
Data Collection Techniques

GEORGE A. MORGAN; ROBERT J. HARMON

In this column we provide a context for many of the types of data collection techniques used with human participants. We will not discuss methods for assessing diagnostic status, but we will provide some information about developing or evaluating a questionnaire, test, or other data collection technique.

Research approaches or designs are approximately orthogonal to the techniques of data collection, and thus, in theory, any type of data collection technique could be used with any approach to research. However, some types of data collection are more commonly used with the experimental approaches. Others are more common with comparative or associational (survey) approaches, and still others are more common in qualitative research.

Table 1 gives an approximation of how common each of several data collection techniques are within each of these three major groupings of research approaches. Note that we have ordered the data collection techniques along a dimension from observer/researcher report to self-report measures. The observer report end includes observations and physiological recordings that are probably less influenced by the participants' desire to look good, but they are affected by any biases the observer may have. Of course, if the participants realize that they are being observed, they may not behave naturally.

At the other end of this dimension are measures based on self-reports of the participants, such as interviews and questionnaires. In these cases, responses are certainly filtered through the participants' eyes and are probably heavily influenced by factors such as social desirability.

Concern about faulty memories or socially desirable responses lead researchers, especially those who use experiments, to be suspicious about the validity of the self-reports. On the other hand, observer reports are not necessarily valid measures. For example, qualitative researchers point out that cultural biases may lead observers to misinterpret their observations. In general, it is advisable to select instruments that have been used

in other studies if they have been shown to be reliable and valid with the planned types of participants and for purposes similar to that for the planned study.

TYPES OF DATA COLLECTION TECHNIQUES

Direct Observation

Many researchers prefer systematic, direct observation of behavior as the most accurate and desirable method of recording the behavior of children. Using direct observation, the investigator observes and records the behaviors of the participants rather than relying on reports from parents or teachers. Observational techniques vary on several dimensions.

Naturalness of the Setting. The setting for the observations can vary from natural environments (such as a school or home) through more controlled settings (such as a laboratory playroom) to highly artificial settings (such as a physiological laboratory). Qualitative researchers do observations almost exclusively in natural settings. Quantitative researchers use the whole range of settings, but some prefer laboratory settings.

Degree of Observer Participation. This dimension varies from situations in which the observer is a participant to situations in which the observer is entirely unobtrusive. Most observations, however, are done in situations in which the participants know that that observer is observing them and have agreed to it. Such observers attempt to be unobtrusive, perhaps by observing from behind a one-way mirror.

Amount of Detail. This dimension goes from global summary information (such as overall ratings based on the whole session) to moment-by-moment records of the observed behaviors. Ob-

TABLE 1
Data Collection Techniques Used by Research Approaches

Data Collection Techniques	Research Approach		
	Quantitative Research		
	Experimental & Quasi-Experimental	Comparative, Associational, & Descriptive Approaches	Qualitative Research
Research report measures			
Physiological recordings	++	+	-
Coded observations	++	++	+
Narrative observations	-	+	++
Participant observations	-	+	++
Other measures			
Standardized tests	+	++	-
Archival measures/documents	-	+	++
Content analysis	-	+	++
Self-report measures			
Summated attitude scales	+	++	-
Standardized personality scales	+	++	-
Questionnaires (surveys)	+	++	+
Interviews	+	++	++
Focus groups	-	-	++

Note: Symbols indicate likelihood of use (++ = quite likely; + = possibly; - = not likely).

viously, the latter provides more detail, but it requires considerable preparation and training of observers.

Standardized Versus Investigator-Developed Instruments

Standardized instruments cover topics of broad interest to a number of investigators. They usually are published, are reviewed in a *Mental Measurements Yearbook* (1938–2000), and have a manual that includes norms for making comparisons with broader samples and information about reliability and validity.

Investigator-developed measures are ones developed by a researcher for use in one or a few studies. Such instruments also should be carefully developed, and the report of the study should provide evidence of reliability and validity. However, there usually is no separate manual for others to buy or use.

The next several sections utilize this distinction. Some tests, personality measures, and attitude measures are developed by investigators for use in a specific study, but there are many standardized measures available. There are standardized question-

naires and interviews, for example, those for diagnostic classification, but most are developed by an investigator for use in a particular study.

Standardized Tests

Although the term test is often used quite broadly to include personality and attitude measures, we define the term more narrowly to mean a set of problems with right or wrong answers. The score is based on the number of correct answers. In standardized tests, the scores are usually translated into some kind of normed score that can be used to compare the participants with others and are referred to as norm referenced tests. For example, IQ tests were normed so that 100 was the mean and 15 was the standard deviation.

Achievement Tests. These are designed to measure knowledge gained from educational programs. There should be reliability and validity evidence for the type of participants to be studied. Thus, if one studies a particular ethnic group, or children with developmental delays, and there exists an appropriate

standardized test, use it. In addition to saving time and effort, the results of your study can be compared with those of others using the same instrument.

When standardized tests are not appropriate for your population or for the objectives of your study, it is better to construct your own test or re-norm the standardized one rather than use an inappropriately standardized one. If you develop your own test, you should determine reliability and validity before using it.

Aptitude Tests. In the past, these were often called intelligence tests, but this term is less used now because of controversy about the definition of intelligence and to what extent it is inherited. Aptitude tests are intended to measure general performance or problem-solving ability. These tests attempt to measure the participant's ability to solve problems and apply knowledge in a variety of situations.

In a quasi-experiment or a study designed to compare groups that differ in diagnostic classification (e.g., ADHD), it is often important to control for group differences in aptitude. This might be done by matching on IQ or statistically (e.g., using analysis of covariance with IQ as the covariate).

The most widely used individual aptitude tests are the Stanford-Binet and the Wechsler tests. The Stanford-Binet test produces an intelligence quotient (IQ), which is derived by dividing the obtained mental age by the person's actual or chronological age. A trained psychometrician must give these tests to one person at a time, which is expensive in both time and money. Group aptitude tests, on the other hand, may be more practical for use in research in which group averages are to be used.

Standardized Personality Inventories

Personality inventories present a series of statements describing behaviors. Participants are asked to indicate whether the statement is characteristic of their behavior, by checking yes or no or by indicating how typical it is of them. Usually there are a number of statements for each characteristic measured by the instrument. Some standardized inventories measure characteristics of persons that might not strictly be considered personality. For example, inventories measure temperament (e.g., Child Temperament Inventory), behavior problems (e.g., Child Behavior Checklist), or motivation (e.g., Dimensions of Mastery Questionnaire). Notice that these instruments have various labels, i.e., questionnaire, inventory, and checklist. They are said to be standardized because they have been administered to a wide variety of respondents, and a manual provides information about these norm groups and about the reliability and validity of the measures.

These "paper-and-pencil" inventories are relatively inexpensive to administer and objective to score. However, the validity of a personality inventory depends not only on respondents' ability to read and understand the items but also on their understanding of themselves and their willingness to give frank and honest answers. Although good personality inventories can provide useful information for research, there is clearly the possi-

bility that they may be superficial or biased, unless strong evidence is provided for construct validity.

Another type of personality assessment is the projective technique. These measures require an extensively trained tester, and therefore they are expensive. Projective techniques ask the participant to respond to unstructured stimuli (e.g., ink blots or ambiguous pictures). It is assumed that respondents will project their personality into their interpretation of the stimulus, but, again, one should check for evidence of reliability and validity.

Summated (Likert) Attitude Scales

Likert initially developed this method as a way of measuring attitudes about particular groups, institutions, or concepts. Researchers often develop their own scales for measuring attitudes or values, but there are also a number of standardized scales to measure attitudes such as social responsibility. The term Likert scale is used in two ways: for the summated scale that is discussed below and for the individual items or rating scales from which the summated scale is computed. Likert items are statements related to a particular topic about which the participants are asked to indicate whether they strongly agree, agree, are undecided, disagree, or strongly disagree. The summated Likert scale is constructed by developing a number of statements about the topic, usually some of which are clearly favorable and some of which are unfavorable. To compute the summated scale score, each type of answer is given a numerical value or weight, usually 1 for strongly disagree, up to 5 for strongly agree. When computing the summated scale, the weights of the negatively worded or unfavorable items are reversed so that strongly disagree is given a weight of 5 and strongly agree is 1.

Summated attitude scales, like all other data collection tools, need to be checked for reliability and validity. Internal consistency reliability would be supported if the various individual items correlate with each other, indicating that they belong together in assessing this attitude. Validity could be assessed by determining whether this summated scale can differentiate between groups thought to differ on this attitude or by correlations with other measures that are theoretically related to this attitude.

Questionnaires and Interviews

These two broad techniques are sometimes called survey research methods, but we think that term is misleading because questionnaires and interviews are used in many studies that would not meet the definition of survey research. In survey research a sample of participants is drawn (usually using one of the probability sampling methods) from a larger population. The intent of surveys is to make inferences describing the whole population. Thus the sampling method and return rate are very important considerations. Salant and Dillman (1994) provide an excellent source for persons who want to develop and conduct their own questionnaire or structured interview.

Questionnaires are any group of written questions to which participants are asked to respond in writing, often by checking or circling responses. Interviews are a series of questions presented orally by an interviewer and are usually responded to

orally by the participant. Both questionnaires and interviews can be highly structured, but it is common for interviews to be more open-ended, allowing the participant to provide detailed answers.

Open-ended questions do not provide choices for the participants to select; rather, they must formulate an answer in their own words. This type of question requires the least effort to write, but they can be difficult to code and they are demanding for participants, especially if responses have to be written or concern issues that the person has not considered.

Closed-ended items ask participants to choose among discrete categories and select which one best reflects their opinion or situation. Questions with ordered choices are common on questionnaires and are often similar to the individual items in a personality inventory or a summated attitude scale. These questions may in fact be single Likert-type items which the respondent is asked to rate from strongly disagree to strongly agree.

Two main types of interviews are telephone and face-to-face. Telephone interviews are almost always structured and usually brief, whereas face-to-face interviews can vary from what amounts to a highly structured, oral questionnaire with closed-ended answers to in-depth interviews, preferred by qualitative researchers. In-depth interviews are usually tape-recorded and transcribed so that the participant's comments can be coded later. All types of interviews are relatively expensive because of their one-to-one nature.

SUMMARY

We have provided an overview of techniques used to assess variables in the applied behavioral sciences. Most of the methods are used by both quantitative/positivist and qualitative/constructivist researchers but to different extents. Qualitative researchers prefer more open-ended, less structured data collection techniques than do quantitative researchers. Direct ob-

servations of participants is common in experimental and qualitative research; it is less common in so-called survey research, which tends to use self-report questionnaires. It is important that investigators use instruments that are reliable and valid for the population and purpose for which they will be used. Standardized instruments have manuals that provide norms and indexes of reliability and validity. However, if the populations and purposes on which these data are based are different from yours, it may be necessary for you to develop your own instrument or provide new evidence of reliability and validity.

The authors thank Nancy Plummer for manuscript preparation. Parts of the column are adapted, with permission from the publisher and the authors, from Gliner JA and Morgan GA (2000), *Research Methods in Applied Settings: An Integrated Approach to Design and Analysis*. Mahwah, NJ: Erlbaum. Permission to reprint or adapt any part of this column must be obtained from Erlbaum.

References

- Mental Measurements Yearbooks (1958–2000), Lincoln, NE: Buros Institute of Mental Measurements, University of Nebraska, Vols 1–14
- Salant P, Dillman DA (1994), *How to Conduct Your Own Survey*. New York: Wiley

Dr. Morgan is Professor of Education and Human Development, Colorado State University, Fort Collins; and Clinical Professor of Psychiatry, University of Colorado School of Medicine, Denver. Dr. Harmon is Professor of Psychiatry and Pediatrics and Head, Division of Child and Adolescent Psychiatry, University of Colorado School of Medicine, Denver.