CHAPTER 8

Bipolar Scales

James C. McCroskey,
West Virginia University
Virginia P. Richmond,
West Virginia University

One of the oldest methods of measurement, which currently remains very popular, is bipolar scaling. The essence of this method is the establishment of concepts that are opposite of one another, with degrees or steps between these extreme poles.

Every student has been subjected to the most basic of bipolar scales—the true-false test. This type of bipolar scale recognizes no degrees or steps between the opposing poles: the statement is either "true" or "false." However, the test-taker sees shades or degrees of truth or/and falsity, which makes choosing between the extreme polar options very difficult.

A similar problem occurs when we are asked to evaluate a policy or another person on a "good-bad" bipolar scale. Few policies (or people for that matter) are totally "good" or totally "bad." Thus, it is best to have a several-step scale between good and bad. We might, for example, have the following options: very good, fairly good, equally good and bad, fairly bad, very bad. These options form a typical five-step, bipolar scale. The number of options may vary substantially, depending on a large number of factors, but all bipolar scales are essentially similar to this one.

While most bipolar scales employ some type of verbal description of the opposing poles, and often include verbal descriptions of the intermediate steps, this is not an essential characteristic of this method. For example the poles can be identified by colors, with black and white being the opposing poles, and the steps in between can be represented by various shades of gray. Similarly, researchers have used drawings of smiling and frowning faces to establish bipolar scales when working with pre-literate children. The steps between the extremes are created by varying the degree of smile or frown on the face.

PROCEDURE

The development of measures based on the method of bipolar scaling is both extremely simple and very complex. It is simple in that any person of below average intelligence can slap a few bipolar scales together, assign numbers to responses on them, add up the numbers and claim to have a measure. Indeed, examples available in the published literature of the social sciences indicate this not only can be done but it is being done. On the other hand, developing a solid measure with this procedure is a fairly complex and often a time-consuming process.

The development of a measure employing bipolar scales involves three initial steps: (1) determining what is to be measured; (2) determining how many different concepts must be measured; and (3) determining the scales to constitute the measure. We will consider each of these in turn.

Determining what is to be measured. Measures of this type typically are designed to probe such things as feelings, perceptions, attitudes, and/or beliefs which people have. They may be self-reports of one's own internal feelings, beliefs, or attitudes. They may be self-reports of one's perceptions with regard to something or someone in the external environment. An example of the former would be Jack reporting how competent as a communicator he perceives himself to be. An example of the latter would be Jack reporting how competent as a communicator he perceives Bill to be. The critical test of the validity of the former is whether it measures what Jack really thinks about himself. The critical test of the latter, depending on what the researcher has decided he/she is trying to measure, could be whether it measures what Jack really thinks about Bill or whether it measures the degree of Bill's competence as a communicator. The same scale could be both an excellent measure of Jack's perceptions and a worthless measure of Bill's competence.

The importance of determining what is to be measured as the first step in the measurement process cannot be overestimated. If this task is done carefully, it removes the potential for a lot of confusion and error later. For example, if it is clear what is to be measured, it usually is clear exactly how the validity of the measure can be determined. Equally important, once it is clear what is to be measured it becomes possible to conduct a search of the literature to determine whether an appropriate measure already exists. It is totally wasteful of research effort to take the time to develop a new measure when an appropriate measure already exists. Exerting such effort will normally only be appropriate when there is need for more than one alternative measure of the same thing.

To exemplify, colleagues of the first author of this chapter have avoided the need to develop measures of attitude, belief, and credibility for many years because of parallel projects which the author conducted in the mid-1960s. After spending several months developing Likert-type attitude scales on two topics for a research project in persuasion, he became convinced there had to be a better way. After becoming familiar with the work on measurement of meaning

(Osgood, Suci, & Tannenbaum, 1957), he decided to attempt to develop a generalized scale to measure attitudes toward a wide variety of topics. This work, which need not be reported in detail here, involved 40 bipolar scales and 154 topics. The outcome of this project indicated that 6 of the 40 bipolar scales were among the ten judged best for measuring attitude for each of the 154 topics. In addition, the internal reliability (Alpha) of the six scales, presented in a seven-step response format and summed as a single measure, was above .90 for each of the topics. The six scales are good/bad, wrong/right, harmful/beneficial, fair/unfair, wise/foolish, and positive/negative.

These scales can be used for measuring general attitudes toward such issues as capital punishment, foreign aid, current speed limits, and the like, with considerable confidence that they will generate a reliable and valid measure of attitude that will correlate highly with other carefully developed and tested measures. In most instances, these scales may be used to measure people's attitudes without additional testing. Figure 8.1 illustrates how the measure could be presented to research subjects. Note that half of the items need to be reversed for scoring. The resultant scores may range from 6 to 42, with 24 representing the hypothetical neutral score. Higher scores represent more positive attitudes.

The scales presented in Figure 8.1 cannot, of course, be used to measure all kinds of attitudes and beliefs. The researchers determined they were not appropriate for measuring feelings toward belief statements such as "The U. S. should reduce taxes." For these kinds of beliefs, the following bipolar scales were found to produce a much more appropriate measure: true/false, right/wrong, yes/no, disagree/agree, correct/incorrect. A measure based on these scales was found to be highly reliable ($>.90$) when used with over 75 belief topics and thus can be used with confidence for such topics in place of the scales used in Figure 8.1.

To summarize this section, it is critical that the researcher planning to use bipolar scales determine at the outset precisely what is to be measured. Once this is done it may be determined that such measures already exist and are available for use. If, however, it is determined that no appropriate measure is available, the researcher should proceed to the next step.

Determining how many things are to be measured. Instruments may measure only a single variable or several variables. The former type is referred to as **unidimensional**, the latter as **multidimensional**. While bipolar scales may be used to develop either type of measure, it is of vital importance to know whether a measure is multidimensional or unidimensional.

To understand the concept of multidimensionality in measurement, consider physical space. We know that space has length, width, and height. Those are the three dimensions of physical space. Knowing any one of these dimensions of a space tells us quite a bit about that space. Knowing all three gives us a great deal of knowledge about the space. For example, if we know the

Directions: On the scales below, please indicate your feelings about _____. Circle the number between the adjectives which best represents your feelings about _____. Numbers "1" and "7" indicate a very strong feeling. Numbers "2" and "6" indicate a strong feeling. Numbers "3" and "5" indicate a fairly weak feeling. Number "4" indicates you are undecided or do not understand the adjectives themselves. Please work quickly. There are no right or wrong answers.

Good	1	2	3	4	5	6	7	Bad*
Wrong	1	2	3	4	5	6	7	Right
Harmful	1	2	3	4	5	6	7	Beneficial
Fair	1	2	3	4	5	6	7	Unfair*
Wise	1	2	3	4	5	6	7	Foolish*
Negative	1	2	3	4	5	6	7	Positive

*Scoring should be reversed for these items.

Figure 8.1. General Attitude Scale

length of a room is 16 feet, this gives us some information. If we also know the width of the room is 14 feet and the height is 8 feet, we have a much better picture of the room. It becomes an average size room in our minds. If, however, someone told us the size of the room was 38 feet (16 + 14 + 8), we really would have no idea at all of what the room looked like.

The person describing the room size as 38 feet has no understanding of the concept of physical space. He or she is adding length to width to height. That is like adding five apples to seven oranges to four watermelons. We have 16, but we don't know 16 what. The basic idea here is that we cannot simply add across dimensions to obtain a score on a measure. Each dimension must have a separate score. While there are complex mathematical methods for generating meaningful numbers representing composites of numerous dimensions, simple addition is not one of those methods. Thus, determining the dimensionality of a construct prior to developing a measure for it can greatly reduce the problems of measurement development.

Consider, for example, the concept of "interpersonal attraction." This has to do with how much we are attracted to another person. Many researchers have treated this concept as unidimensional, and several have used bipolar scales to measure it. Based on the substantial theoretical work of others, McCroskey and McCain (1974) demonstrated that this concept definitely is *not* unidimensional. Although they did not use bipolar scales in their research, the approach they used illustrates what researchers need to do when they suspect the concept they want to measure is multidimensional.

Based on the work of other writers in the area of interpersonal attraction, McCroskey and McCain (1974) decided that there probably were three dimensions of interpersonal attraction—physical, task, and social attraction. They reasoned that we see some people as beautiful or handsome, but would really not like to be around them socially or work with them. Further, they presumed that we know some people we would like to socialize with because they are fun, but they are less than prize physical specimens and are goof-offs at work. Additionally, they presumed that we know some people who are very good to work with, but whom we do not want to spend time with away from work. While they recognized that some people are positively attractive to us on all three dimensions and some are repulsive on all three, it is quite possible to be attractive on one or two dimensions and repulsive on one or two others at the same time. Thus, these dimensions may operate independently.

McCroskey and McCain (1974) proceeded to construct scales to measure each of the hypothesized dimensions of interpersonal attraction (in this case they used Likert-type scales rather than bipolar scales, but the rest of their procedure is appropriate for either type of scale). After developing the scales, they randomly placed scales from all three presumed dimensions on a single instrument and had subjects respond to the scales in terms of a person with whom they interacted in a study. They then analyzed the data by means of a statistical procedure known as "factor analysis." The results confirmed the existence of the three presumed factors of attraction and the scales that had been created formed three distinct, but related, measures.

Factor analysis is a very sophisticated form of data analysis and we will not attempt to explain it fully here. For our purposes it is enough to say that factor analysis determines what groups of scales are substantially correlated with each other but not highly correlated with scales in other groups. These "groups" of scales represent the dimensions inherent in a measure. If there are no distinct groups, the measure is unidimensional. If there are two or more groups of scales, the measure is multidimensional.

Although in the McCroskey and McCain (1974) study, the factor analysis results supported the existence of the three dimensions they had presumed to exist, such supportive findings do not always occur. For example, in earlier research McCroskey (1966) had concluded from reading classical writers such as Aristotle and early social psychologists (Hovland, Janis, and Kelly, 1953) that source credibility (or as he called it at the time "ethos") was composed of three dimensions—competence (or intelligence), character (or trustworthiness), and good will (or intentions toward the receiver). After developing measures using bipolar scales as well as Likert-type scales and subjecting his data to factor analysis, he was forced to conclude the good will or intention dimension simply did not exist separate from the character or trustworthiness dimension. His findings, however, resulted in separate measures for each of the other two hypothesized dimensions (see Figure 8.2).

The importance of factor analysis in the process of developing a measure

Competence

Reliable	1	2	3	4	5	6	7	Unreliable*
Uninformed	1	2	3	4	5	6	7	Informed
Unqualified	1	2	3	4	5	6	7	Qualified
Intelligent	1	2	3	4	5	6	7	Unintelligent*
Valuable	1	2	3	4	5	6	7	Worthless*
Inexpert	1	2	3	4	5	6	7	Expert

Character

Honest	1	2	3	4	5	6	7	Dishonest*
Unfriendly	1	2	3	4	5	6	7	Friendly
Pleasant	1	2	3	4	5	6	7	Unpleasant*
Selfish	1	2	3	4	5	6	7	Unselfish
Awful	1	2	3	4	5	6	7	Nice
Virtuous	1	2	3	4	5	6	7	Sinful*

*Scoring should be reversed for these items.

Figure 8.2. Measures of Competence and Character

using the bipolar scaling methodology should not be understated. Many researchers have totally destroyed the value of their work by ignoring the dimensionality issue. Some have published research which treated every scale as an independent dimension, hence an independent measure. Single scales may be highly unreliable; thus the amount of error associated with a score on a single scale may be extremely high. Hence, differences between scores may be much larger or smaller than they appear by looking at the results generated by such scales. In contrast, some have published research in which they have summed data from scales representing several dimensions to get a single score (remember our apples, oranges, and watermelon?). Their scores, hence the entire results of their research, become instantly worthless. Imagine, for example, how you would interpret a score on interpersonal attraction that summed across physical, task, and social scales. You might decide to ask someone for a date, when he/she simply hoped you would make a job offer!

Determining the scales to include in the measure. The third and final step of the procedure for developing measures using bipolar scales is the "nuts and bolts" step. When we start this step we presume we know what it is we want to measure and its dimensions. We may discover the dimensionality of the measure is different than we presumed at the outset so that we must go through a major revision of the measure later. However, at this point we assume we know what it is we need to measure and that no appropriate measure currently is available.

Bipolar scales themselves are at the heart of the bipolar scaling methodology. On the face of it, bipolar scales seem extremely easy to construct. All you need is two words (pictures, colors, etc.) that are opposites. The rest is simple. You create some steps—3, 5, 6, 9, or whatever—in between them.

Unfortunately, it is not as simple as that, although numerous researchers have done exactly so—to the detriment of their later research. As communication teachers constantly stress to their students, words do not have meaning. Meaning is in the minds of people. Thus, concepts that seem to a researcher to be opposites of one another may not seem so to research subjects. What is the opposite of hard? To a researcher concerned with rocks, it might be "soft." To a student who has just completed an exam in another class, it might be "easy."

It is very tempting to select opposite words for scales by going to the dictionary or thesaurus. Both of these sources treat words as if they had meaning and are insensitive to context differences. Even one of the classic research efforts employing bipolar scales (Osgood, Suci, & Tannenbaum, 1957) was partially guilty of using this approach. Actually, the only way to be sure that words may be used appropriately to form bipolar scales is to be certain that research subjects, like those with whom the completed instrument will be used, will respond in opposite ways to the words.

The most direct method of determining this is to select a substantial number of words that appear to form bipolar opposites and appear to be related to whatever it is the researcher wishes to measure. Once selected, these words should be randomly placed to form an instrument like the one in Figure 8.3. You will notice the words are not placed in the form of bipolar scales but rather stand alone. The research subjects are asked to indicate on a five-point scale the degree to which each word describes the concept to be measured. Data on this working instrument from at least 100 research subjects (preferably 200 to insure stability of correlations) should be submitted to simple correlational analyses.

Words that have a perfect negative correlation (-1.00) are, of course, ideal for use in the formation of bipolar scales. Unfortunately, you are unlikely to find any such perfect correlations. As we noted before, single scales often are not very reliable and, consequently, cannot produce valid high correlations. As a rule of thumb, the correlation between words should be $-.50$ or stronger to qualify those words for use as the anchors of a bipolar scale.

If the observed correlation is smaller than $-.50$ it is an indication that one or more significant problems may be present. The most common is that different subjects are interpreting one or both words in different ways. Another is that the two words are generating responses to different things. In some instances, you may even find words you thought were opposites to be *positively* correlated. When this occurs it almost always indicates one of the words is being interpreted differently than intended. If "good" and "bad" are positively correlated, the subjects may be responding to "baaaaad" not "bad!"

Once you have selected your bipolar opposites, you must determine how many steps you want to place on your scale between them. It is customary to

Directions: Please indicate the degree to which you think each of the following words apply to _____ on a continuum of 1 to 5. Presume "5" represents *very well* and "1" represents "*very poorly*." The numbers "4," "3," and "2" represent the steps in between these extremes.

_____ good
_____ right
_____ harmful
_____ fair
_____ negative
_____ foolish
_____ beneficial
_____ wrong
_____ bad
_____ unfair
_____ wise
_____ positive

Figure 8.3. Preliminary scale items

use seven steps, although some research conducted by the present authors has indicated five steps produces the same results. There is no reason to expect major differences if one uses nine steps. Although in some instances researchers have even used 100 steps successfully, too many steps can generate false precision. It is doubtful that many people can make a distinction between 42 and 43 on the continuum between good and bad!

In general, adults and better educated subjects can handle more steps. Young children often can handle only three—each pole and "I don't know." Surveys conducted by telephone also may suffer from a large number of steps—the respondents may not be able to remember them all. In general, our advice is to use seven or five steps unless there is some very good reason to use some other number.

The final concern in developing the initial instrument is how many scales to include. Within limits, the general rule is the more the better. Of course, people will get exhausted if you ask them to respond to too many scales at one time, and your data will be worthless. Hence, it often is important to reduce the number of scales. The issues to be considered in determining how many scales must be included are primarily the *intervality* of the score on the measure and the *internal reliability* of the measure.

The use of parametric statistical analyses requires the assumption that scores analyzed are at least at the interval level. While visual examination of a bipolar scale makes it appear obvious that such scales have equal intervals, research indicates clearly that the individual scales do not (McCroskey,

Prichard, & Arnold, 1968). In fact, the two steps at the extreme poles are seen in the minds of respondents as twice as far from their neighbors as any other two steps. Thus, a single seven-step scale should probably be scored 1, 3, 4, 5, 6, 7, 9 rather than 1, 2, 3, 4, 5, 6, 7 as usually is done. This is not a serious problem, however, since this same research indicates summed scores for six scales do meet the intervality assumption. Thus, if you make certain you have at least six scales, that will be sufficient to meet the assumption of parametric statistics.

The second factor of concern is that of internal reliability. In general, measures based on the bipolar scaling approach have very high internal reliability. It has been our experience in working with bipolar scales for over 25 years that there are some very consistent patterns of association between the number of scales and the internal reliability of scores representing a summation of those scales. In general it appears that two or three scales will produce reliability of .5 to .7; four or five scales will produce reliability of .6 to .8; six to seven scales will produce reliability of .7 to .9; and eight or more scales will usually produce reliability above .9.

Mathematically, number of scales and internal reliability must be positively related if other things are equal. If you obtain reliability below the range noted above for your given number of scales, it most likely is an indication that you have multidimensionality in your measure. This is a sign that you must return to step two of this procedure; you probably have added apples and oranges to obtain your score.

Exactly how high you require your reliability to be before you are satisfied with your measure is a matter of judgement. The lower your reliability, the lower the precision in your instrument and the more likely you will have so much error in your data you will draw a false conclusion. In our view, there seldom is reason to accept internal reliability below .80 when using the bipolar scaling approach. It is simply too easy to improve the measure and thus raise the reliability. This rule should not be universely applied, however. In some cases, for example, increasing the number of items to gain better reliability might lead to exhaustion of the subjects when many measures are being completed. By adding more scales in such an instance one would simply be trading one source of error for another. Obviously, that would serve no useful purpose.

RELIABILITY OF METHOD

Reliability relates to consistency of measurement. Two types of consistency are of concern. One relates to consistency across time—if we administer the measure today and again next week, will we obtain similar (consistent) scores? This is often referred to as test-retest reliability. The second type of consistency relates to consistency within the measure. Is the score on one subset of scales within the measure highly correlated (consistent) with the score from another subset of scales? This usually is referred to as internal reliability.

The type of reliability with which a researcher may be concerned will depend on the purpose for which the measure was developed. In some cases it is assumed that what is being measured is something likely to be highly stable across time, such as a person's attitudes on some public policy questions. In other cases, it is clear that what is being measured is subject to substantial change from time to time, such as a person's mood.

When the item to be measured is subject to frequent change, the only type of reliability with which we are concerned is internal reliability. This type of reliability may be assessed through **split-half** techniques or analysis of variance procedures. But whichever method is used, internal reliability of measures employing bipolar scales, if developed properly, usually is very good. Reported reliabilities above .80 are common. Those above .90 are not unusual.

Determining the reliability across time of measures based on bipolar scaling is more complex, as it is with other types of measures. The actual computation process is easy—a simple correlation is computed. This may be between two scores on the same instrument obtained at two different times or it may be between two scores on two different, but presumably equivalent, instruments collected at two different times. The latter often is referred to as "alternate forms" reliability. Either method will provide a good estimate of the reliability of the measure if it can be established that there is no good reason to think the item measured would have changed over time. This method of measuring reliability is useless if the target of the measure is changeable, since the fact that two scores may have a low correlation may be a function of either unreliability of the measure or change in what is being measured, and there is no way to sort out these contributing factors. Although bipolar scaling has been used more commonly to measure things subject to frequent change, such as feelings and perceptions, when used to measure stable characteristics the reliability across time generally has been good.

Before leaving the topic of reliability, we should consider the placement of scales on a measure, because such placement can impact reliability. In the early development of a measure, when multidimensionality is suspected, scales should be randomly assigned to positions on the measure. This will reduce the possibility that a false "dimension" will be produced by factor analysis simply because several somewhat related scales were answered sequentially. Later, after the question of dimensionality has been settled, items on a given dimension should be grouped together. In this way respondents can focus on one thing at a time. Doing so will tend to increase reliability somewhat.

In this regard, you are cautioned to avoid one type of scale assignment that will appear to increase reliability, but will not actually do so. If you turn back to Figure 8.1 you will notice that some scales have the "good-positive" pole on the left and some have it on the right. If you put all of the positives on one side and all of the negatives on the other, reliability estimates are likely to go up. Unfortunately, this probably is a function of what is known as "response bias" rather than increased reliability. When people are responding to this type of

measure they tend to get into a pattern and may not even pay attention to the individual scale to which they are responding.

The authors have seen this error made on instruments used to measure students' evaluations of classroom teachers in several universities. Unfortunately, when this is done, however the student responds to the first few items will predict with considerable accuracy how he/she responds to the later items. To break this response bias, approximately half the scales should have the "positive" pole on the left and half on the right. While this might slightly reduce the reliability estimate for the measure, it will not reduce the real reliability and might actually increase it.

VALIDITY OF THE METHOD

As noted in chapter 5, validity of a measure has to do with whether it actually measures what it is supposed to. Assessing the validity of measures based on bipolar scaling often is very difficult. This difficulty stems not from the measuring procedure but from the nature of what is being measured. Attitudes, feelings, perceptions, and the like exist in people's minds. They cannot be observed directly. Hence, measures, such as ones based on bipolar scaling, often are used to estimate them. Once such measures are developed, researchers commonly try to validate them by observing people's behavior. Such an approach often is overly simplistic.

Consider the question of attitude measurement. As noted in the previous chapter, attitudes are not always consistent with behavior. Thus, if scores on a measure based on bipolar scales do not correlate highly with observed behavior, we might be tempted to say the measure is not valid. Such a conclusion could be completely false. If a behavior is not related to an attitude, the fact a score based on a measure of that attitude does not correlate with the behavior is quite meaningless. It tells us nothing about the validity of the measure.

While we do not want to discount completely any consideration of the correlation between scores on a measure and observable behavior as a method of estimating validity, such data often will be inconclusive or misleading. Thus, it usually is necessary to consider other data in making validity judgements.

Content or **face validity** frequently are important to consider. For example, we should be able to tell something about the validity of a bipolar scaling measure just by looking at the bipolar scales employed. Which of the following scales would you think might be appropriate to measure a person's attitude toward his/her mother? (1) kind-cruel, (2) wise-foolish, (3) hot-cold, (4) timely-late, (5) good-bad, (6) fast-slow. If you picked 1, 2, and 5 your choices have been supported by research (Osgood, Suci, & Tannenbaum, 1957). If you would say those three should be better than the others, you are suggesting a scale using those three would have "face" validity—they look like they should be related to

attitude toward one's mother. While it would be dangerous to base all decisions on the validity of a measure of this type on such "face" concerns, this often is a good place to start.

Concerns with **concurrent validity** are very common among researchers using the bipolar scaling methodology. Often there are complex measures already available of something of interest, which have been determined to be valid. The concern of the research may be the development of simpler measures. Bipolar scaling often is used when researchers desire simpler measures. If the bipolar scaling measure correlates well with the more complex, and previously validated, measure, it is concluded the bipolar scaling measure also has validity.

In general, bipolar scaling measures are very reliable. Any reliable measure is a valid measure of *something*; the question is whether it is a valid measure of the factor the researcher wants to measure. If the procedures described earlier in this chapter are followed, the odds are good the resulting measure will be valid for its intended purpose.

ADVANTAGES AND DISADVANTAGES OF PROCEDURE

There are two major advantages to the use of measures based on the bipolar scaling approach. First, they tend to be highly reliable and valid if properly developed. Second, they consume little subject time so that more items can be measured at one sitting without exhausting the subjects. The first advantage is a concern of all researchers. The second is particularly important to those researchers who must measure a substantial number of items at one sitting. When subjects are asked to fill out too many scales, their later responses may be little more than random.

The major disadvantage of measures based on the bipolar scaling approach is that they appear so easy to use many researchers succumb to the temptation to take shortcuts and become very sloppy. Instead of carefully designing and pretesting their measures, they may simply slap some bipolar scales together and call them a measure. They administer the "measure," add the scores across the scales, and analyze the results. As a result they may be adding across several dimensions of response, have inadequate reliability, and have no validity at all. Or they may even take every single bipolar scale as an independent measure and analyze scores for each scale separately. As noted previously, such scales typically do not even represent interval scaling, hence are not amenable to the statistical analyses to which they are typically subjected. The weakness of the bipolar scaling approach is that shoddy researchers can use it. It is not an exaggeration to say measures based on the bipolar scaling approach are no better nor worse than the researcher using them.

SUMMARY

This chapter has reviewed one of the oldest methods of measurement, the bipolar scaling technique. The three initial steps in the development of a measure employing bipolar scales are determining (1) what is to be measured, (2) how many different variables must be measured, and (3) the scales to constitute the measure.

Determining what is to be measured involves questions of validity. A measure is only of value if it measures something of importance. Once that "something" is identified, the literature must be surveyed to determine if appropriate measures already exist. This developmental stage is critical in that it establishes at the outset whether the measure ultimately developed will be of any value to researchers.

Determining how many variables are to be measured primarily involves the distinction between unidimensional and multidimensional constructs and measures. While determining unidimensionality and multidimensionality initially is considered conceptually, ultimately it must be determined empirically as well. Factor analysis frequently is employed in this process.

A decision on the scales to include in a measure presumes the researcher knows what it is she/he wants to measure, and its dimensions. Bipolar scales themselves are at the heart of the bipolar scaling methodology. The researcher must determine empirically that the scales truly are bipolar and decide how many steps will be employed between the bipolar extremes. Next, he or she must determine empirically the dimensionality of the measure, obtain estimates of the reliability and validity of the measure, and repeat these processes until the completed measure is acceptable for research purposes.

STUDY QUESTIONS

1. Describe the bipolar scaling technique. Discuss its advantages and disadvantages for social scientific research.
2. Discuss why the authors suggest, "the importance of determining what is to be measured as the first step in the measurement process cannot be overestimated."
3. Review the work of Osgood, Suci, and Tannenbaum. Describe the significance of their research for the field of communication.
4. Distinguish between "unidimensional" and "multidimensional." Identify why this distinction is necessary in bipolar scaling. Review the McCroskey/McCain interpersonal attraction measure and discuss the nature of the measure.
5. Explain why it is necessary for the researcher to determine clearly the dimensionality of a measure before using it in studies. Discuss consequences if this procedure is not followed.
6. Review the various methods of determining the words (pictures, etc.) which are used to form bipolar scales. Discuss the issues related to each method.

7. Discuss how a researcher might determine if he/she has found a useful bipolar scale.
8. Discuss the "sophistication" of the measure when administering it to adults versus children.
9. Review and discuss the reliability and validity issues as related to bipolar scaling. Discuss the issue of "scales must have equal intervals."
10. Explain why content or face validity are important to consider in bipolar scaling. Identify some bipolar scales that might be used to measure your attitude toward your teachers.

ANNOTATED READINGS

Kerlinger, F. N. (1973). *Foundations of behavioral research* (2nd ed.). New York: Holt, Rinehart and Winston.

This text is designed to assist the student in understanding the nature of scientific behavioral research. It is specially designed for persons in the sciences and education. It is divided into the following section: language and approach of science; sets, relations, and variance; probability, randomness, and sampling; analysis, interpretation, statistics, and inference; analysis of variance; designs of research; types of research; measurement; methods of observation and data collection; multiple regression; and factor analysis.

Kerlinger, F. N. (1979). *Behavioral research: A conceptual approach.* New York: Holt, Rinehart and Winston.

The purpose of this book is to assist students in comprehending scientific research and its components. It looks at research and applies scientific principles from a wide array of fields. It reviews the following topics: nature of science: scientific concepts; forming hypotheses; relations and explanations; probablity; design; experimental versus nonexperimental; multiple regression; factor analysis; canonical correlation; and concludes with misconceptions and controversies.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957). *The measurement of meaning.* Urbana, Ill., University of Illinois Press.

This book is a report on the authors' research program on measurement of meaning. It discusses their ongoing program of research and reviews more than fifty studies. It is the text which introduced the field to semantic differentiation through the use of bipolar scaling. This is the most extensive single research program based on bipolar scaling extant in the literature.