail to define a concept clearly. This problem occurs for various reasons under various circumstances. Sometimes, a concept is so common that we see no need to define it.

Sometimes, users of a concept are so vague about its boundaries that they make sweeping claims without being aware that they have gone way beyond what the concept captures. There are still other times when users of a concept neglect to define it conceptually and simply use its label to represent a measurement procedure (e.g., when being asked what is meant by the concept "IQ," a researcher replies, "IQ is what my IQ test measures—no more and no less"). This lack of attention to conceptual meaning of operational measures is rampant in the applied research conducted by various commercial and policymaking agencies.

Concept explication is a central task in theory building. It also must be conducted in the context of a theory and serve the purpose of carrying out some empirical tests called for by the theory. This claim can be understood in two ways. First, explication, by linking conceptual and operational definitions of a concept, makes a theory testable or at least more amenable to testing. Second, explication advances or improves a theory in that it makes a theoretical proposition more precise and may even lead to construction of a new theory.

But what is *theory*? Perhaps the most useful single sentence definition is as follows: "Theory is a set of organized propositions that provide an explanation for some recurrent phenomena of research interest." But this highly condensed statement needs further elaboration. The term *explanation* implies that theory should convey a *sense of understanding* to the person reading about the theory. That is, the consumer or user of the theory should feel "Okay, I see the point and it makes sense." The term *organized* implies that the *reasoning* behind the explanation is stated in a logically deductive and/or inductive fashion. In brief, theory is more than a predictive statement or "hunch." It tells us *why* the statement is made and what logical assumptions and arguments it is based on.

The term *recurrent phenomena* implies what is being explained is sufficiently abstract to recur (i.e., to be observable in similar form in other times and places). Phenomena cannot recur in total detail (no two events are identical), but *conceptually similar* phenomena can be seen as repeating if defined with adequate abstraction. Theory thus differs from pure description in that description may be specific in detailing a set of actions that have happened in a particular time and place while theory states under what conditions conceptually similar phenomena will or might recur.

It is in this sense that a good theory has predictive power. Although *prediction* states the likelihood that an event (or phenomenon) will occur under some specific conditions, *explanation* states *why* and *how* a conceptually similar type of phenomenon occurs. A good explanation leads to specific predictions, but a prediction is not necessarily predicated on explanation.

To bring *theory* closer to the actual conduct of research, let's define it by using the concept types presented earlier. Defined in this way, theory is a collection of interrelated statements asserting how and why two or more variable concepts are related as they vary among a class of objects. The "class of objects" is comprised of the *units of analysis* whose variation we are trying to explain. Examples of units of analysis are: individuals, media organizations, nations, families, newspapers, and so on. The "two or more variable concepts" represent variables—the characteristics on which the units of analysis vary. Examples of the characteristics on which individuals vary are: gender, frequency of television news exposure, political participation, and the like. Examples of the characteristics on which newspaper articles vary are: length, position, number of times affect-laden adjectives are being used, and so on. The relationship between variables may be described by some phrase denoting a relational concept such as "is positively associated with," "is caused by," and so on. Other elements of a theory include statements that function as premises, such as basic assumptions or postulates that serve as a logical basis for us to build the relational statement between variables and narratives of how the stated relationship takes place in the real world. Consider the following theoretical statement: "Exposure to violent television content makes it more likely that adolescent children will display antisocial, aggressive behavior."

"Adolescent children" is a label of a class concept, and the individuals belonging to this category are the units of analysis. It is their behavior, more specifically, "antisocial, aggressive behavior," that we are interested in explaining. "Antisocial, aggressive behavior," then, is considered an effect of TV viewing, or a *dependent variable.* It is a variable concept, referring to the frequency or intensity of engaging in antisocial behavior. "Exposure to violent television" in this statement is the causal, or *independent variable.* It is based on a simple postulate that adolescent children vary in the amount of TV violence that they watch. The relational concept, expressed in the phrase "makes it more likely," links the independent and dependent variables and suggests a causal and positive relationship between the two (the more TV violence an adolescent child watches, the more likely that child will display aggressive behavior).

This example involves four concepts (one class, one relational, and two variable concepts). The statement by itself is only a hypothesis, not a theory. To make it into a theory, we need a set of *premises,* or *assumptions,* that will provide a logical basis for the hypothesized relationship (i.e., telling us *why* and *how* watching televised violence and engaging in aggressive behavior may be related to each other). This logical basis is essential to developing a sense of understanding, the ultimate goal of developing a theory. For example, one set of premises may be that adolescents observe aggressive behavior depicted on TV and imitate it in their daily lives

(Bandura, 1977). Another set of premises may be that intense violent actions on TV heighten physical arousal among viewers, and people in a physically aroused condition are more likely to engage in aggressive behavior (Zillmann, 1971). The two sets of premises provide different reasons why television violence viewing and aggressive behavior may be related. Thus, both convey a sense of understanding. But they may not produce equally strong supportive evidence under various conditions. The choice between the two explanations has to rely on empirical evidence testing the specific predictions directly derived from one but not the other theory and the exact nature of the empirical phenomena under investigation. For us, the point is to show that by *theory*, we generally mean a whole collection of theoretical statements linked by concepts and assumptions.

To bring the statement in our example into research, both variable concepts need to be explicated. Conceptual definitions with clear linkages to operational definitions must be developed. It may be necessary to specify dimensions of each concept as part of the explication process. Various indicators could be used for operational definitions. For violent television viewing, for example, we could *observe* and rate adolescents watching television; we could *measure* it by asking them how frequently they watch various shows that we know to contain violent acts and scenes; or we might experimentally *manipulate* violent viewing by showing a violent show to a group of adolescent children (experimental group) and a nonviolent show to another group of adolescent children (control group).

Once adequate indicators of these concepts have been developed, a *research hypothesis* can be stated that links the operational definitions of the two concepts. The research hypothesis must be parallel with but more specific than the more abstract theoretical statement. Presumably, many research hypotheses involving various operational procedures and measures could be generated from the same theoretical statement. A given research hypothesis, therefore, is apt to be only a partial representation of a given theoretical statement and an even smaller portion of a given body of theory. Testing a given research hypothesis is thus only a partial and indirect test of a more general body of theory.

Theory is thus tested indirectly through research hypotheses and even more indirectly because a research hypothesis is tested through a corresponding *null hypothesis* (a statement asserting that there is *no*, i.e., null, relationship between the variables specified in the research hypothesis). That is, we gather empirical data to estimate the likelihood that the hypothesized relationship occurs only by chance. Notice that we would be comparing two competing theories in such an empirical test: One states that exposure to violent TV shows may, due to the reasons of imitation and/or arousal, increase the likelihood of aggressive behavior among adolescent children; the other asserts that whatever relationship between the specified exposure and behavior is due only to chance. The evidence ascertained in testing a hypothesis is based on the likelihood that the null hypothesis may be supported by the data. The evidence rejecting a null hypothesis based on the conventional significance level of $p \leq 0.05$ (i.e., the chance that the data support the null hypothesis is 1 out of 20 or smaller) does not "prove" the research hypothesis; rather, it only enhances our confidence in tentatively retaining it until contradictory evidence is obtained.

Figure 2.2 presents an illustration of how the theoretical statement may be developed into research and null hypotheses through the meaning analysis discussed here.

For the purpose of illustration, the researcher has stipulated two dimensions of the "exposure to violent television content" (exposure to adventure-drama shows and exposure to crime-detective shows) and two dimensions of "aggressive behavior" (overt physical acts and verbal abuse). This is clearly an incomplete list of possible dimensions. For each dimension, two indicators are shown, although many others are possible. The indicators for the two dimensions of exposure to violent shows might be a list of such shows with each adolescent respondent being asked to indicate how often he or she has watched it. The aggressive behavior indicators might be frequencies of adolescent children engaging in various physical acts and verbal abuse behaviors, as rated by the adolescent respondents themselves, by teachers, and/or by parents. Indices can be formed by combining these various indicators of each dimension, and their reliability and validity can be assessed—central tasks of the empirical analysis part of concept explication.

An empirical test involving the indicators shown in the example may yield results that partially support the broad theoretical statement. It may also provide clues as to alternative operational measures of the two variable concepts. In any event, what we have shown here is only one of many iterations between theory and empirical evidence. Conceptualization—a process of linking the conceptual and empirical worlds—is a continuous process and involves two-way traffic between them. Specific to concept explication, this means that each explication of a concept may inform us about how to gather observations in certain ways (operationalization), the empirical analysis of which in turn may demand that we modify or improve our explication for further research. Meaning analysis and empirical analysis, then, complement each other in concept explication. Figure 2.3 illustrates this iterative process.

In this illustration, concept explication is depicted as going through an iterative (repeated) process of informal observations, preliminary explication, formal observations, and further explication that involve both meaning and empirical analyses. The goal is to develop better conceptual and operational definitions with stronger linkages between them.

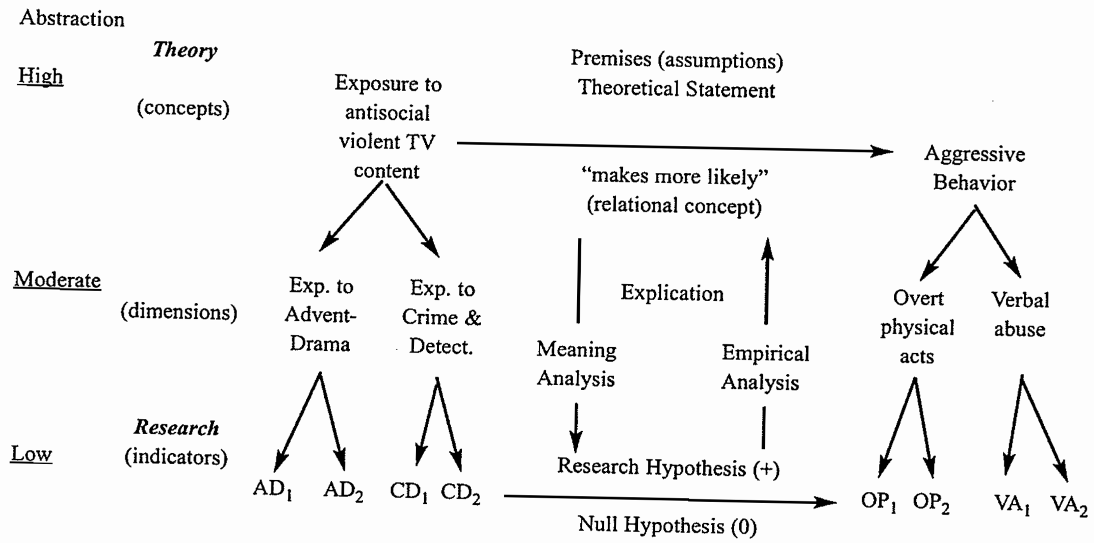

FIGURE 2.2. An illustration of the relationship between theory and research.
AD_1 and AD_2 refer to indicators 1 and 2 for frequencies of watching "adventure and drama shows" on TV; CD_1 and CD_2 refer to indicators 1 and 2 for frequencies of watching "crime and detective shows" on TV. OP_1 and OP_2 refer to indicators measuring frequencies of engaging in "overt physical acts" that are considered aggressive, and VA_1 and VA_2 refer to indicators of "verbal abuse."

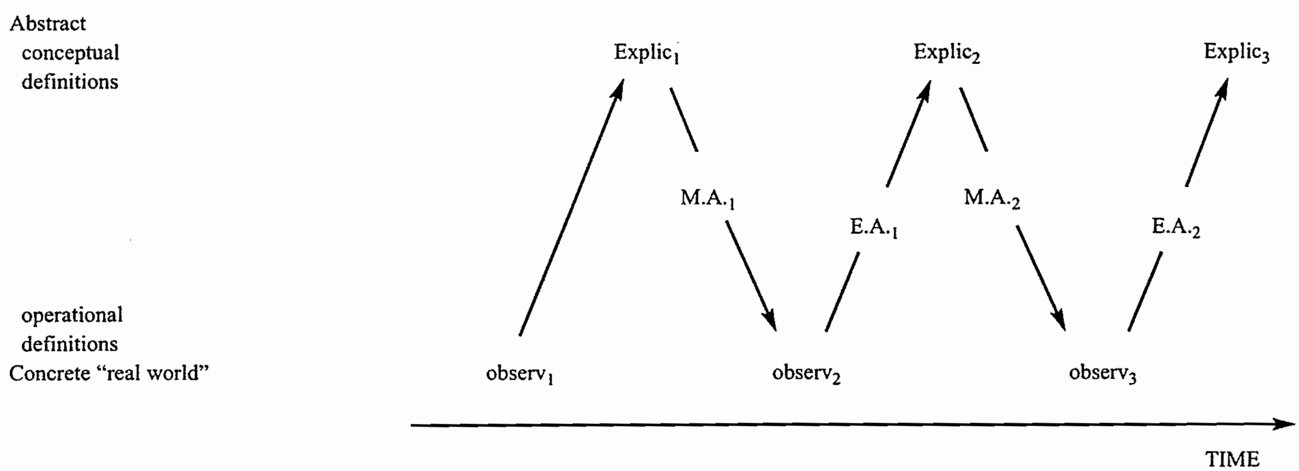

FIGURE 2.3. The interactive process of concept explication.
Explic = explication; observ = observation; M.A. = meaning analysis; E.A. = empirical analysis. (Adapted from Donohew & Palmgreen, 1989, p. 46.)

A complete explication of a complex research concept obviously takes a great deal of time and effort. As a result, it is unlikely that even the most diligent researchers will be able to fully explicate the many research concepts germane to a particular research problem. Fortunately, some concepts have been used so frequently in the past and/or are so concrete that the connection between their conceptual and operational definitions is well understood. "Level of education" of the respondent (number of years of formal schooling completed) and "age" are examples of concepts needing little further explication. Even so, what historical forces and life experiences are deposited in the variations of each of these variables may still demand careful theoretical considerations (Chaffee, 1991), and consequently, findings involving these concepts may still be open to varying interpretations. It would be a gross mistake not to explicate these seemingly "natural" concepts if one were to examine life experiences of people in different age cohorts and/or with experiences in some common defining events. Most often, the researcher's task is to explicate the focal concepts that are either less familiar or ambiguous in previous research or need to be defined specifically for a fresh look at some phenomena or hypothesis. Such efforts ought to be central tasks in research planning and fundamental for the analysis following data gathering.

EMPIRICAL ANALYSIS

The Cyclical Process of Concept Explication

Scientific research is a cyclical process alternating in focus between theory and empirical evidence (Wallace, 1971). A crucial part of it is *concept explication,* a cyclical process in its own right. When we start from a theoretical perspective or idea or concept, we work our way down from the more abstract conceptual definition to more concrete empirical indicators. The procedure was discussed earlier as *meaning analysis*.

Meaning analysis is limited to logical reasoning and by the creativity of a researcher-theorist and the quality of that researcher's assumptions. It takes us only so far toward providing valid answers to our research questions, but it also sets a stage for us to evaluate our logical reasoning with empirical data. By specifying indicators of a concept, meaning analysis leads to statements of some *within-concept* hypotheses about the relationships among these indicators and between the indicators and their conceptual counterpart. In addition, it also leads to a stipulation of some hypotheses about the relationships between the indicators of different concepts (between concepts). As illustrated in Figure 2.2, the *between-concept* hypotheses are derived from the corresponding theoretical statements. The *within-concept* hypotheses are developed in meaning analysis, based on related theories (auxiliary theories) as to how a concept may be represented by its empirical indicators. Both types of hypotheses must be supported by empirical data before any claim to valid knowledge can be made.

The process of examining the fit between the *within-concept* hypotheses and empirical data is called *empirical analysis.* The goal is to create empirical variables that are valid representations of the underlying construct based not only on the logical analysis done in meaning analysis but also on empirical evidence. Empirical analysis is thus an integral part of concept explication. In the ongoing process of concept explication, evidence obtained from empirical analysis may *inductively* lead to revising the conceptual and operational definitions of a concept or to developing a new theoretical construct.

Figure 2.4 is a schematic depiction of the complete conceptualization process that shows how meaning analysis and empirical analysis are related.

Two distinct but related tasks are involved in empirical analysis: (a) the development of indices or scales from the indicators of a concept and (b) the empirical evaluation of each index or scale and its constituent indicators.

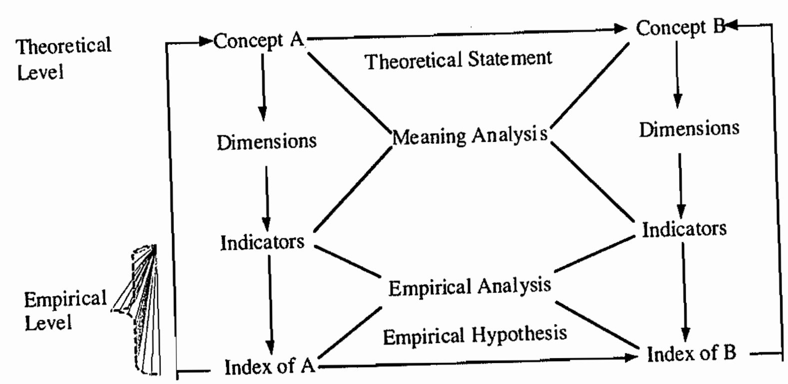

FIGURE 2.4. An illustration of a complete conceptualization process.

Constructing Indices and Scales

The first task of empirical analysis is to create a scale or index on which to locate the units of observation (e.g., individual respondents) by assigning each unit some value according to certain rules. We say "assigning values" instead of "assigning numbers" because values do not necessarily assume the natural numerical sequence; rather, they reflect the kind of theoretical assumptions about the nature of the characteristic being measured. For example, when we use "1" to stand for male and "2" for female, the two values do not stand for their additive nature as if they were in a natural number sequence. They simply reflect the categorical difference that we assume between gender categories.

A rule of simple addition or average of each respondent's values on each of the indicators used to measure the same concept generates the person's total *index score*. *Scale values* are produced based on more complex rules of combination or initial data gathering. This distinction between a *scale* and an *index* (see Babbie, 2004, for an introductory discussion) is less important than the understanding that they are both composite measures representing some particular concept or dimension.

Sometimes we might use only one indicator (e.g., a single questionnaire item) to measure a concept. This might happen if there is a "natural" dichotomy (e.g., male vs. female) or if it is unclear how more than one question might be phrased to measure one attribute (e.g., education in terms of years of formal schooling). We are accustomed to scales in the physical world that yield only one observation or indicator. For example, we measure length by applying a ruler of appropriate length already marked in inches or centimeters. Normally, a single thermometer with marks of Fahrenheit or Celsius is sufficient to record room temperature. Some physical concepts are complex, of course, and may require some specific rules of combining different observations (e.g., mass as quantity of matter measured by the weight of a body divided by the acceleration owing to gravity), but most of our direct experience is with relatively straightforward physical scales.

The concepts that most often concern us in social science (e.g., a person's opinions, behaviors, and perceptions) are complex and not so easy to capture with a single question or observation. Nor is it obvious what system metric (inches or centimeters?) can be used to assign values to each *unit of observation*. We are much better off, for most concepts, if we develop multiple *indicators* and combine them in some way into an index. Multiple indicators are preferred for several reasons:

1. If we have several indicators, it is more efficient (less time and space) to combine them into an index than to analyze and present them separately. Data reduction is a fundamental part of empirical analysis.
2. The random measurement errors found in each indicator tend to be canceled out when these indicators are appropriately combined into an index—(i.e., an index tends to be, presumably, more reliable than its constituent indicators).
3. Use of indices allows us to operate at a more abstract level of theorizing, thus making for more widely applicable theory and for more important policy implications.
4. The more indicators of a concept we have, the less likely we are to miss entirely or undersample a domain of a concept (i.e., multiple indicators enhance our confidence in content validity).
5. Having multiple indicators yields necessary data for us to assess the quality of the operational definition of a concept empirically, using the two basic criteria, reliability and validity.
6. And, multiple indicators help us to evaluate the meaning of the concept after data analysis (e.g., we may disaggregate a concept into its constituting indicators and then assess the relationships with the predicted antecedents or effects at this lower level of abstraction to see how internally coherent a concept may be) and to improve its redefinition and measurement.

In the physical world, the construction of a measuring device follows the laws of physical science and mathematics. However, such laws do not apply straightforwardly to the social and behavioral phenomena that we study. Nevertheless, in mass communication research, scale construction is also guided by theoretical and empirical work that has accumulated over the years. Several scaling methods developed in other disciplines are widely used, including Thurstone scales, Likert scales, Guttman scales, Semantic Differential scales, and so on. Because discussion of these scales can be found in other readily available sources (e.g., Babbie, 2004; Kerlinger & Lee, 1999; McIver & Carmines, 1981), we do not dwell on them in this chapter.

When going through the arguments for using indices, careful readers will sense tensions between specificity of indicators and abstractness of indices on the one hand and between reliability and validity as two criteria for assessing an index on the other. We already addressed the first tension in the discussion of the four criteria used to assess scientific concepts. It is essentially the tension between abstractness and the other three criteria: clarity in meaning, operationalizability, and precision. We may view the hierarchical system of moving from the most abstract level of a concept down to a less abstract level of its constituting dimensions and then further down to empirical indicators of each dimension (see Fig.2.4) as a logical and pragmatic way of dealing with the tension. Meaning analysis deals with the tension in part by working out the "vertical" linkages between an abstract concept and its empirical referents. When assessing indices in terms of reliability and validity, we encounter the same tension manifested in a different form (i.e., the tension between reliability and validity.) As we go through the example of empirical analysis, we discuss how to handle such a tension with empirical evidence.

Metrics, Scales, and Levels of Measurement

A scale, whether it involves a single indicator or an index, must have some kind of system for assigning values to the objects being measured. For example, to measure length, we can use the metric system (e.g., centimeter, meter, or kilometer) or the British/U.S. system (e.g., inch, yard, or mile). When we create a scale to measure some characteristic of persons (or other units of analysis in social sciences), we also need some kind of measurement system. Different systems reflect our specific assumptions about the nature of the characteristic we are measuring. In general, four levels of measurement are differentiated, from the lowest to highest: nominal, ordinal, interval, and ratio. They form a hierarchical system in that a measure at a higher level includes not only all the information in the measure immediately below it but also more. For example, an *interval scale* can be used not only for classification (a task that can be accomplished with a *nominal scale*) and rank ordering (a task must involve at least an *ordinal scale*), but also for assessing the distance between two adjacent attributes of a variable. A *ratio scale* can do all these and provide us with a true zero point. An example of a ratio scale is when we ask respondents how likely they are to vote in an election, using a scale from 0% (*definitely not*) to 100% (*definitely will*). Variables such as gender or race are assumed to have only qualitatively different attributes, represented by the different categories we use to code these variables (e.g., assigning "1" to male and "2" to female). They are measured on a nominal scale.

Two points need to be emphasized. First, in many situations in social science research, the response alternatives used in a given indicator or set of indicators may have a totally arbitrary metric. For example, we may use a 10-point scale (e.g., *little* to *very close*) to measure respondents' attention to various TV shows. A 10 or any other point on the scale has no meaning in the physical world except that it is higher than any number under it and, we assume, the interval between any pair of adjacent values on this scale is a constant. With this scale, we can examine the *relative* position of many respondents on the scale and see how such a position relates to their positions on other indicators and concepts. However, using such a scale to take a ratio of attention between two individuals would amount to a noninterpretable number.

Second, the choice among the four levels of measurement not only reflects our assumptions about the characteristic being measured but also limits our choice of analysis strategies. Some statistical techniques and tests are appropriate only to some particular levels of measurement because of the underlying assumptions made in such techniques or tests. For example, we would be violating some basic statistical assumptions if we used Pearson's correlation coefficient (r) to assess correlation between gender and race, both categorical variables measured on a nominal scale. Alternatively, we would be throwing away much data if we used X^2 to test the significance of the relationship in a contingency table involving levels of education and levels of attention paid to TV news when both are measured on an interval scale.

Empirical analysis should be thought of as a general strategy that embodies a set of broad principles rather than as a fixed set of procedures. Choices among the strategies are constrained by the level of measurement, research design, and measurement models employed, but they also reflect the preferences of the researcher to some extent. In the subsequent discussions, we will discuss the procedure of empirical analysis through one major example of a summated rating scale (a particular measurement model), the explication of the concept "public affairs media attention." The example draws data from a 1987 survey that interviewed a probability sample ($n = 480$) of adult residents in Dane County, Wisconsin. The indicators of attention all used the familiar 11-point scale (1 = *little attention*, 10 = *very close attention*, and nonusers of a given medium were scored as 0). Because these are continuous measures and the assumption of equal intervals is reasonable, we can use statistics (e.g., Pearson's r) and analytical techniques (e.g., linear regression) appropriate to interval measures.

Because we have multiple indicators for the concept under examination (three types of content across two media), we can examine reliability in terms of *homogeneity* (i.e., internal consistency). The approach basically involves assessing the level of agreement among separate indicators. The logic is the same as assessing intercoder reliability by calculating the level of agreement among independent coders. However, because we only have cross-sectional data (data from only one point in time rather than multiple time points as in a panel design), we are unable to examine *stability* as another form of reliability (Heise, 1969; Wiley & Wiley, 1970).

Within the constraints of measurement and design, empirical explication proceeds generally from reliability issues concentrating on the indicators of a concept and index building outward to the examination of validity, first in terms of discriminating the concept from other similar concepts and then as to its predicted antecedents and effects. Most of what we demonstrate is what Chaffee (1991, pp. 51–62) calls "univariate research" (i.e., involving only a single construct), even though some statistical techniques involved are multivariate in nature.

The Example: Public Affairs Media Attention

Let's say we are interested in the role of attention in enhancing or restricting mass media effects. More specifically, the focal concept is "public affairs media attention." Our conceptual definition is the following: "a conscious process of focusing concern on public affairs content ('hard news') during the exposure to news media." This means attention involves allocating mental resources to absorbing information and comprehending the reports of public affairs in news media.

Operationally in this particular example, "news media" is defined as newspapers and television news; news magazines and radio news, although appropriate, were excluded only due to space limits in the questionnaire. "Public affairs content" is operationally defined as including three categories: international affairs, national government and politics, and local government and politics. The levels of "focused concern" are measured by a 10-point response scale (1 = *little attention* and 10 = *very close attention*) in answering the question, "When you read the following kinds of newspaper stories (When you are watching television and the following kinds of stories appear), how much attention do you pay to them?" "A conscious process" is reflected in the fact that we use self-report (vs. unobtrusive observational or instrument-recorded) data.

Notice that we phrased these questions in ways that would avoid the potential problem of displacement among items (or the lack of independence in scoring each indicator). For example, watching a network's evening news is likely to displace watching another network's evening news because both networks have their evening newscasts scheduled at the same time. Consequently, if how frequently watching evening news on each network is asked in a survey, with two indicators of "TV news viewing" (e.g., watching evening newscast on ABC and watching evening newscast on CBS), then how one answers one question will be constrained by how the other was answered. Such a displacement relationship among indicators contradicts a basic premise of the analysis discussed in this chapter, that we are devising a summated rating scale to measure attention. A summated rating scale assumes that items used comprise a random subset of the universe of a concept and that they form a scale when individuals' responses to these items can be summed (or averaged). Based on this assumption, then, one subset of items from the universe of the underlying concept is interchangeable with another subset from the same universe, and thus there is an empirical homogeneity among the items from the universe. This qualification is very important because it limits the range of applicability of the logic and techniques discussed in this example.

Table 2.3 shows means and standard deviations of the responses on the six indicators. We can see that these items yield well-spread distributions

TABLE 2.3. Descriptive Statistics of Public Affairs Attention Items

ATTENTION TO	MEAN	STANDARD DEVIATION
1. Newspaper international news	6.18	2.67
2. Newspaper national news	6.06	2.64
3. Newspaper local news	5.95	2.50
4. Television international news	6.54	2.77
5. Television national news	6.24	2.66
6. Television local news	5.72	2.43

with means located only slightly above the midpoint of the scale. Both are preliminary indications that the questionnaire items worked pretty well in capturing variance and providing the basis for us to find out whether it is the right kind of variance.

We cannot emphasize enough the importance of inspecting descriptive statistics and the out-of-range values at the level of indicators *before* proceeding. Such inspection gives us some initial sense of how well each indicator might work and what are the features of its distribution when compared with the other indicators. Familiarity with the data at this level will help us identify the "bad apples" in the pool of indicators—items that, for example, simply do not yield much variance to differentiate units of analysis or have a distribution too peculiar to be usable in any analysis involving the multivariate normality assumption. It will also inform our decisions in subsequent analysis on, for example, whether an item should be excluded from entering an index due to its empirical performance despite apparent face validity—one of many decisions that we have to make in handling the tension between reliability and validity. For example, if any of the items in this group of six—all were measured on a 10-point scale—had a mean of 2 or 9 and standard deviation of 1 or smaller, it would likely be a poor indicator for us to measure public affairs media attention due to its skewed distribution and lack of variance. In other words, a seemingly valid item, such as a survey question, could be a bad indicator due to its empirical characteristics. Although dropping such an item may be a decision made on empirical grounds, it could also signal the need for more careful conceptualization in future research.

The conceptualization and operationalization of public affairs media attention is significant not only for what is included but also for what is *not* included. Assessing these six items and the attention concept that they are purported to be measuring, we also place them in a data matrix that involves other closely related measures: attention to different types of media content, more specifically, features about ordinary people and sports in newspaper and crime shows and sitcoms on TV, as well as exposure to each of these content categories. All together, the matrix involves 16 media-related items. Then, analyzing intercorrelation among these items, we may assess the following *within-concept* hypotheses, as they are summarized in Fig. 2.5:

1. *Content-specific attention hypothesis*. The homogeneity among the six indicators (1 to 6 in Fig. 2.5) should be sufficiently high as to overcome differences between media (newspaper vs. television) and between news content types (international vs. national versus local). If this is the case, we can speak meaningfully of "public affairs media attention" rather than having to deal with two separate media dimensions ("public affairs *newspaper* attention" and "public affairs *television* attention") or three separate content dimensions ("*international* news media attention," "*national news* media attention," and "*local* news media attention").
2. *General attentiveness hypothesis*. We might postulate that whenever one is exposed to media, regardless of content types, one is giving relatively similar levels of attention. That is, degrees of attentiveness refer to individual differences (between persons) that do not vary in response to media content types. If so, the six public affairs media attention indicators (1 to 6 in Fig. 2.5) and the four soft news/entertainment indicators (7 to 10 in Fig. 2.5) may be correlated so highly that they can be grouped into a single attentiveness index. Alternatively, if we assume that attention is a conscious response to a content type, meaning the content-specific attention hypothesis is empirically valid, the six public affairs attention indicators should yield zero or significantly lower correlations with the four indicators of "soft news/entertainment attention" relative to the correlations among the items within each set, thus giving evidence of *discriminant validity*.
3. *Hypothesis of distinctness of attention*. Based on our conceptual definition, attention to public affairs media ought to be conceptually and empirically distinct in that it differs from its closest neighbor, *exposure* to the corresponding content categories, which denotes the behavioral manifestation of individuals' uses of media in relation to the corresponding categories. If so, the correlation among indicators of public affairs media attention will be sufficiently homogeneous as to be capable of being discriminated from the indicators of public affairs media exposure. These are measures of how frequently (same 0 to10 scale with

	ATTENTION					EXPOSURE		
	PUBLIC AFFAIRS			SOFT NEWS & ENTERTAINMENT		PUBLIC AFFAIRS		
	International News	National News	Local News	Features/ Crimes	Sports/ Sitcoms	Intl.	Natl.	Local
Newspaper	1	2	3	7	8	11	12	13
Television	4	5	6	9	10	14	15	16

The groupings of items for the within-concept hypotheses are:

- Attention to public affairs media: 1–6
- Attention to soft news and entertainment: 7–10
- General attentiveness (cross-media): 1–10
- General attentiveness specific to newspapers: 1, 2, 3, 7, 8
- General attentiveness specific to TV: 4, 5, 6, 9, 10
- Public affairs exposure: 11–16

FIGURE 2.5. Media-related items grouped based on the within-concept hypotheses

0 being *never read* and 10 being *read frequently*) respondents read international, national, or local news in newspapers or watch each type of news on TV (indicators 11 to 16 in Fig. 2.5).

Index Building and Testing Within-Concept Hypotheses

The development of a Public Affairs Media Attention index, the first task of empirical explication, coincides with testing the first within-concept hypothesis, that Indicators 1 through 6 meet the criterion of homogeneity, a first indication of the reliability, or unidimensionality, of the index. The level of homogeneity among items constituting an index can be expressed in several ways, including *average inter-item correlation*, *corrected item-total correlation*, and *Cronbach's alpha*. The empirical assessment of the index also includes comparing this index, which is based on the hypothesis that attention is content-specific, with the other conceptual possibilities specified by Hypotheses 2 and 3, as well as some of their variations. When doing so, we are also bringing in specific criteria of validity. A particularly relevant one is that of content validity. That is, through such comparisons, we assess the extent to which high homogeneity may be achieved with a sacrifice of content validity, meaning the inclusion only of the indicators on a particular segment of a concept. Therefore, at no time can a decision made in empirical analysis be based on homogeneity reliability alone. Furthermore, at no time can such a decision be made without assessing conceptually whether an index represents the concept it is meant to represent, whether excluding an indicator from this index sacrifices a significant portion of the underlying concept, and whether empirical evidence of multidimensionality (e.g., the indicators under investigation form two or more distinct factors rather than one as hypothesized) has any theoretical merit (e.g., improved clarity in the meaning of the concept being examined and precision in stating and observing the relationships between the concept and other concepts).

Therefore, the analysis discussed here was carried out in accordance with the logic of hypothesis testing. More advanced statistical techniques based on such logic are available (e.g., confirmatory factor analysis via LISREL). However, this logic can be demonstrated more easily by analyzing patterns of a correlation matrix involving the relevant items.

Hypothesis 1: Content-Specific Attention. The primary task at this stage is to assess the internal consistency (homogeneity) of the proposed index called Public Affairs Media Attention. The matrix of correlation coefficients representing the associations among the six indicators of this index is shown in Table 2.4. At the bottom of the matrix, we also show the three measures of homogeneity.

Note that the 15 coefficients in the upper right triangle of the matrix are the mirror image of those in the lower left portion of the matrix. In interpreting the size of the coefficients, we can be guided by knowing that, for a sample of this size ($n = 480$), a coefficient of .09 or greater is significantly different from 0 at the $p \leq .05$ level (i.e., a coefficient of .09 or greater is unlikely to be obtained by chance if two variables are not related). However, to make strong statements about homogeneity we expect at least moderate coefficients of, say, .20 or greater. These decision rules are not precise and fixed thresholds for decision making. They are either social conventions (e.g., $p \leq .05$) accepted in the scientific community or rules of thumb based on a whole host of factors, including sample size, population heterogeneity, and the nature of data (e.g., more random error in self-reported data than in physiological data gathered by advanced equipment; within self-reported data, more random error in measures of subjective attributes such as opinions and feelings than in measures of more "objective" attributes such as age and gender).

The *average inter-item correlation* for each item (sum of the five coefficients in that column divided by 5) is satisfactory for all items (the first row in the second portion of the table), with relatively even averages ranging from .44 for both local news measures to .58 for television national news.

TABLE 2.4. Pearson Correlation Coefficients for Public Affairs Media Attention Indicators

Attention to Public Affairs	NEWSPAPER			TELEVISION			Overall
	Intl	Natl.	Local	Intl.	Natl.	Local	
1. NP international news	--	.76	.42	.56	.51	.21	
2. NP national news	.76	--	.60	.47	.63	.33	
3. NP local news	.42	.60	--	.22	.36	.59	
4. TV international news	.56	.47	.22	--	.81	.48	
5. TV national news	.51	.63	.36	.81	--	.60	
6. TV local news	.21	.33	.59	.48	.60	--	
Average inter-item correlation[a]	.49	.56	.44	.51	.58	.44	.50
Corrected item-total correlation[b]	.64	.73	.54	.66	.78	.56	.65
Alpha if item deleted[c]	.84	.82	.85	.83	.81	.85	.86

[a]Average inter-item correlation is an average of the correlation coefficients of an item with all the other items in the same index. For example, the average inter-item correlation (0.49) for attention to newspaper international news is the average of the five coefficients in the first column.

[b]Corrected item-total correlation is an average of the correlation coefficients of an item with the index comprised of the other items. For example, the corrected item-total correlation for newspaper international news attention is the correlation between this item and the index comprised of the other five items.

[c]Alpha if item deleted is a measure of reliability coefficient of the index that is comprised of the other items. For example, deleting newspaper international news attention, alpha for the index without this item would be .84.

NP = newspaper.

A *corrected item-total correlation* is the association of an indicator with the total index created from summing the other indicators. For example, the corrected item-total correlation for attention to newspaper international news is 0.65. This is the correlation between this item and an index based on the other five items (Items 2 to 6). Note that for each of the indicators these correlation coefficients are considerably higher than the corresponding averages in the row above for average inter-item correlation (e.g., 0.65 vs. 0.50). The average of 0.65 is quite satisfactory. The results illustrate the point made earlier that composite indices are more reliable than each individual indicator by itself.

Cronbach's alpha of the Public Affairs Media Attention index, as indicated in the last column of the bottom row of Table 2.4, is .86. Based on conventional standards for alpha, the index can be judged as highly reliable. The coefficients shown in the last row for the six individual indicators show what alpha levels would be if the corresponding indicator were to be removed from the total index. Because in all cases the alphas would *decrease* rather than increase, we can conclude that we should retain all items in the index. That is, every item makes its own contribution to the overall index. However, we need to keep in mind one property of Cronbach's alpha. That is, it increases as more items that are positively correlated are added to the overall index, potentially disguising weak items or low correlations among the indicators of an index. Therefore, item-total correlations need to be examined together with alpha.

Thus, all three measures indicate an acceptable degree of internal unidimensionality of our Public Affairs Media Attention index. Taking the same approach, we also analyzed the intercorrelations among the items within each content type (two indicators in each of the three content types, e.g., attention to international news in newspaper and attention to international news on TV) and within each medium, respectively (three indicators in each of the two media, e.g., attention to newspaper international, national, and local news). These analyses did not yield any indication (e.g., substantially higher average inter-item correlation among the items *within* each content type or medium and substantially lower average inter-item correlation *between* content types or media than what is shown in Table 2.4) of the presence of more than one dimension. The evidence clearly supports the hypothesis that the six attention indicators form a unidimensional index, Public Affairs Media Attention. Additionally, let's keep in mind that item analysis also assesses each individual item (indicator). The evidence shows that no item among the six is particularly poor in terms of its contribution to the overall index.

Hypothesis 2: General Attentiveness. The second within-concept hypothesis asserts that attention reflects general individual differences and does not respond to content or media types. Or, the general attentiveness may vary between media types due to individual variations in their responsiveness to different media. These are two alternative models of attentiveness. The first specifies a unidimensional concept of attention across media and content types. The second specifies medium-based dimensionality of the concept of attention. Given the measures that we have here, it means we would observe a TV attention dimension and a newspaper attention dimension.

To differentiate these two models and test each of them against the Public Affairs Media Attention concept, we analyze an expanded correlation matrix that includes four more attention items, two for newspaper "soft" news (features about ordinary people and sports and entertainment) and two for television entertainment (crime and adventure shows and situation comedies). The criterion of content validity for the general attentiveness models would call for such a broader inclusion to capture different segments of the concept. The empirical analysis part is to investigate whether the expanded correlation matrix contains any pattern to support either of these conceptual models of attention. The matrix shown in Table 2.5 thus involves 10 items (added Items 7 to 10 indicated in Fig. 2.5).

First, we can look at the average inter-item correlation for general attentiveness. To do so, we will take an average across the 10 columns in the corresponding row. This leads to an estimate of .24 for general attentiveness, showing rather low internal consistency of the hypothesized concept. Such evidence indicates two possibilities. One is that the additional items are poor measures due to insufficient variance or too much random error. The other is that attention is not unidimensional across content types. Closely examining the individual average inter-item correlation shows that the weakness is concentrated in the four nonpublic affairs items that were added. They have low inter-item correlations ranging from .14 to .17. Such a pattern suggests that attention to the two broad classes of media content is *unlikely* to be unidimensional. There may be two different dimensions in the concept of media attention: attention to public affairs content and attention to soft news and entertainment content. This case illustrates that content validity has to be assessed in terms of the unique dimension of an abstract concept that the variable represents. Further evidence of validity for differentiating the index of attention to public affairs from that of attention to soft news or entertainment has to be obtained by testing the hypotheses of the relationships of the two indices with some criterion variables (e.g., attention to public affairs correlates positively with knowledge on current affairs while attention to soft news or entertainment does not).

Next, we can look at the statistics on medium specific attentiveness. Dividing the 10 indicators into five newspaper (Columns 1–5) and five television indicators (Columns 6–10) raises the average inter-item (internal) correlation modestly to .30 for newspaper and .31 for television. To see whether this modest increase justifies the two medium-specific attention indices, we need to test *discriminant power* by taking the difference between the within-medium internal average to the cross-media external average (correlation of each newspaper indicator with each television indicator and vice-versa). For each medium-specific index, subtracting the *external correlation* (.20 for newspaper and .19 for television) from the average *internal correlation* leaves only .10 and .12 as the respective esti-

TABLE 2.5. Pearson Correlation Coefficients for 10 Attention Indicators (Arranged to Test General Attentiveness)

	Newspaper						Television					
	Intl.	Natl.	Local	Feat.	S&E	Avg.	Intl.	Natl.	Local	C&A	Com.	Avg.
Attention to newspaper												
1. International news (Intl.)	--	.76	.42	.04	.02	--	.56	.51	.21	–.05	–.09	--
2. National news (Natl.)	.76	--	.60	.08	.07	--	.47	.63	.33	–.05	–.07	--
3. Local news (Local)	.42	.60	--	.37	.24	--	.22	.36	.59	.05	.08	--
7. Features on people (Feat.)	.04	.08	.37	--	.42	--	–.10	–.02	.02	.12	.25	--
8. Sports & entertainment (S&E)	.02	.07	.24	.42	--	--	.05	.05	.11	.32	.33	--
Attention to television												
4. International news (Intl.)	.56	.47	.22	–.10	–.05	--	--	.81	.48	.09	.03	--
5. National news (Natl.)	.51	.63	.36	–.02	.05	--	.81	--	.60	.11	.07	--
6. Local news (Local)	.21	.33	.59	.20	.11	--	.48	.60	--	.20	.21	--
9. Crime and adventure (C&A)	–.05	–.05	.05	.12	.32	--	.09	.11	.20	--	.47	--
10. Situation comedies (Com.)	–.09	–.07	.08	.25	.33	--	.03	.07	.21	.47	--	--
General attentiveness												
Average inter-item correlation	.26	.31	.33	.15	.17	--	.28	.35	.31	.14	.14	.24
Medium-specific attentiveness												
Average internal correlation	.31	.38	.41	.23	.19	.30	.35	.40	.37	.22	.20	.31
Average external correlation	.23	.26	.26	.09	.15	.20	.22	.31	.25	.08	.10	.19
Discriminant power[a]	.08	.12	.15	.14	.04	.10	.13	.09	.12	.14	.10	.12

[a]Discriminant power is the difference between average internal and average external correlation coefficients. It measures the degree to which an item is capable of separating the two different indices under consideration.

mates of the discriminant power of each. The figure is roughly a third of the average internal correlation for each of the medium-specific attentiveness indices, suggesting a poor distinction between the two indices proposed to measure medium attention across content types.

Other indicators of homogeneity, item-total correlations and Cronbach's alpha also show nonfeasibility of "general attentiveness" as a concept. Item-total correlations averaged only .45, much less than the .65 shown earlier for Public Affairs Media Attention in Table 2.4. When we looked at alpha coefficients for two medium-specific attentiveness indices, we found .68 for newspaper and .69 for television. Although these coefficients reach a more acceptable level, they were not as high as the coefficient for Public Affairs Media Attention, showing that removal of the two sets of nonpublic affairs content indicators would actually increase the alpha levels. This suggests that the nonpublic affairs content indicators hurt the internal consistency of the hypothesized general attentiveness attention index or the medium-specific attentiveness indices. In terms of the internal consistency criterion (reliability), Public Affairs Media Attention is empirically superior to either general attentiveness or medium-specific attentiveness. Additional analyses are necessary to assess these different measurement models in terms of the other criterion, validity.

Notice that the general attentiveness hypothesis, or medium-specific attentiveness hypothesis, states *conceptual* alternatives to the first one, which says that public affairs media attention is a unidimensional construct. To further differentiate these conceptual models and to demonstrate the utility of the content-specific conceptualization, we re-analyze a correlation matrix that is shown in Table 2.5. This time, we rearrange the matrix to reflect the conceptual model of content-specific attention. The idea is to investigate whether the correlations contain any pattern to clearly *distinguish* between public affairs versus soft news/entertainment attention across the two types of media. In other words, we will examine evidence of *discriminant validity* for Public Affairs Media Attention. The matrix analyzed is shown in Table 2.6.

The discriminant power of each index (difference between average internal and average external correlation coefficients) gives us an indication of the discriminant validity—the degree to which two concepts with a close kinship can be empirically distinguished. The overall average external correlation of .07 suggests little overlap between the two attention indices. Taking an average of the six discriminant power coefficients (the last row), we have an estimate of .43, close in magnitude to the average internal correlation of the same six indicators (.50). Now we can take a look at the corresponding estimates for the other four attention indicators. We find that the discriminant power for attention to soft news and entertainment is .25, still rather large in comparison with the average internal correlation of the

TABLE 2.6. Pearson Correlation Coefficients for 10 Attention Indicators (Rearranged to Test Content-Specific Attention)

	Public Affairs							Soft News/Entertainment				
	Newspaper (NP)			Television			Avg.	Newspaper		Television		Avg.
	Intl.	Natl.	Local	Intl.	Natl.	Local		Feat.	S&E	C&A	Com.	
Attention to public affairs												
1. NP international news (Intl.)	--	.76	.42	.56	.51	.21	--	.04	.02	-.05	-.09	--
2. NP national news (Natl.)	.76	--	.60	.47	.63	.33	--	.08	.07	-.05	-.07	--
3. NP local news (Local)	.42	.60	--	.22	.36	.59	--	.37	.24	.05	.08	--
4. TV international news (Intl.)	.56	.47	.22	--	.81	.48	--	-.10	-.05	.09	.03	--
5. TV national news (Natl.)	.51	.63	.36	.81	--	.60	--	-.02	.05	.11	.07	--
6. TV local news (Local)	.21	.33	.59	.48	.60	--	--	.20	.11	.20	.21	--
Attention to soft news & entertainment												
7. NP features on people (Feat.)	.04	.08	.37	-.10	-.02	.20	--	--	.42	.12	.25	--
8. NP sports & entertainment (S&E)	.02	.07	.24	-.05	.05	.11	--	.42	--	.32	.33	--
9. TV crime and adventure (C&A)	-.05	-.05	.05	.09	.11	.20	--	.12	.32	--	.47	--
10. TV situation comedies (Com.)	-.09	-.07	.08	.03	.07	.21	--	.25	.33	.47	--	--
Average internal correlation	.49	.56	.44	.51	.58	.44	.50	.26	.36	.30	.35	.32
Average external correlation	-.02	.01	.19	-.01	.05	.18	.07	.10	.07	.06	.04	.07
Discriminant power	.51	.55	.25	.52	.53	.26	.43	.16	.29	.24	.31	.25

same set of indicators (.32), even though it is substantially smaller than the figure for attention to public affairs. Clearly, the evidence strongly favors our first within-concept hypothesis that the concept of public affairs attention is unidimensional and separable from attention to soft news and entertainment content.

Hypothesis 3: Distinctness of Attention. The third within-concept hypothesis actually goes beyond the concept of attention. It asserts the conceptual and empirical independence of media *attention* and *exposure*. At the conceptual level, the two may be distinguished with the former referring to heightened mental concentration on the media while the latter refers to behavioral contacts with the media. How well can this conceptual distinction be borne out empirically? Let's focus on attention and exposure to public affairs content in answering this question in part because we have already demonstrated the utility of distinguishing two content types in conceptualizing attention.

Table 2.6 shows that attention to news and attention to entertainment are correlated at .07 only, meaning the two indices are mostly independent from each other. When we look at attention and exposure to public affairs, we anticipate less independence. The reason is that attention must be contingent on exposure in that some minimal level of attention is a psychological condition during exposure. In addition, for newspapers, the act of reading requires some levels of attention. Thus, we expect attention and exposure to be positively correlated. However, if the two are distinct concepts, the strength of this association should not be so strong as to overcome the internal consistency of Public Affairs Media Attention. Statistics on the discriminant validity of public affairs attention and exposure are shown in Table 2.7.

Before looking at the statistics, we should explain a bit more about the measures used here. One problem with developing operational indicators of television public affairs exposure is the difficulty in finding measures comparable to the international, national, and local news types. These categories are appropriate for measuring attention in both newspapers and television. They are also meaningful categories when asking questions on frequencies of reading newspaper news because of the differentiation of newspaper sections that are largely based on such categories. However, when asking about frequencies of watching television news, such distinctions may cause difficulties among respondents because audiences cannot select which TV news stories to watch in the same manner as they can use headlines to guide their newspaper reading. We watch what is presented but can only increase or decrease our focus on different news items presented to us in a newscast. News stories that fall into these categories normally are beamed at us in a continuous flow of one TV newscast. Thus, somewhat

TABLE 2.7. Pearson Correlation Coefficients of Attention and Exposure Items

	Public Affairs Attention							Public Affairs Exposure						
	Newspaper			Television			Avg.	Newspaper			Television			Avg.
	Intl.	Natl.	Local	Intl.	Natl.	Local		Intl.	Natl.	Local	D&S	NNN	LSN	
Attention to Public Affairs														
1. NP international news (Intl.)	--	.76	.42	.56	.51	.21	--	.71	.60	.31	.08	.11	–.05	--
2. NP national news (Natl.)	.76	--	.60	.47	.63	.33	--	.62	.72	.47	.16	.11	.02	--
3. NP local news (Local)	.42	.60	--	.22	.36	.59	--	.34	.47	.73	.16	.12	.21	--
4. TV international news (Intl.)	.56	.47	.22	--	.81	.48	--	.47	.40	.22	.24	.32	.19	--
5. TV national news (Natl.)	.51	.63	.36	.81	--	.60	--	.46	.52	.35	.29	.30	.23	--
6. TV local news (Local)	.21	.33	.59	.48	.60	--	--	.24	.32	.51	.32	.31	.45	--
Exposure to Public Affairs														
11. NP international news (Intl.)	.71	.62	.34	.47	.46	.24	--	--	.71	.42	.19	.16	.01	--
12. NP national news (Natl.)	.60	.72	.47	.40	.52	.32	--	.71	--	.54	.18	.13	.05	--
13. NP local news (Local)	.31	.47	.73	.22	.35	.51	--	.42	.54	--	.28	.14	.25	--
14. TV documentary & specials (D&S)	.08	.16	.16	.24	.29	.32	--	.19	.18	.28	--	.28	.25	--
15. TV national network news (NNN)	.11	.11	.12	.32	.30	.31	--	.16	.13	.14	.28	--	.50	--
16. TV local station news (LSN)	–.05	.02	.21	.19	.23	.45	--	.01	.05	.25	.25	.50	--	--
Average internal correlation	.49	.56	.44	.51	.58	.44	.50	.30	.32	.33	.24	.24	.21	.27
Average external correlation	.29	.35	.34	.31	.36	.36	.33	.47	.51	.43	.21	.21	.18	.34
Discriminant power	.20	.21	.10	.20	.22	.08	.17	–.17	–.19	–.10	.03	.03	.03	–.07

different measures of exposure must be used for television news. Here we have used three news program types to strive for specificity of indicators: documentaries and news specials, national early evening network news, and local newscasts in the evening by local television stations.

We should now focus on the last two rows of Table 2.7. As can be seen, many of the items on public affairs attention have rather high average correlations with the public affairs exposure items (average external correlation). In Column 7, the overall average, .33, is more than half of the average internal correlation (.50) and twice as large as the discriminant power (.17). Together, although indicating some degree of distinction between attention and exposure indices, the figures tell us clearly that the main story here is the overlap between attention and exposure to public affairs content. It means one or both of the following two things. One, as far as public affairs content is concerned, attention and exposure are not distinguishable concepts. Further analysis is needed to test this measurement model. Second, if we treat them as distinguishable variables, to estimate the *unique* or *net* effect of public affairs attention, we *must* control for exposure to the same content.

The story does not end here. The average internal correlation and discriminant power of the public affairs exposure items seem to be saying the same thing. Although these items all have moderate internal consistency (ranging from .21 to .33), due to their overlaps with the items of pubic affairs attention, their discriminant power is either *zero* or *negative*, indicating either a lack of discrimination or greater overlap than distinction between exposure and attention. The evidence suggests that *at this level of aggregation across different content categories and media*, public affairs media exposure and attention are *not* empirically distinguishable constructs.

If that is the case, we can examine the data in more detail to gain a better understanding of how such overlap takes place. For attention measures, we can see that the discriminant power is .20 or greater for four of six indicators, quite acceptable for survey data. However, the other two items, attention to local news in newspapers and on television, are marginal in strength of discriminant power (.10 and .08). These two items also have lower internal correlation (.44 for each) with the other four items. But if we look at the correlation between local news exposure and attention, we find it at .73 for newspaper and .45 for television. In addition, the average of correlation among all four local news items is as high as .43. Such evidence suggests that in the complex constellation of the items on public affairs attention and exposure, attention and exposure to local news items may go together to form a distinct local news consumption index. It is possible that conceptually, one's consumption of local news reflects a distinct motivational condition and local orientation rather than need for more information in general or orientation to different media. Here, the evidence suggests a different within-concept hypothesis to explore.

Now look at the details of the exposure items. The overall Public Affairs Exposure index has a *negative* discriminant power (minus .07), and all three newspaper Public Affairs Exposure measures are also negative. Part of the problem, as we can see from the table, is due to the relatively low association of exposure across media (average of .15 between newspaper and television exposure). Another problem is the strong association (average = .55) between exposure and attention with respect to newspaper news (supporting our initial conception that reading requires some degree of attention). This implies that, in contrast to our findings for attention, Public Affairs Exposure probably should be divided into two indices: Newspaper Public Affairs Exposure and Television Public Affairs Exposure. That is, reading newspaper news and watching TV news may be two distinct types of behavior that should be represented by two different concepts. Such a conceptual model will have significant theoretical implications. For example, when assessing effect of newspapers, Newspaper Public Affairs Attention and Newspaper Public Affairs Exposure are largely, although not completely, interchangeable predictors. The same cannot be said of exposure and attention to public affairs content on TV (Chaffee & Schleuder, 1986). More detailed analysis reveals the following. With the six measures of exposure and attention to *newspaper* public affairs content, the internal correlations for attention and exposure to newspaper public affairs are .59 and .56, respectively. Remember that the average correlation between newspaper exposure and attention items is at .55; clearly, the six newspaper indicators can be comfortably grouped together to form an internally consistent index. The same cannot be said of exposure and attention to public affairs content on TV. For television, the three exposure indicators yield an average correlation of .34 among them; the three attention indicators are correlated at .63. The average correlation between attention and exposure measures is .29. It is clear that for television, exposure and attention to public affairs content, although overlapping, can be differentiated. More generally, the evidence suggests that exposure and attention to media cannot be plugged into a neat matrix composed by media types (e.g., TV vs. newspaper), content types (e.g., public affairs vs. soft news or entertainment), and behavioral types (exposure vs. attention).

The evidence does not give a clear answer to our question concerning the third within-concept hypothesis—attention is a distinct characteristic that individuals engage in while being exposed to media public affairs content. It sends us back to the drawing board to rethink how attention, especially the type of mental focus measurable via survey interviewing, should be conceptualized. From the analysis presented here, we have gained some insights, which are clues to how we may reconsider the dimensionality of attention. Going through this example in such detail also shows clearly how concept explication is a cyclical process.

A SUMMARY

We have come a long way from the first step in conceptualization—trying to understand what is a concept. This is a two-part journey: meaning analysis and empirical analysis. We have shown that meaning analysis is a logical process of specifying the meaning and dimensions of a concept and developing operational definitions for that concept. We have also shown the logic of empirical analysis applicable to creating and assessing a summated rating scale (or index) and its component indicators. We have presented our discussion as steps for testing the within-concept hypotheses that have developed through the meaning analysis in order to underscore the guiding role of theoretical reasoning in the whole process. Throughout our discussion, we used familiar communication concepts as examples. In the empirical analysis section, we went through some steps of evaluating one such concept, attention to public affairs content in the news media and how it is measured.

Why should we spend so much time and effort investigating the conceptual and empirical properties of one concept? The answer is twofold. At the general level, this is required for developing scientific knowledge. As Sartori (1984) said, part of what science does is to develop a "*special and specialized language*" to "correct the defects of ordinary language" (pp. 57–58). More specific to our example, this is needed for the development of our field. After the first 60 years of mass communication research, communication variables are still often used in an imprecise and sloppy fashion. This has led to conflicting and sometimes uninterpretable findings about media effects. There have been proposals offered to deal with this problem. For example, as Bartels (1993) showed, weak media exposure measures could lead to substantial underestimations of the effect size of media messages; therefore, a measurement model needs to be part of the empirical model of media effects. Another political scientist, Zaller (1996; see also Price & Zaller, 1993), asked us not to be limited to behavioral measures of media exposure, given their poor empirical performance as predictors of information acquisition and the attitude change conditioned on such acquisition. Sometimes, we also see careless neglect in conceptualization. For example, a well-publicized study is set to convince us that somehow watching television is like having a vacuum cleaner suck information out of our brain, because the more time people spent on watching

TV, the less they know (Morgan, Lewis, & Jhally, 1992). All these proposals and problems suggest that lack of sophistication and precision in concept explication persist in our field. A hard and humiliating fact is that, after more than 60 years of research, despite some fruitful efforts (e.g., Chaffee & Schleuder, 1986; McLeod & McDonald, 1985), we as communication researchers still do not have theoretically powerful conceptions of media uses that would have clear operational definitions for different theoretical purposes and clear connections among varying operational measures. With greater attention to concept explication, it may be possible that in the next 50 years, the field will produce better conceptual building blocks. Efforts along this line would greatly improve theory building in our field and provide a more informed basis for communication and information practices.

As our discussion shows, concept explication is no easy task. It runs through the entire cyclical process of scientific research. Before bringing a research project into the field, careful conceptualization will provide clearer guidance on what to measure and how to measure it, and consequently, will make the resulting data more useful. At the practical level, it will improve the efficiency of research, meaning that one does not collect unusable data. Of course, the matter goes well beyond that, because the matter is directly related to how we can interpret evidence supporting or rejecting a research hypothesis.

As Fig. 2.4 shows, two types of relationships are simultaneously present in a hypothesis-testing study, namely, between-concept and within-concept relationships. Ambiguities in within-concept relationships render confusion in and over the between-concept hypothesis. In other words, concept explication is the basis for us to establish both conceptually and empirically that we are testing the right hypothesis derived from a particular theory under concern. Results may be interpreted as evidence in support of a between-concept hypothesis *only if* the validity of the measures involved is established. Conversely, null results, meaning evidence in support of the null hypothesis, might suggest either lack of support for the research hypothesis or problems in within-concept relationships (lack of validity and/or reliability in the measures used) or both. Results that are contradictory to a well-established empirical criterion or a well-developed theory should be a call for a reconsideration of how the key concepts are explicated and measured. For example, if the evidence shows no relationship between news media exposure and knowledge of current affairs, it could indicate potential problems in our measurement of one or both of the variables. If we observe a significant negative relationship between a media exposure variable and the knowledge on current affairs, rather than jumping to the conclusion that more exposure leads to less knowledge, we should go back to examine how each variable is conceptualized and measured. A decision on whether to reject a null hypothesis, therefore, is not just a statistical one based on the conventional criterion of significance; rather, it is also a conceptual one in that it requires a solid foundation in concept explication.

In addition, as we showed in our empirical analysis example, any measure is limited, and we need to be aware of what theory (or which aspect of a theory) of some particular recurring phenomenon we can test with a particular set of measures. Let us continue to use the concept of "attention" as an example. We define the concept as "a conscious process of focusing concern" on a particular type of media content during the exposure. This conception would be appropriate for the method we use to measure the concept: gathering self-report data in a sample survey. When developing the questionnaire items (indicators) to measure attention to various types of media content, we are assuming that audiences are capable of using the same content type categories that we rely on in the measurement. We further assume that audiences are capable of reliably reporting their conscious efforts in allocating their mental resources across these categories. These assumptions limit the utilities of self-report measures in a theory of attention. For example, such measures would not yield any data on variations in involuntary attention, such as momentary changes in mental orientations in response to changing external stimuli (Lang, 1990; Reeves et al., 1985). Nor would the self-report measures be appropriate when studying those who cannot report their behavior or psychological conditions reliably (e.g., young children). In brief, although it is clear that choosing a measurement strategy is a theory-guided decision, the other side of the coin is that one is limited by the chosen measures in what inferences one can make about a theoretical claim involving the concept being measured. This is why we stated that any research hypothesis is only a partial representation of a theoretical statement.

Furthermore, the measure that we choose also specifies the variance that is available for data analysis. This is more than an issue of choosing among different levels of measurement. It is also an issue of the *range* as well as the *kind* of variation in a sample that is captured by an empirical variable. For example, attention to news may be measured on a yes–no (nominal) scale. If so, we would be able to examine differences only among those who claim to have paid attention to news and those who report they have not. To test a hypothesis on the effect of different *levels* of attention, we will need a scale higher than the nominal one. We can measure attention on a three-point ordinal scale, *little*, *some*, *a lot*, or, as in our case, a 10-point interval scale. We assume that each of these scales, despite their differences, captures the full range in the "real world," from the minimum to the maximum that respondents could possibly recognize, differentiate, and reliably report. If at either the lower or the upper end, a scale misses a portion of the systematic variation in the characteristic being measured, it is said to have been "truncated," and it would generate a highly skewed distribution. A truncated scale could produce seriously distorted estimates of its relationship with other variables. This does not necessarily mean the wider the range of a scale the better (e.g., a 10-point rather than a 3-point scale). Rather, it means that one needs to *properly define* the lower and upper ends of a scale.

Sometimes, the systematic variation in the characteristic being measured may only occur at a lower or higher portion of a theoretically possible range. For example, when measuring the importance of protecting individuals' civil liberties, the proper scale (let us continue to use a 10-point scale) might not be the one with the lower end anchored in *not at all important,* as would be the case when measuring some other characteristics, because on this value, it would be very hard to find some cases at the *not important* end of the possible *importance* scale. We might instead want to find a proper lower end so that the resulting scale would not be too restricted in variance and the sample distribution would not be too skewed. Whether defining the lower end to be *somewhat important* or *important* would depend on both pragmatic (i.e., the efficient use of the scale in terms of generating a good sample distribution) and theoretical (e.g., need to compare different lower ends between peace times and crisis times) considerations. At the empirical analysis stage, then, one needs to inspect the descriptive statistics of each indicator to confirm the expectations that each item captures the right variance.

Clearly, concept explication requires a researcher to "travel" back and forth between the conceptual and empirical domains, making a series of theoretical as well as practical trade-offs along the way. We start to make such trade-offs right from the onset of a research project. For example, should we develop a concept of "media attention" that would be so general that it could be employed to discuss attention to the media by all populations in all situations? Although we do strive toward generality in scientific research, trying to explicate "media attention" at that level of abstraction could render it too far removed from its empirical referents to be operationalizable. A trade-off must be made in locating the appropriate level of abstraction. To take another example, in measuring a complex construct, should we develop multiple indicators? The logic of assessing content validity and reliability in terms of internal consistency requires us to do so. However, limited by funding and other practical constraints, we may not be able to develop a sufficient number of indicators for all the constructs in a study. A trade-off must be made between methodological desirability and practical feasibility. The third example of trade-offs concerns the decision making in empirical analysis. Although the logic of evaluating an indicator or an index is clear, it is not so when making a decision on a *particular* indicator or an index because such a decision needs the integration of the information of different kinds and no single statistical criterion could be relied on exclusively. Thus, as we showed, Cronbach's alpha must be used with caution because of its unique property of varying with the number of indicators included in an index. An average internal correlation coefficient would be better interpreted *together* with a corresponding average external correlation coefficient. More specifically, would an average internal correlation of .40 be considered "sufficiently strong" as an indication of internal consistency of an index? It might if an external correlation separated this index from a closely related one at .30 (discriminant power), leaving only .10 as the overlap between the two (external correlation). If the two indices overlap at .20 or higher, with the same internal correlation, the evidence would become much less clear-cut. We also need to bring other practical considerations into our decision making. For example, are these two indices based on self-report data on respondents' internal psychological characteristics? How do they compare with other alternative measures that have been employed in the field? Although the first question suggests that we have realistic expectations of the measurement that we employ, the second question asks us to try to use *the best available*, although not necessarily ideal, measures.

Another trade-off a researcher faces arises when there is a discrepancy between prior theorizing (meaning analysis) and subsequent empirical evidence about a concept. As part of explicating our concept via meaning analysis, we have specified multiple indicators to represent our conceptual definition, but the empirical analysis of internal consistency reveals that one or more putative indicators of the concept fail to correlate with the other indicators that are highly correlated. To evaluate the evidence, we have to keep in mind that in this case, the conceptual model established through meaning analysis does *not* make the indicators alternatives to one another, as would have been the case with, say, frequency of watching ABC evening news versus frequencies of watching CBS or NBC evening news. Therefore, the criteria of content validity and homogeneity reliability do not logically point toward opposing decisions with regard to dropping indicators that have weaker associations with other indicators in the same index. Then, how should we assess the tension between content validity, a criterion for conceptual inclusion of the items, and internal consistency (homogeneity reliability), a criterion for empirical correlations among these items?

An illustration of this dilemma is provided by our example of attention to public affairs media used in this chapter. The six-indicator index had an average inter-item correlation of .50 (Table 2.4). The two local news attention indicators are less reliable, however, in that the average correlation of each with the four indicators of international and national attention is only .40. Excluding the two local indicators would improve reliability (but not much in alpha because of its unique property of increasing with the number of positively correlated items), raising the average inter-item correlation of the four international and national news indicators to .62. Improved reliability as a criterion argues for dropping the items, but our prior concept explication argued that local news was an important aspect of public affairs content and thus important to include according to the criterion of content validity.

How do we resolve this dilemma of a trade-off between homogeneity reliability and content validity? There are three things we could do with the two weaker local news items: We could drop them from our index and exclude them from further analyses. We could divide our public affairs attention concept into two dimensions, international–national news attention with four items and local news attention with two items. We could keep them as part of the six-item index. To choose among the three alternatives, several steps are needed. First, we consider the possibility of weaknesses in the measurement of local news items. This seems unlikely in this example because the introduction to the attention questions in the survey is the same for all six items. Furthermore, although the two local news items are correlated with each other at .59, their average correlation with the four nonlocal news items is at .40. Thus, weakness of measurement is not indicated, and there is no reason to simply drop the items on those grounds. Second, we would examine the empirical feasibility of using local news attention as a separate two-item index. The correlation of .59 between the two items indicates that a local news attention index is feasible. However, the overlap between this local news attention index and the index of attention to international and national news (average correlation coefficient .40) suggests that they may work against each other when both indices are included in a predictive model of some effect variable (e.g., knowledge of current affairs). It could lead to an underestimation of the effect of attention to public affairs media content. Third, we return to our conceptual understanding of public affairs attention to determine if the local news attention and international–national news attention indices would have different antecedents or effects, as predicted by some well-formulated theories. In this case, we would predict that the two indices differ in their relationships with many plausible antecedents and effects only in degrees (e.g., education should be more strongly related to international–national attention than to local news attention) but not in kind or direction. Therefore, although we could proceed to group all six indicators into one index (i.e., treating them as indicators of a single concept), or to create two separate indices (i.e., as two dimensions of the concept), of public affairs media attention), there is *no compelling theoretical basis* for having two separate indices. Nor do we expect significant gains in theoretical understanding from the separation. Whichever we decide, it is always wise

to do some post hoc analyses by disaggregating an index to examine the relationships between each indicator and a set of antecedents and effects. This may help us to better understand the concept for future research.

All these examples suggest that research is not free from flaws and limitations. Careful conceptualization would help researchers treat such flaws and overcome certain limitations. The more skilled and careful we are in conceptualization, the better able we are to handle various flaws and limitations in research practices. This chapter was simply set to inform our readers of the logic and certain practical steps of handling such flaws and limitations.

ACKNOWLEDGMENT

This chapter, originally prepared in September 1998, is adapted from a two-part reading for Journalism 658: Communication Research Methods, taught at the University of Wisconsin–Madison. The ideas discussed here have been central components of the collective intellectual endeavor at the Mass Communications Research Center (MCRC). Our debt to the late Steven H. Chaffee is duly recognized. Many individuals who honed their research skills at MCRC left their influences on the development and articulation of these ideas. In alphabetical order, they are David Amor (Knox College), Charles Atkin (Michigan State University), Lee Becker (University of Georgia), Brian Beil (public relations), Jane Brown (University of North Carolina, Chapel Hill), Carl Bybee (University of Oregon), Annie Chi (Shih Hsin University, Taiwan), Sun Yuel Choe (Ewa University, Korea), Katie Daily (post doctoral), Jean Durall (public relations), William Elliott (Marquette University), William P. Eveland, Jr. (Ohio State University), Ronald Faber (University of Minnesota), Carroll Glynn (Ohio State University), Zhongshi Guo (Hong Kong Baptist University), Edward M. Horowitz (Cleveland State University), Stacy Huang (Shih Hsin University, Taiwan), Donald Johnson (public relations), J. David Kennamer (Virginia Commonwealth University), Gerald Kosicki (Ohio State University), Robert Krull (Rensselaer Polytechnic Institute), Nojin Kwak (University of Michigan), William Leutscher (evaluation research), Daniel McDonald (Ohio State University), Patricia Moy (University of Washington), Garrett O'Keefe (Colorado State University), Gary Pettey (Cleveland State University), Donald Pember (University of Washington), Maria Powell (post-doctoral studies University of Wisconsin–Madison), Donna Rouner (Colorado State University), Dianne Rucinski (opinion research), Dietram A. Scheufele (University of Wisconsin–Madison), Pamela Shoemaker (Syracuse University), Mira

Sotirovic (University of Illinois UC), Catherine Steele (Evaluation Research), Vernon Stone (University of Missouri, retired), Sewen Sun (National Chung Cheng University, Taiwan), Albert Tims (University of Minnesota), Leonard Tipton (University of Florida), Sjef van den Berg (University of Connecticut), Daniel Wackman (University of Minnesota), James Watt (Rensselaer Polytechnic Institute), Scott Ward (University of Pennsylvania), Dean Ziemke (evaluation research), Jessica Zubric (University of Wisconsin doctoral program). However, the two authors take sole responsibility for any error or inadequacy in this chapter.

GLOSSARY

Abstractness: A characteristic of scientific concepts. It refers to the degree to which a concept is distanced from some specific events or phenomena observed at some particular time and place. Alternatively, we can think of this as indicating the range of objects or subjects (units of analysis) whose characteristic is represented by the concept.

Clarity in meaning: A characteristic of scientific concepts. It refers to the degree to which the intended meaning of a concept is clearly stated and conveyed. It also means that sharing the intended meaning of a concept among parties of communication is achievable through some appropriate label for the concept and careful definitions of it in its specific usages.

Class concept: A concept that denotes some collection of objects or subjects by summarizing the commonality among them.

Concept: A term or phrase used to represent an abstract idea. It denotes a characteristic common to all the objects or subjects in a particular class. It is the basic building block of thought and theory.

Concept explication: The process by which abstract concepts are linked to their real-world referents so that they can be observed by appropriate methods. It involves both deductive and inductive reasoning. It can be further divided into meaning analysis and empirical analysis.

Conceptual definition (theoretical, constitutive): It involves verbal descriptions of the essential properties that are to be included in the intended meaning of a concept. In research practice, it often involves specifying the logical dimensions of a concept.

Content validity: An approach to assessing validity of an instrument. It involves a researcher making a judgment as to whether the measurement instrument misses an important dimension or component of a complex construct. For example, if television viewing is conceptualized as including amount of viewing and selection of programs to watch, then an instrument measuring TV viewing that includes only a single question on time spent watching TV would not have sufficient content validity.

Construct (scientific concept): A concept that has been consciously invented, constructed, or adopted for purposes of theory building and research. It should be clearly defined and may be broken down into less abstract (more specific) dimensions. It should meet four standards: abstractness, clarity in meaning, operationalizability, and precision.

Construct validity: An overarching conception of the criterion of validity in empirical analysis. It refers to the hypothesized relationship between an abstract concept and its empirical referents (or conceptual and operational definitions) and between that concept and other concepts theoretically linked to it. It requires the combined power of logical reasoning and empirical evidence to assess construct validity of a measure.

Deduction: A mode of logical reasoning that involves deriving statements (or hypotheses) from more general theoretical propositions (premises or assumptions). For example, from the general law of gravity, we derive a hypothesis concerning the movement of a free-falling object. In concept explication, we follow the deductive reasoning to move from an abstract construct to its empirical referents (indicators). Very often, deduction and induction are intermingled in concept explication.

Dependent variable: In either theoretical statements or research hypotheses, it is the variable concept whose fluctuation (variation) is to be explained. It is assumed to be affected by the independent variable to which it is linked by a relational statement.

Dimension (subconcept): Distinguishable components of a more abstract concept that have higher levels of internal coherence than has the concept. Each dimension of a concept is a concept in its own right, only less abstract than the concept of which it is a constituent part. Various dimensions of a concept may have different antecedents and consequences and if so, theoretical statements must be stated for the constituent dimensions rather than for the more abstract concept.

Discriminant power: A statistical measure of the degree to which indicators of one concept are distinct from those of another concept. Operationally, it is the difference between *internal* and *external correlations*. It is treated as evidence of *discriminant validity* of the measures under consideration.

Empirical analysis: Statistical and logical procedures involved in using research data to evaluate the quality of conceptual and operational definitions of a concept. More specifically, this is a process of assessing empirical indicators developed in meaning analysis, creating indices whenever necessary, and assessing their empirical properties. It provides an empirical basis for further development and explication of a concept.

Explanation: A set of logically organized statements specifying why and how some observed events or phenomena have occurred. Because it lays out the specific conditions under which conceptually similar phenomena occur or recur, it can be used to derive specific predictions.

External correlation: A statistical measure of the relationship between indicators measuring two different variables. For example, when we have six indicators measuring Media Public Affairs Attention and another six measuring Media Public Affairs Exposure. The correlation between the two sets of indicators would be an external correlation. It is a measure of the degree of overlap between two indices.

Fallacy of reification: The logical error of attributing concrete observations to an empty conceptual entity or equating a functional relationship in some specific situations with generalizable conceptual abstraction.

Homogeneity: A particular approach to reliability that involves intercorrelation among the indicators designed to measure the same concept in a summated rating scale. In other words, the indicators to be grouped into one index must yield high correlations among them before we can say that they are homogeneous enough to be grouped in this way.

Independent variable: In either theoretical statements or research hypotheses, it is a variable concept used to explain variation or change in another variable. It is assumed to affect or influence the other variable to which it is linked by a relational statement.

Index: A composite variable that involves combining two or more indicators based on certain rules. It is often considered (but needs to be demonstrated empirically) to be more general and reliable than each indicator. It is also considered to fare better in content validity than each indicator separately.

Indicator: Concrete observable manifestations of a concept or a dimension (e.g., questionnaire items or observable behavior). It is concrete in that it can be directly or indirectly recorded via some research technique. Its relationship with the concept that it is purported to "indicate" is assumed to be invariant within some specified range of time, space, and domain. Furthermore, such a relationship is understood within a culture.

Induction: A mode of logical reasoning that involves developing general propositions from concrete empirical observations. For example, based on observing a large number of white swans, we may conclude that "all swans are white." As a mode of logical reasoning, induction is flawed, as it is characterized by the term of "inductive fallacy" (i.e., overgeneralization or discrepancy in scopes between an abstract and inclusive statement and limited evidence on which the statement is based). Therefore, inductively generated propositions must be tested with independently gathered empirical evidence. Induction is an important logical procedure in concept explication, as it is shown in Fig. 2.3. Very often, induction and deduction are intermingled in concept explication.

Inter-item correlation: A statistical measure of the relationship between indicators measuring the same concept. It is a measure of internal consistency or homogeneity of an index. It is often averaged across all pairs of indicators.

Interval scale: A level of measurement that involves not only classification and rank ordering of the attributes of a variable but also a fixed distance between each pair of adjacent attributes. For example, a 10-point scale may be used to measure attention to TV network news. The distance between 1 and 2 on this scale is assumed to be the same as that between 9 and 10. By using an interval scale, the units of observation can be evenly arrayed on the scale.

Item-total correlation: A statistical measure of the relationship between an indicator and the index of which the indicator should be a part but is not yet included. Here is how we can calculate it. First, we form an index by combining all its constituting indicators except the one that we wish to examine. (Including the item being evaluated would inflate the item-total correlation.) Second, we treat this "incomplete" index (total) and the indicator (item) under examination as separate variables and obtain the correlation coefficient between them. Item-total correlation tells us how much of a unique contribution an indicator (item) may have to an index.

Manipulation: Procedure used in experimental research to induce variations in some specified variable(s) as captured by two or more conditions into which subjects are randomly assigned. It is intended that the differences in responses between the subjects in various conditions can be causally attributed to the induced variations.

Meaning analysis: Logical and epistemological procedures used in developing connections between conceptual and operational definitions of a concept. It involves specifying the essential properties and thus boundaries of the recurring phenomenon denoted by a concept. Along with empirical analysis, it comprises the iterative process of concept explication.

Measurement: Assignment of values (numbers or symbols) to units of observation according to rules. The rules are those specified in the operational definitions and reflect the standards used by scholars in a given field. Measurement includes nominal or categorical assignment as well as more powerful forms: ordinal, interval, and ratio measurement.

Mega-concept: A composite concept representing multiple characteristics of a class of object or subjects, often compounded with affect and/or values. It needs to be decomposed into several concepts (e.g., democracy into a class concept representing a type of political system, or into a variable concept indicating the degree to which democratic ideals are practiced) in order to be useful in empirical research.

Nominal scale: A level of measurement that involves only classification of units of observation. It gathers information on qualitative difference between attributes of a variable. A variable that uses a nominal scale is variably called "categorical variable" or "discrete variable." Examples are gender, race, religion, and occupation.

Null hypothesis: A statistical statement that asserts that there is no relationship beyond chance between the variables in the research hypothesis. In other words, it is a statement that any relationship or difference shown is simply due to chance or sampling error.

Observation: The recording of data from units of observation pertinent to a given variable and compatible with its operational definition. In its specific usage, the term refers to a research method (i.e., field study) in which an observer records data either as a nonparticipant or participant in an ongoing process involving a group of people. More broadly, it refers to various forms of obtaining data from units of observation: direct observation, interviews, self-administered questionnaires, recording with an instrument, and collecting archival or archaeological data, and so on.

Operational definition: Procedures by which a concept is to be observed (as in participant observation), measured (as in sample surveys), or manipulated (as in experiments). It details the rules, specific steps, equipment, instruments, and scales involved in measuring (etc.) a concept. All three types of operational procedures can be generally called either observation or measurement, using these two terms in their broadest sense.

Operationalizability: A characteristic of scientific concepts. It refers to the degree to which the linkage between the concept and its real-world referents can be specified and its variation across a class of objects or subjects can be observed through the available means for empirical research. It is contingent upon clarity in conceptual definition of a concept and/or practical feasibility (e.g., availability of apparatus) in carrying out the stipulated measures.

Ordinal scale: A level of measurement that involves not only classification but also rank ordering of the attributes of a variable. As a result, units of observation can be compared as to possessing more or less of the particular characteristic measured by the variable. For example, attention to TV network news shows may be measured on a 3-point scale of *very little*, *some*, and *a lot*.

Precision: A characteristic of scientific concepts. It refers to the degree to which a concept is clearly defined, with clear boundaries from other concepts, and its intended meanings are agreed on by commuunicators (scientists) and receivers (consumers of scientific research or policy makers).

Prediction: A statement about the likelihood that some specific event or phenomenon will occur under specific conditions in some designated future time.

Premise (assumption): Propositions that supply reasons for a theoretical statement. Two or more premises are needed to draw a logical conclusion. Often, a premise is a general statement that has achieved a law-like status in that it is generally held to be true and does not require further empirical tests. Assumption is its synonym when applied more generally in theory and research. Assumptions should be testable with empirical evidence, and often they are based on a considerable body of evidence from previous research.

Primitive term: Word or phrase whose meaning is widely shared and incapable of further definition except by using synonyms. It is used to define other theoretical terms in a theory.

Process concept: A concept that characterizes a particular way in which a recurring phenomenon takes place. Usually, it points to certain logically determined outcomes of some effect variable. It is a key element in a theory that helps us to answer the question of "how," an important part of understanding.

Ratio scale: The highest level of measurement which not only differentiates various attributes of a variable, ranks them, and shows equal distance between each pair, but also records complete absence of the characteristic measured by the variable. For example, education measured by years of formal schooling may range from 0 to 18 or more. "Zero" here means total absence of any formal schooling, the characteristic of the units in a population measured.

Referent: Conceptually implied and empirically observable counterparts of a concept in the real world.

Relational concept: A concept that makes connections between other forms of concepts. The connections can be comparative (e.g., larger than, as long as), associative (e.g., positively related, negatively related), or causal (e.g., leads to, caused by, has an impact on . . .).

Research hypothesis: Assertion about the relationship between two or more variables stated in terms of concrete operational definitions. Their relationship is predicted from the logic contained in the corresponding theoretical statement and in the explication of the variables contained in the statement. A research hypothesis is stated before any empirical evidence is examined. The assertion describes the direction (i.e., negative or positive), type (linear or nonlinear), nature (i.e., causal or associative), and/or the magnitude (strength) of the relationship.

Scale: A term often used in two different but related ways when discussing concept explication, especially operationalization and empirical analysis. One is that a scale is a composite measure that combines some indicators based on the patterns of intensity or strength among them. For example, a Guttman scale often involves several items that can be arranged from the least to most intense so that respondents who choose "yes" in answering the most intense item should or are expected to choose the same answer for all the items with lower intensity levels. A good analogy is that if a person can answer a difficult math question correctly, that person should have a higher likelihood of answering the less difficult math questions correctly than another person who cannot answer the difficult math question correctly. The second meaning of scale is the system of assigning values to the

units of observation. For example, the Likert scale is a system of assigning any number from 1 to 5 to a unit based on the direction and strength of a respondent's attitude (labeled as *strongly disagree*, *disagree*, *feel neutral*, *agree*, and *strongly agree*). Two or more items measured on a Likert scale, if they are designed to measure the same concept or dimension of a construct, may be added together to form an index.

Singular concept: A concept that denotes some particular object or subject (e.g., a person's name or the name of a university).

Theoretical statement: Assertion about the relationship between two or more relatively abstract variable concepts or their dimensions. It should be clear as to the units of analysis under concern and the conditions under which the statement should hold. Theoretical statements are more abstract and general than hypotheses. A theoretical statement can be tested indirectly by testing (via null hypotheses) a variety of research hypotheses that are logically compatible with it.

Theory: An explanation of some recurrent phenomena of research interest. It is a collection of (a) theoretical statements about the relationship between two or more variable concepts or their dimensions within a domain or collection of units of analysis under specified conditions; (b) the premises or assumptions providing the reasoning behind the theoretical statement; (c) two or more variable concepts and their conceptual and operational definitions; and (d) specific research hypotheses connecting operational definitions of the concepts included in the theory.

Unit of analysis: Element that is studied and analyzed through empirical data. Variation among the units of analysis on some characteristic is captured by a variable, and it is the focus of any research project. Units of analysis commonly used in mass communication research include individuals, media organizations, media systems, dyads, groups, families, and communities.

Unit of observation: The singular object or element that we actually observe. It may or may not correspond to the unit of analysis in a specific research project. Where the two units do not correspond (e.g., coding each half-hour TV show to garner data for comparing TV networks in their levels of violence in these shows), the data obtained from the units of observation may be aggregated to provide measures of variables for the units of analysis. Alternatively, the unit of observation stands for the authoritative source of information on a unit of analysis (e.g., the head of household is interviewed for the data on each household).

Variable concept: A concept that distinguishes objects or subjects in terms of the degree to which they possess some designated characteristic (e.g., level of education, age, television viewing hours per day). Variable concept enables relationships to be stated in statistical or deterministic terms (correlation or causality).

Within-concept relationship: The conceptual linkage between an abstract concept and its real-world referents (indicators), which could be a specific object or subject, a class of objects or subjects, or observable variations among a class of objects or subjects over a characteristic. It is the relationship hypothesized and tested when assessing construct validity.

Working definition: A convenient operational procedure linking a concept to its real-world variations for a specific research purpose. It does not necessarily imply a conceptual meaning of the concept and thus may not be generalizable beyond the specific research purpose.

REFERENCES

Anderson, D. R., & Burns, J. (1991). Paying attention to television. In J. Bryant & D. Zillmann (Eds.), *Responding to the screen: Reception and reaction processes* (pp. 3–25). Hillsdale, NJ: Erlbaum.

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Babbie, E. (2004). *The practice of social research (*10th ed.*)*. Belmont, CA: Wadsworth.

Bartels, L. M. (1993). Messages received: The political impact of media exposure. *American Political Science Review, 87*, 267–285.

Cappella, J. N., & Jamieson, K. H. (1997). *Spiral of cynicism: The press and the public good*. Oxford, England: Oxford University Press.

Chaffee, S. H. (1991). *Communication concepts 1: Explication*. Newbury Park, CA: Sage.

Chaffee, S. H., & Schleuder, J. (1986). Measurement and effects of attention to media news. *Human Communication Research, 13*, 76–107.

Donohew, L., & Palmgreen, P. (1989). Theory construction. In G. H. Stempel III and B. H. Westley (Eds.), *Research methods in mass communication (*2nd ed., pp. 30–48). Englewood Cliffs, NJ: Prentice Hall.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*, 51–68.

Gamson, W. A. (1996). Media discourse as a framing resource. In A. N. Crigler (Ed.), *The psychology of political communication* (pp. 111–132). Ann Arbor: University of Michigan Press.

Giddens, A. (1984). *The construction of society: Outline of the theory of structuration*. Berkeley: University of California Press.

Goffman, E. (1974). *Frame analysis*. New York: Harper & Row.

Hage, J. (1972). *Techniques and problems of theory construction in sociology*. New York: Wiley.

Heise, D. R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review, 34*, 93–101.

Hempel, C. G. (1952). *Fundamentals of concept formation in empirical science*. Chicago: University of Chicago Press.

Iyengar, S. (1992). *Is anyone responsible? How television frames political issues*. Chicago: University of Chicago Press.

Jo, E., & Berkowitz, L. (1994). A priming effect analysis of media influence: An update. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (pp. 43–60). Hillsdale, NJ: Erlbaum.

Kerlinger, F. N., & Lee, H. (1999). *Foundations of behavioral research (*4th ed.*)*. Belmont, CA: Wadsworth.

Lang, A. (1990). Involuntary attention and physiological arousal evoked by structural features and emotional content in TV commercials. *Communication Research, 17*, 275–299.

Lasswell, H. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 32–51). New York: Harper.

Mason, S. F. (1962). *A history of the sciences*. New York: Collier Books.

McCombs, M., & Ghanem, S. I. (2001). The convergence of agenda-setting and framing. In S. D. Reese, O. H. Gandy, Jr., & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 67–94). Mahwah, NJ: Erlbaum.

McIver, J. P., & Carmines, E. G. (1981). *Unidimensional scaling*. Beverly Hills, CA: Sage.

McLeod, J. M., & McDonald, D. G. (1985). Beyond simple exposure: Media orientations and their impact on political processes. *Communication Research, 13*, 3–33.

Merton, R. K. (1958). *Social theory and social structure*. New York: The Free Press.

Morgan, M., Lewis, J., & Jhally, S. (1992). More viewing, less knowledge. In H. Mowlana, G. Gerbner, & H. I. Schiller (Eds.), *Triumph of the image: The media's war in the Persian Gulf* (pp. 216–233). Boulder, CO: Westview.

Pan, Z., & Kosicki, G. M. (1993). Framing analysis: An approach to news discourse. *Political Communication, 10*, 55–75.

Pan, Z., & Kosicki, G. M. (2001). Framing as a strategic action in public deliberation. In S. D. Reese, O. H. Gandy, Jr., & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 35–66). Mahwah, NJ: Erlbaum.

Patterson, T. E. (1993). *Out of order*. New York: Knopf.

Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Dubuque, IA: Wm. C. Brown.

Pratkanis, A., Breckler, S. J., & Greenwald, A. G. (Eds.). (1989). *Attitude structure and function*. Hillsdale, NJ: Erlbaum.

Price, V., & Zaller, J. (1993). Who gets the news? Alternative measures of news reception and their implications for research. *Public Opinion Quarterly, 57*, 133–164.

Reese, S. D. (2001). Prologue—Framing public life: A bridging model for media research. In S. D. Reese, O. H. Gandy, Jr., & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 7–31). Mahwah, NJ: Erlbaum.

Reeves, B., Thorson, E., Rothschild, M., McDonald, D., Hirsch, J., & Goldstein, R. (1985). Attention to television: Intrastimulus effects of movement and scene changes on alpha variation over time. *International Journal of Neuroscience, 25*, 241–255.

Sartori, G. (1984). Guidelines for concept analysis. In G. Sartori (Ed.), *Social science concepts: A systematic analysis* (pp. 15–85). Beverly Hills, CA: Sage.

Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication, 49*, 103–122.

Singleton, R. A., Jr., & Straits, B. C. (1999). *Approaches to social research* (3rd ed.). New York: Oxford University Press.

Snow, D. A., Rochford, E. B., Worden, S. K., & Benford, R. D. (1986). Frame alignment processes: Micromobilization and movement participation. *American Sociological Review, 51*, 464–481.

Tuchman, G. (1978). *Making news: A study in the construction of reality*. New York: The Free Press.

Wallace, W. L. (1971). *The logic of science in sociology*. Chicago: Aldine Atherton.

Wiley, D. E., & Wiley, J. A. (1970). The estimation of measurement error in panel data. *American Sociological Review, 35*, 112–117.

Wyer, R. S., Jr., & Srull, T. K. (1989). *Memory and cognition in its social context*. Hillsdale, NJ: Erlbaum.

Zaller, J. (1996). The myth of massive media impact revived: New support for a discredited idea. In D. C. Mutz, P. M. Sniderman, & R. A. Brody (Eds.), *Political persuasion and attitude change* (pp. 17–78). Ann Arbor: University of Michigan Press.

Zillmann, D. (1971). Excitation transfer in communication-mediated aggressive behavior. *Journal of Experimental Social Psychology, 7*, 419–434.

PART 2