

How to Measure Attitudes

From: Simonson, M.R. (1979). Attitude measurement: Why and how. *Educational Technology*, 19, 34-38.

When reviewing the literature that deals with attitude change and instructional technology, it is very apparent that attitude measurement is often done very poorly. Simonson (1979a) commented on the sad state of attitude measurement in the educational technology literature, and more recent reviews have not revealed any improvements in testing methodology (Simonson & Maushak, 1995). The move to more qualitative-based research (see 40.2) and measurement has not changed this situation, and may be contributing to a decline in the quality of attitude testing (see 6. 1).

Before beginning this discussion of attitude measurement, it is important once again to establish a frame of reference for this review. Attitude research is largely conducted by those called *empiricists*, *objectivists*, and *reductionists*. They tend to take the approach of the scientific empiricist who believes that there are laws of nature that the scientist must discover. The vast body of attitude and attitude-change literature is authored by those attempting to "discover the answer" and to determine "truth." These researchers usually apply quantitative approaches in their research designs (see 39.4).

Those advocating naturalistic inquiry (see 40.2) may be uncomfortable with the approach taken by this chapter. A general question often asked by qualitative researchers, "What is going on here", does not readily translate to results of the kind summarized in this chapter and the type of measurement techniques recommended next. Certainly, it would be unwise to discount qualitative techniques for examining the critical issues of the field. Just as certainly, the vast body of literature about attitudes and attitude measurement were generated by scientists who applied quantitative approaches to measurement.

Problems with attitude measurement are of three types. First, researchers are not clearly defining their attitude variables. In other words, they are not operationalizing the constructs that they are setting out to measure. This problem is heightened by the failure of many to include attitude hypotheses or research questions in their research designs. Rather, attitude constructs are often included as post-hoc components of research studies. Qualitative researchers also tend to show little interest in attitude constructs.

Second, attitudes are not measured well. Certainly, quantitative measurement of attitudes has evolved into a fairly exact process (Henerson, Morris & Fitz-Gibbon, 1987). However, reports about the methods used to develop measures of attitudes are reported in only a minority of the research studies found in the literature. Simonson (1979a) reported that only 50% of the studies reviewed reported on the validation of attitude measures, and only 20% reported descriptive information about their attitude tests. Most measures then, and today, tended to be locally prepared and used only once-in the specific study reported. Researchers who were otherwise extremely careful to standardize their achievement measures did not do the same for their tests of attitudes.

One alarming trend was the use of single items to measure attitudes. Researchers reported using a single item to determine a person's attitude (e.g., Do you like chemistry?), and then used the responses to this question in powerful statistical analyses. Apparently, reliability and validity concerns were not worrisome to these researchers.

Finally, attitude measurement has tended to be of only peripheral importance to researchers. Often, as stated above, attitudes are relegated to post-hoc examinations, often conducted without controls or design considerations being taken into account. As a matter of fact, it is obvious that attitude study is not an area of interest or importance in mainstream instructional technology research. Of the hundreds of studies published in the literature of educational communications and technology since Simonson's review (1979a) of attitude research, less than 5% examined attitude variables as a major area of interest. This lack of interest was discouraging, especially when contrasted with the wealth of attitude research in the literature of social psychology.

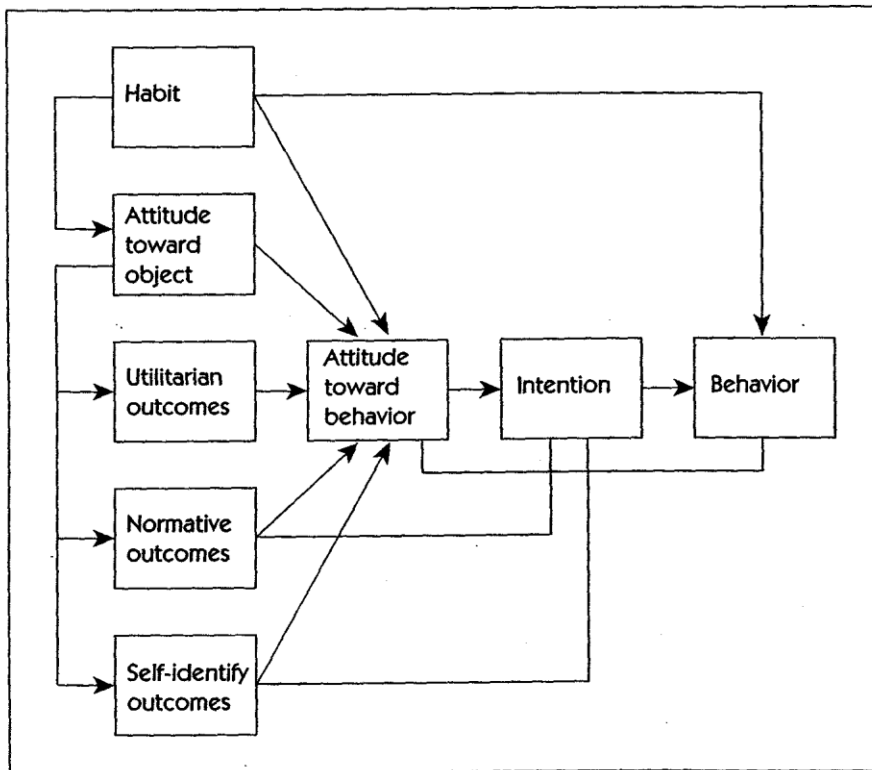


Figure34-1. A composite model of the attitude-behavior relationship (Eagly & Chaiken, 1993).

One reason attitudes may be studied so rarely is the difficulty many have in clearly identifying how attitudes should be measured. The characteristics of attitude contribute to this perception of difficulty, as does the recent move away from quantitative research procedures. In a recent review of the indexes of five textbooks dealing with methods of qualitative analysis, the term *attitude* was not found in any, even in the recently published *Handbook of Qualitative Research* (Denzin & Lincoln, 1994).

Since attitudes are defined as latent, and not observable in themselves, the educator must identify some action that would seem to be representative of the attitude in question so that this behavior might be measured as an index of the attitude. This characteristic of attitude measurement is justifiably one of the most criticized of this area of educational evaluation. However, there are several generally recognized procedures used to determine quantitatively an individual's, or group's, attitude toward some object or person. It is those procedures that are described below. Two excellent sources for information on attitude measurement should be reviewed by those interested in quantitatively testing for attitudes. First is Himmelfarb's (Eagly & Chaiken, 1993) comprehensive review of the basic concepts and ideas behind attitude measurement. It also contains an explanation of the various techniques for quantifying attitude positions. Himmelfarb's discussion is a scholarly explanation of attitude measurement.

For those interested in more specific procedures for attitude measurement, Henerson, Morris, and Fitz-Gibbon's (1987) manual is excellent. It would be unfair to call the manual a *cookbook* because it is more than that. It does contain step-by-step, cookbook-like, procedures for validly and reliably developing measures of attitudes. It is a must reference for those interested in quantifying attitudes as part of a research study, but who do not wish to become attitude measurement experts. Henerson, Morris, and Fitz-Gibbon even include a section labeled "alternative approaches to collecting attitude information" designed to appeal to the qualitative researcher.

34.5.1 Characteristics of Quantitative Attitude Measurement

Before procedures for measuring attitudes are discussed, there are several general characteristics of measurement that should be considered in order to determine if an evaluation technique is an effective one. Good tests have these characteristics. Basically, a quantitative approach to attitude measurement requires that measures be:

- Valid. The instrument must be appropriate for what needs to be measured. In other words, a valid test measures the construct for which it is designed. A test of "attitude toward chemistry" will have items that deal directly with the concept of chemistry.
- Reliable. The measure should yield consistent results. In other words, if people were to take a reliable test a second time, they should obtain the same, or nearly the same, score as they got the first time they took the test, assuming no changes occurred between the two testings.
- Fairly simple to administer, explain, and understand. Generally, the measures that yield a single score of an attitude position epitomize the intent of this characteristic, although the single score may be deficient in meeting the intent of other characteristics of good measurement. Most tests of single attitudes have about 10 to 30 items, are valid, and have reliability estimates above .80.
- Replicable. Someone else should be able to use the measure with a different group, or in a different situation, to measure the same attitude. Replicable tests of attitude should be usable in a variety of situations. In other words, a test of computer anxiety should measure the existence of that construct in college students, parents, elementary school students, and even stockbrokers.

34.5.2 Categories of Attitude Measurement Techniques

There are four widely used and accepted categories, or approaches, for collecting attitude information. These approaches are:

- Self-reports, where the members of a group report directly about their own attitudes. Self-reports include all procedures by which a person is asked to report on his or her own attitudes. This information can be provided orally through the use of interviews, surveys, or polls, or in written form through questionnaires, rating scales, logs, journals, or diaries. Self-reports represent the most direct type of attitude assessment and should be employed, unless the people who are being investigated are unable or unwilling to provide the necessary information. Questions like "How do you feel about XT" where X is the attitude construct under investigation are often asked in self-reports.
- Reports of others, where others report about the attitudes of a person or group. When the people whose attitudes are being investigated are unable or unlikely to provide accurate information, others can be questioned using interviews, questionnaires, logs, journals, reports, or observation techniques. Parents of children can be asked how their children feel about X, where X is the attitude construct under investigation.
- Sociometric procedures, where members of a group report about their attitudes toward one another. Sociometrics are used when the researcher desires a picture of the patterns within a group. Members of groups can be asked questions like "Who in your group fits the description of XT" where X is the attitude position being studied.
- Records, which are systematic accounts of regular occurrences, such as attendance reports, sign-in sheets, library checkout records, and inventories. Records are very helpful when they contain information relevant to the attitude area in question. For example, when a researcher is trying to determine if a schoolwide program to develop a higher level of school pride is working, the school's maintenance records might give an index of the program's effectiveness. If school pride is improving, then vandalism should decline, and maintenance costs should be lower. The amount of trash picked up from the school's floors might yield relevant information, too. Students who have 'school pride are less likely to throw trash on the floor.

Within each of these categories, there are strategies for measuring attitude-related behaviors. Most commonly, attitude measurement is accomplished by one of the following techniques:

- Questionnaires and rating scales. Questionnaires and rating scales are instruments that present information to a respondent in writing and then require a written response, such as a check, a circle, a word, a sentence, or several sentences. Attitude rating scales are special kinds of questionnaires. They are developed according to strict procedures that ensure that responses can be summed to yield a single score representing one attitude. Questionnaires and rating scales are often used because they permit anonymity, permit the responder time to answer, can be given to many people simultaneously, provide uniformity across measurement situations, permit relatively easy data interpretation, and can be mailed or administered directly. Their main disadvantage is they do not permit as much flexibility as do some other techniques.

- Interviews. Interviews are face-to-face meetings between two or more people in which the respondent answers questions. A survey is a highly structured interview. Often surveys are conducted over the telephone, an approximation of face-to-face interviewing. A poll is a headcount. Respondents are given a limited number of options and asked to select one. For example, word-of-mouth procedures, such as interviews, surveys, and polls, are useful because they can be read to people who cannot read or who may not understand written questions. They guarantee a relatively high response rate, they are best for some kinds of information especially when people might change their answers if responses were written, and they are very flexible. There are two major problems with interviews. First, they are very time consuming. Second, it is Possible that the interviewer may influence the respondent.
- Written reports, such as logs, journals, and diaries. Logs, journals, and diaries are descriptions of activities, experiences, and feelings written during the course of the Program. Generally they are running accounts consisting of many entries prepared on an event, on a daily or weekly basis. The main advantage of this approach is that reports Provide a wealth of information about a person's experiences and feelings. The main problem is in extracting, categorizing, and interpreting the information. Written reports require a great deal of time by both the respondent and the researcher.
- Observations. These procedures require that a person dedicate his or her attention to the behaviors of an individual or group in a natural setting for a certain period of time. The main advantage of this approach is its increased credibility when pretrained, disinterested, unbiased observers are used. Formal observations often bring to attention actions and attitudes that might otherwise be overlooked. Observations are extremely time consuming, and sometimes observers produce discomfort in those they are observing. The presence of an observer almost always alters what is taking place in a situation.

A specific strategy for attitude measurement should be chosen which is appropriate for the type of attitude construct of interest, the type of learner, and the situation being examined (Henerson, Morris & Fitz-Gibbon, 1987). The procedures summarized above are those most often used. Others strategies are available, but attitude researchers are cautioned to select a technique appropriate to their research questions and a technique they are competent to carry out.

34.5.3 A Recommended Process for Attitude Measurement

Attempts at measurement, including the evaluation of attitude, require that a systematic process be followed. Using structured procedures increases the likelihood of an effective measurement taking place. Guidelines for attitude measurement usually recommend that at least six steps be followed (Henerson, Morris & Fitz-Gibbon, 1987):

1. Identify the construct to be measured. A construct is simply defined as the attitude area of interest. It is usually best to identify specific attitude constructs. Narrow attitude constructs such as "desire to take a course in chemistry" are probably better than "liking of chemistry," and "importance of knowing about the chemical elements" might be an even better attitude to measure. A learner can conceivably have an attitude position toward any object, situation, or person. When mediated instruction is designed, those attitudes that are important to the learning

activity should be clearly identified and defined. An example of an attitude that an instructional developer might be interested in would be "attitude toward learning about titrations; by video."

2. *Find an existing measure of the construct.* Once a certain attitude construct has been identified, an attempt should be made to locate an instrument that will measure it. Published tests are the first choice for measuring attitudes because they have usually been tried out in other instructional situations and include some statement of test validity and reliability. Additionally, instructions for administration of published tests often are available. The use of standardized measures simplifies the job of attitude evaluation.

The most obvious disadvantage to using a predesigned test is that it may not be evaluating the specific attitude being studied. Even if this is the case, it may sometimes be possible to extract valuable information from an instrument designed to test an attitude position similar to the one of specific instructional interest.

Possibly the best source of published tests is the research literature. Researchers who have conducted attitude research will often have developed or identified measures of their dependent variables that can be used in new experimental situations. If the research literature does not yield an appropriate measure of an attitude construct, then published indexes of tests can be reviewed. *Mental Measurements Yearbooks*, and *Tests in Print* are general sources for tests of all kinds. Often, standardized tests, such as those listed in general indexes, can be used to provide direction to the development of more specific attitude tests.

3. *Construct an attitude measure.* If no existing test of the relevant attitude is available, and a quantitative measure is needed, then it is necessary to construct a new test. Of the many types of attitude measurement possible, one widely used technique that seems to possess most of the characteristics of a good measure is the Agreement, or Likert-type, Scale. This technique involves the use of statements about the attitude that are either clearly favorable or unfavorable. Each student responds to each test item according to his or her perceived attitude "intensity" toward the statement. Often, students are asked to answer test items using a five-point scale that has responses varying in the amount of agreement to the statement from "strongly disagree" to "strongly agree." Advantages of this technique are ease of scoring and ease of summarizing the information obtained.

When a test is constructed, it is critical that validity and reliability information be collected for the measure. Of these two concepts, validity (i.e., appropriateness of instrument) is the most difficult to determine.

Validity for a test depends on a number of factors, such as the type of test and its intended use. Basically, there are four categories of validity:

- *Construct validity.* This concept refers to the extent to which the measure accurately represents the attitude construct whose name appears in its title. This can be determined by:

a. *Opinions of experts.* Experts are asked to review the test, and their reactions to it are used to modify the test, or if they do not have negative reactions, then the test is considered valid.

b. *Correlations to other measures of the same construct.* In some situations there may be other, often more complex, measures of the same variable that are available. Validity can be determined by asking a sample of learners to complete both the complex and the simpler versions and then correlating their scores. This procedure was used by Maurer (1983) when he validated his Computer Anxiety Index by correlating student's scores on it to Spielberger's (1970) much more complex and expensive State Anxiety Index.

c. *Measures of criterion group subjects (those who have been proved to possess the construct).* Maurer (1983) validated his computer anxiety index also using this technique. He observed learners and identified those who possessed the obvious characteristics of the computer anxious person. He then examined their Computer Anxiety Index scores and determined that their Index scores were also high, indicating that it was validly measuring computer anxiety.

d. *Appeals to logic.* Many times, particularly when the attitude can be easily defined, audiences will accept an instrument as logically related to the attitude, as long as they know it will be administered fairly.

- *Content validity.* This refers to the representativeness of the sample of questions included in the instrument. Content validity is usually determined by careful analysis of the items in the test. There is no simple process to determine content validity other than a close, thoughtful examination of each item separately, and all items collectively.
- *Concurrent validity.* This refers to the agreement of a test with another test on the same topic that was administered at approximately the same time. Concurrent validity is determined by correlating the results of the two parallel measures of the same attitude. This correlation coefficient is reported as an index of concurrent validity. For example, if an attitude test measuring "willingness to study chemistry" was administered and scores were obtained, it could be correlated to the instructor's assessments of the "completion rate of chemistry homework assignments" in order to determine an index of concurrent validity.
- *Predictive validity.* This refers to how well a measure will predict a future behavior, determined by comparing the results of an attitude test to a measure of behavior given in the future. This type of validity is usually expressed by a correlation coefficient found by comparing the results of two measures. For example, the results of an attitude test that measured "willingness to take additional chemistry courses" could be compared to actual course enrollment figures to determine the predictive validity of the attitude test.

Determining validity is not simple, however. Every educator who constructs a test of any type should be acutely aware of the need to develop valid instruments. Because there is no single, established method for determining validity, the test originator should exercise great care when constructing, administering, and interpreting tests.

Reliability is the ability of a measure to produce consistent results. It is usually less difficult to determine than validity. Reliability also refers to the extent to which measurement results are free of unpredictable kinds of error.

There are several methods of determining reliability that can be easily used by the attitude test developer. The "Test-Retest" method involves a second administration of the instrument to the target group and correlation of the results. The "Split-Half" method uses a random division of the instrument into two halves. Results from each half are correlated and reported as a reliability coefficient. "Alternate-Form" reliability involves the correlation of the results of two parallel forms of tests of the same attitude construct. In this method, each subject takes each form, and the resulting correlation is reported as a reliability estimate. Internal consistency reliability is a determination of how well the items of an attitude test correlate with one another. Measures of internal consistency, such as the Cronbach-alpha, are often used by attitude test developers (Ferguson, 1971).

Both the Test-Retest and Alternate Form techniques will yield a score between -1.00 and +1.00. The higher the number, the more reliable the test. Reliability coefficients above .70 are considered respectable. Scores above .90 are not uncommon for standardized attitude tests. As with validity, the results of reliability estimation should be reported to the test's consumer (Anastasi, 1968; Cronbach, 1970; Talmage, 1978; Henerson et al., 1987).

4. *Conduct a pilot study.* While it is possible to obtain validity and reliability data during the actual testing portion of the instructional activity, it is preferable to administer attitude instruments to a pilot audience before any formal use is undertaken. This is done to obtain appropriate data, and to uncover minor and potentially troublesome administrative problems such as misspellings, poor wording, or confusing directions. A group of learners similar to those who are the target group for the attitude test should be given the measure. Results should be used to revise the test and to determine validity and reliability information.

5. *Revise tests for use.* Results of pilot testing are used to revise, and refine, attitude instruments. Once problems are eliminated, the measure is ready to be used with its intended target audience.

6. *Summarize, analyze, and display results.* After testing is completed, the resulting data should be interpreted. Attitude test results are handled similarly to any other quantitative test information. Attitude responses should be summarized, analyzed, and displayed in such a manner that results are easily and quickly understood by others.

Descriptive statistics should be reported about the attitude test results. Most often, means, standard deviations, and the range of scores should be reported. In experimental situations, tests of inference are often performed using the results of attitude tests. Most attitude test results can be analyzed using standard parametric tests, such as *t* tests and analysis-of-variance tests. However, attitude data about instructional method or content area are often useful even if they are only averaged and compared to other averages. In other words, did the class average change for "Attitude Toward the Happiness of People in India" after viewing the video, or did the class react favorably to "The Importance of Wearing Seat-Belts" after participating in a hypermedia computer lesson?

Displaying data is another effective method of analysis. Charts, graphs, and bar diagrams are examples of data display techniques that are useful in assisting the reader in developing an understanding of what test results indicate. Whatever the process, the developer of an attitude test should make every effort to decipher the results of the measure and to explain apparent conclusions and implications derived from the test.

Attitude measurement is certainly not an exciting topic, and may be of less interest than other issues discussed in this chapter. However, attitude testing specifically, and identifying attitudes generally, are apparently not understood and probably not valued by many educational technology researchers. Certainly, the trend toward more qualitative approaches to investigation may convince some that attitude measurement, and even attitude identification, are irrelevant to the important issues of the field. However, those who are still approaching research questions from an objectivist perspective will want to be sure that they are correctly following the accepted principles of measurement.