

directly, by providing a means by which people evaluate stimulus attributes (e.g., one might learn via classical conditioning procedures that a particular nationality is "good" and subsequently be inclined to have a positive attitude towards anyone who has this nationality).

See also: ATTITUDE CHANGE; GENETIC INFLUENCES; MERE EXPOSURE; SOCIAL LEARNING.

BIBLIOGRAPHY

- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Forth Worth, TX: Harcourt Brace Jovanovich.
- Page, M. M. (1969). Social psychology of a classical conditioning of attitudes experiment. *Journal of Personality and Social Psychology*, 11, 177-186.
- Staats, A. W. (1983). Paradigmatic behaviorism: Unified theory for social-personality psychology. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 16, pp. 125-79). New York: Academic Press.
- & Staats, C. K. (1958). Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology*, 57, 37-40.
- Zanna, M. P., Kiesler, C. A., & Pilkonis, P. A. (1970). Positive and negative attitudinal affect established by classical conditioning. *Journal of Personality and Social Psychology*, 14, 321-28.

ANTONY S. R. MANSTEAD

attitude measurement and questionnaire design A great deal of research has examined methods of designing questionnaires to measure attitudes and other subjective psychological constructs effectively (*see also* ATTITUDE THEORY AND RESEARCH, SURVEY METHODS), and this research provides a basis for a number of recommendations. We shall begin below by reviewing some of the original attitude measurement techniques, which involved elaborate procedures and multiple items. Then we will review the more recent literature on single-item measurement and the many decisions one must make in designing such measures.

EARLY ATTITUDE MEASUREMENT METHODS: MULTIPLE ITEM INDICES

A number of elaborate attitude measurement techniques were developed beginning in the 1920s (for reviews, *see* Dawes & Smith, 1985; Mueller, 1986; Summers, 1970). One such technique was developed by L. L. Thurstone. In his classic paper, "Attitudes can be measured," Thurstone (1928) argued that attitudes toward objects could be gauged in ways similar to those used to assess perceptions of sensory stimuli such as light and sound. Although this idea does not seem to be particularly revolutionary today, it marked a bold departure from the dominant behaviorist tradition of the time, which held that latent psychological constructs were not legitimate topics of scientific inquiry or assessment.

Thurstone developed several techniques for measuring attitudes, the most popular of which was the method of equally appearing intervals (EAI). EAI scales are developed by generating a large pool of statements, each of which expresses some degree of positivity or negativity toward a target object. A group of judges then classifies each statement into one of eleven categories according to how much positivity or negativity the statement reflects toward the object. The categories are then numbered from 1 to 11, representing increasing positivity, and a scale value is assigned to each statement by computing the median or mean rating of the statement (on the 1 to 11 scale) across the judges. A final item set is then constructed by selecting one statement to best represent each of the 11 intervals. This final set of items can then be administered to respondents, instructing them to indicate with which statements they agree. The median or mean scale value of the statements with which each respondent agrees serves as the index of the respondent's attitude.

Although the EAI method is highly reliable, the extensive preparation necessary to construct EAI scales spurred researchers to investigate other, simpler methods. One popular one proposed by Rensis Likert is the method of summated ratings. This method begins by generating a large number of statements reflecting positivity or negativity toward the target object. Respondents then

indicate the extent of their agreement or disagreement with the statements using 5-point agree/disagree scales. Next, responses to these items are summed to create an overall score for each respondent, and the correlation of each item with the total score is computed. Items that correlate poorly with the total score are deleted, and the remaining items are used to yield a final index of the attitude. Thus, the method of summated ratings does not require a separate group of judges to rate items prior to administering the final scale to the sample of interest.

Another measurement method traditionally used to assess attitudes towards ethnic groups is the method of social distance. According to this method, respondents are given a series of statements reflecting increasingly proximal degrees of social contact with members of a target social group (e.g., shopping in a store where members of the target group shop versus having a member of the target group over for dinner). The respondent's attitude is assessed by determining the most proximal behavior the respondent is comfortable with.

Another method of measuring attitudes is the semantic differential, developed by Charles Osgood and his colleagues. These investigators argued that attitudes can be measured by asking respondents to rate an object on bipolar scales anchored by pairs of adjectives reflecting an evaluative, positive-negative dimension (e.g., good-bad, wise-foolish). Responses to these scales can then be summed to arrive at an overall attitude score for each respondent. Because the same adjective pairs can be used for nearly any attitude object with little if any pretesting, this method has been quite popular among social scientists.

CONTEMPORARY ATTITUDE MEASUREMENT: DESIGNING SINGLE ITEM MEASURES

Unfortunately, it is often not practical to measure attitudes using one of the above methods, because they involve multiple items per attitude and sometimes involve elaborate pretesting. Especially in surveys in which attitudes towards dozens of objects are assessed, it is not feasible to obtain multiple measures of each attitude. Consequently, re-

searchers have turned increasingly often to single item measurement approaches that they believe best reflect the underlying attitude.

An advantage of the multiple item approach is that the particular characteristics of any particular item are unlikely to have a substantial impact on the results of an assessment procedure. But when one relies on only a single item, its characteristics can potentially have tremendous impact on one's conclusions. Therefore, one must design single item measures very carefully in light of one's research goals. Below, we will review the existing literature on designing single items and summarize some of the recommendations supported by these studies (for more details, see Himmelfarb, 1993; Krosnick & Fabrigar, in press; Schuman & Presser, 1981).

OPEN VERSUS CLOSED QUESTIONS

One of the first issues that a researcher must confront when constructing an item is whether to use an open-ended or a closed-ended question. Closed questions provide a list of response options among which a respondent must choose, whereas open questions allow respondents to answer in their own words. Although researchers have typically used closed questions due to the ease of administering and coding them, recent research has suggested that this may sometimes be at a cost to data quality.

One drawback of closed questions is that respondents usually limit their responses to the offered alternatives, even if their optimal answers are ones not mentioned in the question. For example, if asked "What is the most important problem facing the country today: unemployment, inflation, the government's budget deficit, or some other problem?", nearly all respondents will choose one of the first three options rather than generating alternatives. Yet when these same individuals are asked an open-ended question, they are likely to generate a much larger set of important problems. It is possible to avoid this problem, however, if one builds lists of closed question response alternatives based upon pretesting with open-ended questions.

One concern about open-ended questions is that they may disproportionately reflect

concerns or feelings that happen to be on the minds of respondents at the time a question is asked, rather than tapping deeper, more considered beliefs or attitudes. According to a number of studies, momentary SALIENCE of considerations does indeed influence responses to open questions, but salience does so to about the same extent with closed questions. Therefore, this concern does not seem to be a basis for shying away from open questions.

Finally, some research in this area has investigated whether closed questions are more likely than open questions to elicit vacuous responses from respondents who actually have no opinion toward an object. This work has indeed found that respondents are more likely to give answers to closed questions than to open questions when the target is a fictitious object, toward which respondents could not have an opinion. Presumably, the ease of responding to closed questions encourages respondents to provide answers even when these answers are not meaningful. Taken together, then, this literature suggests a number of advantages to open-ended questions, so they may be particularly useful for some attitude measurement tasks.

RATING VERSUS RANKING

When one uses a closed question format, one must choose between a rating format and a ranking format. Rating formats require respondents to report the absolute magnitude of a psychological construct along a continuum (e.g., ranging from "like a great deal" to "dislike a great deal"). Rankings require respondents to order a set of objects according to some criterion (e.g., most preferred to least preferred).

Rating formats are more common in attitude research, presumably because rankings have a number of inherent disadvantages. For example, they yield ordinal and ipsative data, which are less informative and harder to analyze than the interval level data provided by ratings. Furthermore, rankings are a great deal more time consuming and difficult for respondents to complete. Nonetheless, there is some evidence that rankings yield more valid data. Specifically, rankings yield more reliable data than ratings, produce more interpretable factor solutions, and have greater

criterion validity. Furthermore, ratings appear to be more susceptible than rankings to response set biases, due to a failure of some respondents to make fine distinctions among objects in rating tasks. Therefore, rankings have a number of psychometric advantages when one is interested in comparisons of attitudes across sets of objects.

NUMBER OF SCALE POINTS

When using rating scales, one must decide how many scale points to use for each item. Popular single-item attitude measures have ranged from as small as 2-point yes/no or agree/disagree scales to ones as large as 101 points. On one hand, using scales with more points may allow for greater precision. On the other hand, there may be limits to respondents' ability to make fine discriminations, so increasing the number of scale points beyond a certain point may enhance random error rather than enhancing information acquisition. In fact, the various empirical studies on this matter indicate that scales with 5 to 7 points seem to be both more reliable and valid than scales with more or fewer points (*see* Krosnick & Fabrigar, in press).

A related issue is the question of whether one should use rating scales with odd numbers of points (i.e., including a clear midpoint) or whether to use an even number of scale points. Including a midpoint allows respondents with neutral opinions to report them rather than arbitrarily indicating either a positive or a negative attitude. It is also possible, however, that offering the midpoint may reduce validity by providing a response alternative that is easy to select without much thought and therefore discourages respondents from expending the cognitive effort to report a more substantive view based upon their beliefs about an object.

Empirical research on these matters currently supports four conclusions (*see* Bishop, 1987; Krosnick & Fabrigar, in press; Schuman & Presser, 1981). First, respondents are unlikely to report neutral opinions if a midpoint is not explicitly provided, and they are much more likely to report such opinions when a midpoint is offered. Second, the distribution of positive and negative attitudes

expressed can sometimes be very different depending upon whether a response scale includes a midpoint or not. Therefore, the decision regarding whether or not to offer a midpoint may well have a significant effect on substantive research conclusions. Third, a couple of initial studies suggest that the validity of attitude reports is enhanced when a middle alternative is provided. However, additional evidence suggests that middle alternative selection may in fact reflect a desire on the part of some respondents to avoid the cognitive work necessary to formulate and report substantive opinions. Consequently, it is difficult to say at this point whether including a midpoint is desirable.

BRANCHING

Nearly all single-item attitude measures ask respondents to place themselves on a scale ranging from favorable to unfavorable, thus reporting attitude direction and extremity in one step. However, the difficulty of administering long scales during survey interviews over the telephone has led some researchers to employ a branching approach. Accordingly, respondents are first asked whether their attitude is positive or negative or neutral (i.e., direction only). Then, in a follow-up question, respondents expressing positive or negative attitudes are asked how extreme those attitudes are. Respondents who express a neutral attitude initially can be asked a follow-up about whether they would lean in a positive or negative direction. This is called branching because the wording of the follow-up question varies depending upon which initial answer a respondent provides. A number of recent studies indicate that this decomposing of the reporting process into 2 steps enhances the speed and ease with which respondents can report their attitudes as well as the reliability and predictive validity of those reports (Krosnick & Berent, 1993).

VERBAL VERSUS NUMERIC LABELS

It is quite common for researchers to design long rating scales with verbal labels only on the endpoints (e.g., "like a great deal" versus "dislike a great deal") and perhaps also at the

midpoint (e.g., "neither like nor dislike"). Thus, the precise meaning of the other scale points is left implicit. However, a great many scaling studies have been done in recent years to quantify the meaning that people attach to such modifiers as "a great deal," "somewhat," "a little," and so on, so it is conceivable that consensually interpreted and evenly spaced verbal labels can be attached to all points on ratings scales (except, of course, very long ones). A number of studies suggest that this reduces ambiguity in the meanings of those scale points and thereby enhances reliability (see Krosnick & Berent, 1993).

NO-OPINION FILTERS

Regardless of whether open or closed questions are used, a researcher must decide whether to include a no-opinion filter. Such filters sometimes involve asking respondents if they have an opinion toward an object and then only asking what that opinion is if they say they do in fact have one. Alternatively, filtering sometimes involves simply including a "don't know" or "no opinion" category as one of the response options in a single attitude question. Not surprisingly, many more respondents indicate that they have no opinion when a filter is included than when no filter is offered. Again, however, it is not clear whether respondents select a "no opinion" response because they truly have no opinion or because they simply want to avoid the cognitive work necessary to formulate and report an opinion. Because including a no-opinion filter can dramatically alter the proportions of favorable and unfavorable attitudes expressed, the decision about whether or not to include a filter can have important substantive implications.

One set of relevant research has examined whether no-opinion filters enhance validity in the sense of detecting more true non-opinions. This work has focused on the fact that many respondents offer attitudes toward objects that are completely fictitious when asked unfiltered questions. Not surprisingly, including a no-opinion filter does dramatically reduce the number of respondents reporting such attitudes, though filters do not completely eliminate them. This suggests that filtering may enhance validity.

However, other evidence suggests that filtered questions are not superior to unfiltered questions in terms of data quality. First, studies examining test-retest response consistency indicate that reliability does not increase notably when filters are included. Second, associations between attitude reports and other variables do not seem to increase when filters are included. Third, it appears that respondents are especially attracted to no-opinion responses when selecting a substantive response would be especially cognitively burdensome. Therefore, it seems that no-opinion filters may not be desirable in attitude questions.

RESPONSE SETS AND RESPONSE STYLES

Response sets refer to the tendency for an individual to respond to questions in a particular fashion as a result of the structural features of the questions or the data-gathering situation, independent of the content of the questions. In contrast, response styles are response tendencies independent of content that are a function of dispositions of individual respondents, rather than a function of situational factors. A vast literature has investigated response styles and sets, most notably acquiescence and social desirability bias (see Paulhus, 1991).

Acquiescence refers to a tendency to agree with any item, irrespective of its content. This bias often occurs in items using "agree/disagree" or "yes/no" formats. This tendency appears to be a result of both situational demands (e.g., the difference in social status between interviewers and respondents) and PERSONALITY characteristics (e.g., deferential personality). Some researchers believe that acquiescence can be eliminated by administering a large number of items, half of which express positive attitudes and half of which express negative attitudes. Combining across such a set of items may unconfound acquiescence with substantive responses in some cases. However, recent research suggests that this approach may often be unsuccessful, because different items can stimulate different levels of acquiescence, so counterbalancing these levels may be difficult. Fortunately, though, it appears that acquiescence bias can be eliminated by abandoning agree/disagree or yes/no questions and instead using forced choice for-

mat that ask respondents to select one of two opposing substantive points of view.

Social desirability bias refers to a tendency to respond to questions in a way that is socially desirable. This bias can take two forms. In some cases, self-deception is involved, where people are inclined to perceive themselves in inaccurate and socially desirable ways. Alternatively, IMPRESSION MANAGEMENT motives can lead people to distort their presentations of themselves to others in socially desirable directions. Social desirability bias does indeed appear frequently in studies that have looked for it, and its magnitude appears to be a function of both situational and dispositional factors. The impact of social desirability bias may be reduced by conducting pretests to equate response alternatives in terms of their social desirability or to select items that are relatively unaffected by desirability. Additionally, scales measuring social desirability can be used to statistically remove the influence of this bias from other scales. In the case of impression management, assuring anonymity of responses can reduce social desirability biases.

CONCLUSION

Although elaborate procedures for attitude measurement were the norm in the early stages of empirical research in this area, practical considerations have led current researchers to adopt simpler, single-item approaches. However, this move has increased the significance of the structural characteristics of the items used in terms of their potential impact on substantive findings. Fortunately, however, the accumulating body of literature we have reviewed provides numerous insights that aid researchers in designing their measures to tap attitudes as effectively as possible. With continued work in this area will come even clearer recommendations for optimal procedures.

See also: ATTITUDE THEORY AND RESEARCH; IMPRESSION MANAGEMENT; PERSONALITY; SURVEY METHODS.

BIBLIOGRAPHY

Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51, 220-32.

- Dawes, R. M., & Smith, T. L. (1985). *Attitude and opinion measurement*. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. I; pp. 509–66). New York: Random House.
- Himmelfarb, S. (1993). *The measurement of attitudes*. In A. Eagly & S. Chaiken, *The psychology of attitudes* (pp. 23–87). Fort Worth, TX: Harcourt Brace Jovanovich.
- Krosnick, J. A., & Berent, M. K. (1993). Comparisons of party identification and policy preferences: The impact of survey question format. *American Journal of Political Science*, 37, 941–64.
- & Fabrigar, L. R. (in press). *Questionnaire design for attitude measurement in social and psychological research*. New York: Oxford University Press.
- Mueller, D. J. (1986). *Measuring social attitudes: A handbook for researchers and practitioners*. New York: Teacher's College Press.
- Paulhus, D. L. (1991). *Measurement and control of response bias*. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego: Academic Press.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys*. New York: Academic Press.
- Summers, G. F. (1970). *Attitude measurement*. Chicago, IL: Rand McNally.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529–44.

LEANDRE R. FABRIGAR
JON A. KROSNICK

attitude theory and research In an often-quoted passage, Gordon Allport (1935, p. 198) asserted that “the concept of attitude is probably the most distinctive and indispensable concept in contemporary American social psychology.” Few (if any) statements this extreme about social psychology could reasonably be expected to remain valid over a period of nearly 60 years; yet it is arguably true that attitude is still at least one of the most indispensable concepts in social psychology, if not the most indispensable. Despite its prominent status over such an extended

period of time, there is no single definition of attitude that is universally accepted. As noted by Olson and Zanna (1993), attitudes are variously defined in terms of evaluation, affect, cognition, or behavioral dispositions. These different approaches tend to emphasize, respectively, the *evaluation* of attitude objects with respect to their positivity or negativity, the *feelings* of pleasantness or unpleasantness associated with attitude objects, *knowledge* about attitude objects, or predispositions to *behave* positively or negatively towards attitude objects. However, if there is one core feature of attitudes that distinguishes this concept from others, it is that they are evaluative in nature. As Eagly and Chaiken (1993) put it, “Attitude is a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor” (p. 1). Evaluation here can refer to different classes of response, be they affective, cognitive, or behavioral.

Thus the kernel of the definition of attitude is the notion of *evaluative responding* to some entity. Evaluative responses can vary with respect to response category (affective, cognitive, and behavioral) and response mode (verbal or nonverbal), as shown in Table 1 (see Ajzen, 1988). Which category or mode is deemed to be most important in a given research context will of course have implications for ATTITUDE MEASUREMENT.

Table 1. Different types of evaluative response

Response mode	Response category		
	Affect	Cognition	Behavior
Verbal	Expressions of feelings towards attitude object	Expressions of beliefs about attitude object	Expressions of behavioral intentions towards attitude object
Nonverbal	Physiological responses to attitude object	Perceptual responses (e.g. reaction time) to attitude object	Overt behavioral responses to attitude object