PART ONE

CONCEPTS

INTRODUCTION

In this chapter we will

- Explain the purpose and function of a social science instrument.
- Describe nomenclature used to describe instruments.
- List and describe the components of an instrument.
- Outline the steps in the instrument construction process.

We are living in a time characterized as the information age, and we encounter data-gathering instruments in all facets of our lives. For example, we are familiar with polls that gather information about political preferences and voting behaviors. Surveys of potential voters try to predict who will be elected or what proposition will pass. Media commentators remind us of the margin of error associated with a survey or note that an election is still too close to call.

The proliferation of instruments to provide data and information for decision making is not unique to political polls. Survey questionnaires can be used to obtain factual information and to assess attitudes and beliefs across a variety of topics and groups. For example, surveys can assess consumer behaviors, client satisfaction with services, employee attitudes, and the general public's values and beliefs. The federal government is perhaps the greatest consumer and user of survey questionnaires, as these instruments are used to collect data about such topics as criminal activity, educational needs, services

to the mentally ill, and health care utilization, not to mention the data for the U.S. Census.

You may be familiar with other forms of measurement instruments as well. For example, many organizations conduct annual evaluations of employee work performance. Often these evaluation instruments use scales to rate job performance on a number of attributes, such as attendance, ability to work with others, or the ability to complete work tasks in a timely manner. Although the intended purpose of a checklist or rating instrument such as an employee evaluation is different from the purpose of a survey questionnaire, the same principles are used in constructing these instruments.

As you read this introduction, you are likely thinking of instruments that you have personally used or have been asked to complete, including instruments related to your work or study. And as a reader of this text, you are also interested in creating an instrument, perhaps in conjunction with a research project or job-related activity.

Social science instruments are tools for the collection and measurement of data, and the purpose of this book is to describe how good instruments are constructed. We use the adjective *good* because there are indeed standards and guidelines that can produce an efficient and effective instrument rather than a mediocre one. Our goal is to help you identify the components that make for a "good" instrument, one that provides trustworthy information. As you will see, no instrument is perfect, as there are many ways to pose a question. However, by being aware of the conditions that affect your results, you can create an instrument that effectively meets your need for information.

We believe that instrument construction is as much an art as it is a science. Research has demonstrated that some approaches to instrument development can increase the accuracy and dependability of responses. And statistical tests can be used to measure the consistency of people's responses. Nevertheless, much of what you will do involves common sense, interpersonal skills, and to a degree even creativity. You should not shy away from developing your own instrument because you believe the process is too technical.

We compare instrument construction to the activities of a painter creating a work of art. The first stage involves conceptualizing the project, a purely mental process in which the artist begins to visualize the subject and what it should look like when completed. The designer of a questionnaire goes through a similar process as he or she defines the purpose of the study, obtains information about the subject to be studied, and contemplates the items that might be included. The next step is preparatory. The artist may develop a number of sketches to define the subject, conceptualize the composition of the painting, and experiment with the use of color. Similarly, through an iterative and interactive process, the creator

of a questionnaire will draft the instrument, test the items, and revise and modify both individual items and the instrument format.

An artist may set the canvas aside from time to time to evaluate its progress. During such periods, the artist may experiment with subtle changes in composition, colors, or techniques for applying the paint. The questionnaire designer will also fine-tune the instrument, often in response to feedback from content experts or potential users. Items may be reworded, different item types may be tried, and the layout of items within the instrument may be reorganized. In the end both painter and questionnaire designer must reach a point where they are comfortable that their goals have been attained and the product is ready to be unveiled.

Instrumentation

An *instrument* is a mechanism for measuring phenomena, which is used to gather and record information for assessment, decision making, and ultimately understanding. An instrument such as a questionnaire is typically used to obtain factual information, support observations, or assess attitudes and opinions. For example, a survey may ask respondents to list the type of soap they purchase (factual information), recall how often they purchased the item in the past year (an observation), and consider the factors that influenced their purchase, such as smell, touch, or appearance (attitudes toward the product). The term *subjective* describes information that originates within an individual and is reflected by items that measure attitudes, feelings, opinions, values, and beliefs. Information that is *objective* attempts to be free of personal interpretation and is typified by data that are observable.[1]

Some instruments consist of all objective items, like the medical history questionnaire at the end of this chapter. Respondents are asked to provide demographic information such as their weight, height, and age as well as information about their physical health, such as allergies and previous illnesses. Conversely, some instruments are designed to obtain primarily subjective responses, such as information about political preferences. Although political polls include objective demographic questions, the body of the instrument consists of items that require the respondent to express an opinion or attitude.

In the social sciences most instruments are of the paper-and-pencil variety, meaning that the individual completing the instrument is expected to record information on a form. Even when other media are to be used, a paper-and-pencil instrument will probably need to be developed initially. For example, a marriage counselor might use videotape to record the interactions between a husband and wife. However, a written instrument might then be applied to count the number

of times a particular word or phrase or body gesture is used or to rate a type of interaction. A marketing survey organization might collect information over the telephone but record responses with a paper-and-pencil instrument. More and more, instruments are being constructed that can be completed on a computer. This medium has the advantage of reaching many people quickly and the software often allows the user to tabulate results easily. However, even when the process of collecting or entering information varies—pen, pencil, keyboard, mouse, or verbal encoding—the basic construction of the instrument remains the same, regardless of the medium.

We can categorize instruments in several ways. One approach is based on a *mode of administration:* who is responsible for completing the instrument? Some instruments, such as polls or medical history questionnaires, rely on *self-report,* where the respondent supplies the information directly. For example, many interviews, telephone surveys, and some psychometric assessments are initiated by a second party; nonetheless, because the information is provided directly by the respondent, we classify these instruments as self-report.

Another mode of administration is *observation,* where information about an individual is obtained by someone external. Examples of observation instruments are employee performance appraisals, student assessments of faculty, and behavior checklists. These instruments collect data about characteristics intrinsic to an individual even though the respondent is not queried directly. For example, as part of conducting a screening process for attention deficit disorder, a teacher might use a behavior checklist to record the number of times a student is off task.

An external rater or observer is also used when information is needed about things rather than people. For example, a medical records director might develop a checklist to assess whether all the required documentation has been filed in a clinical chart. Similarly, when researchers need information from existing sources, such as medical, personnel, or student records, they may need an instrument designed specifically for data extraction. External raters usually require training to ensure that they obtain the data in the manner required by the instrument developer and are consistent with other observers or raters.

Some instruments use a combination of approaches and formats. The Achenbach questionnaires (Achenbach, 1991) consist of three instruments that provide information about a child's behaviors and emotional state. The Youth Self-Report (YSR), which has various item formats, such as checklists, fill in the blanks, and rating scales, is completed directly by the youngster who is the focus of evaluation. As its title implies, the mode of administration is self-report. The Child Behavior Checklist (CBCL) is completed by the child's parent or guardian and consists essentially of the same items as the YSR, although worded slightly differently. The Teacher Report Form (TRF) attempts to measure the same attributes, but the

questions are formulated to address the child's behaviors in an academic setting. Thus both the CBCL and the TRF are based on parent, guardian, or teacher observation and assessment. The results from each of these three instruments can be used, either individually or collectively, for assessment and diagnostic purposes by a mental health professional.

Both approaches provide the instrument designer with challenges. For example, one of the advantages of constructing an instrument for completion by external raters is the opportunity to interact with and train the observers. This can produce very high levels of consistency between raters. However, this is also time consuming and costly. Conversely, self-report instruments are subject to each respondent's personal interpretation of an item, which may or may not be what you, as instrument designer, had intended. In either case it is important to test and revise the instrument to minimize these potential problems.

We can also try to classify instruments by *use* or *purpose*. Although this approach provides a nomenclature for describing instruments, as we shall see, it is not easy to make distinctions based solely on intended use.

The array of instruments used in the social sciences, ranges from academic tests to survey questionnaires. It is perhaps better to think of them as a continuum (as in Figure 1.1), rather than as distinct categories, as these instruments may be put to use for more than one purpose and often share common elements. For example, results on an achievement test can be used as a measure of cognition in the same manner as an intelligence test, although the latter would more likely be categorized as a psychometric instrument than a test. Additionally, some instruments, such as tax forms, may not fit easily into a single grouping. The overlapping circles in Figure 1.1 illustrate that these are not distinct or exclusive categories.

One major group of instruments listed in Figure 1.1 is tests. "A *test* is a collection of items developed to measure some human educational or psychological attribute" (Worthen & Sanders, 1987, p. 302). One aspect of a test is that a *correct* answer or level of performance is anticipated. Although tests and other instruments share commonalities, the theories and properties underlying test development are sufficiently different from those underlying instruments as to exclude tests from further discussion in this text.

Rating scale is a generic term describing instruments that are evaluative and that make use of an item format where response choices are ordered on a continuum. The system of judging athletic performances in such events as ice skating and diving is an example of a rating scale. A rating scale differs from a ranking in that the instrument designer predetermines the scale and the respondent selects *one value,* such as *strongly agree* on a scale that ranges from *strongly disagree* to *strongly agree.* During the process of ranking, the respondent creates a hierarchy, placing *all* the values in order, for example from *strong* to *weak,* or from *most important* to *least important.*

FIGURE 1.1: CATEGORIES OF SOCIAL SCIENCE INSTRUMENTS.
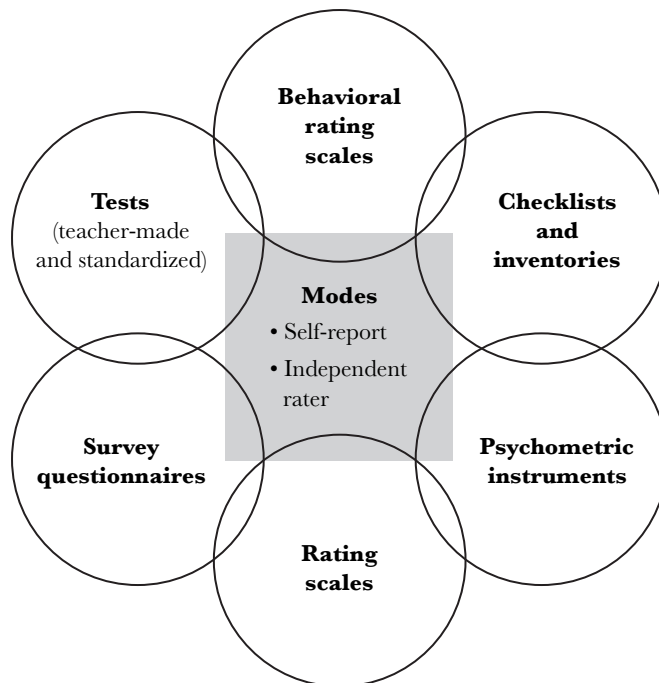
Aiken (1975, p. 12) notes that "rating scales are a primary tool in contemporary assessment methodology, second only to teacher-made achievement tests in frequency of use for rating people, objects, and events."

Rating scales are used to measure attitudes and opinions and also to record direct observation and assessment. Rating instruments are often used to assess the performance of individuals, organizations, programs, and services. For example, professional review organizations often use rating scales to determine if schools or hospitals meet criteria for accreditation or licensure. The following example is adapted from hospital accreditation standards.

EXAMPLE

Treatment plans provide evidence of input from patients and their families (check one):

——Always   ——With a few minor   ——Not consistently   ——Rarely   ——Never
                           exceptions

Performance and behavior rating instruments use rating scales specifically designed to measure an individual's ability to complete a task or perform an activity. Examples of performance instruments are employee job evaluations and the assessment instruments used in rehabilitation and education. These instruments are typically completed by an external observer or rater. Behavior measures may be designed to be descriptive and not evaluative, that is, not judging the value of what is observed but establishing whether it occurred. Such an instrument might use items with a response choice of *present* or *not present.* Behavior measures may also be primarily evaluative, as in a job performance appraisal where the task is to make a qualitative assessment of performance. In that case the response choices might rate an aspect of job performance from *satisfactory* to *unsatisfactory,* or from occurring *none of the time* to *all of the time.*

Checklists are used to determine the presence or absence of an attribute and to count the prevalence of an item or event.[2] These instruments may use a variety of item formats, including scales, rank order, dichotomous choices (*yes* or *no, present* or *not present*), and open-ended questions. One example of a checklist is an instrument used to count the number of computers and computer accessories in a school building. The checklist might be used to indicate if the computer equipment is located where it was originally assigned and to record the property number. The checklist might also include criteria for making qualitative assessments; for example, it might contain a rating scale for evaluating aspects of equipment quality, such as working condition and need for hardware upgrades.

Another example of a checklist is a list of tasks that should be completed to ensure that a process or product is complete. The instructions for building a model airplane might be written as a checklist for instance. Checklists may be organized sequentially for such things as projects where one action must be performed before another can occur—part F of the model airplane may not fit into part G unless it has first been glued to part E.

Checklists may be designed to be self-report instruments or to be completed by an independent observer or rater. Moreover, they may be completed for just one entity (such as one individual) or for multiple entities, possibly chosen by sampling methods. For example, if you are developing a checklist for auditing records (such as medical charts or personnel files) and there are hundreds or thousands of them, you could use random sampling to obtain a representative sample of the records, rather than auditing all of them.

In the broadest terms, an inventory is simply a list of objects, goods, or attributes. The word is also used as a verb to describe the process of compiling the list; one might *inventory* one's supplies, for example. In the social sciences the term *inventory* describes an instrument used to assess a person's interests, characteristics, or skills. One example is the Vineland Adaptive Behavior Scales (Sparrow, Balla, &

Cicchetti, 1984), a developmental inventory that measures adaptive living skills, such as ability to dress oneself and manage a personal budget. An important aspect of developmental inventories is that the items are usually listed sequentially where acquisition of basic skills is a precursor to demonstrating more complex skills; a child typically acquires the ability to hold a spoon, knife, or fork before the ability to cut up food without assistance.

As with other instruments, the mode of administration can be either self-report or observer based. A patient recovering from a stroke might be asked to complete (self-report) an inventory of skills developed as a result of participating in physical therapy. An external rater completing a developmental inventory might observe the same person completing activities of daily living (toileting, personal hygiene, and so forth) and check off whether the individual can or cannot complete the task. An inventory may set a threshold at which observations stop, such as when the individual cannot complete five items in a row. The information obtained can be weighed against normative data to provide a comparative level of functioning. For example, an adaptive living inventory might indicate that, as the result of a stroke, an elderly individual's physical functioning (gross motor skills to carry out the activities of daily living) is significantly diminished in comparison to the functioning of similarly aged adults.

Survey, poll, attitude scale, and *questionnaire* are terms used interchangeably to describe instruments designed to obtain factual information and to assess beliefs, opinions, and attitudes. Questionnaires typically make use of rating scales and open-ended questions. Specific sampling methodologies may be used to obtain responses that are representative of the population of interest (see the following discussion). Questionnaires are typically designed as self-report instruments.

Considerations in the Use of Survey Questionnaires

Surveys are frequently used when information is needed from large numbers of individuals. However, because there are a number of ways of obtaining information, the pros and cons of using a survey as your method of data collection should be considered in each case.

Surveys can be used to explore relationships. You may believe that certain respondents are predisposed to certain decisions—for example, that women, more so than men, rely on appearance and scent when purchasing a particular product. You can obtain data to support this supposition by comparing the demographic data supplied by respondents to their responses on specific items.

Surveys can be used to examine attitudes and beliefs. When assessing attitudes and beliefs there are no right or wrong answers; instead you are interested in the nature of respondents' values, perceptions, and feelings. Surveys are adept

in garnering this information because a number of item formats, including multi-item scales, can be used for both data collection and verification.

Surveys can be used to obtain sensitive information. Respondents are often not comfortable with sharing information about their financial status, legal involvement, or lifestyles in person—through an interview or focus group, for example. Similarly, workers may be reluctant to share information about their employer (such as working conditions) for fear of retaliation if they can be associated with their responses. Surveys that are properly designed and administered and that ensure confidentiality or anonymity can address these obstacles.

Surveys can be combined with other data-gathering approaches. Your survey may provide meaningful results but also suggest that you need to gather additional information not readily assessed by preselected items. For example, you may wish to follow-up your survey that has provided information that can be generalized to the population of interest with a number of interviews that can provide details and explanations the survey was unable to capture.

Despite the many benefits that surveys afford, they also have limitations:

Surveys are mistakenly perceived to be more time and cost effective than other approaches to information gathering. When interventions are evaluated using randomized experimental designs, it may take months or even years to produce results. Qualitative approaches present a similar problem; it may take many months to conduct interviews or videotape situations. In some cases, such as anthropological research, a study may be carried out over a number of years. In comparison to these approaches, surveys are often assumed to be fairly efficient to produce, administer, and analyze. In fact, development of an effective survey may take months by the time one ensures that the instrument produces reliable and trustworthy information. The U.S. General Accounting Office (1986) estimates that it takes about nine months to develop a questionnaire, administer the survey, analyze and interpret the data, and report the results. In some complex situations, the GAO notes, the survey process can take up to eighteen months to complete. For example, if the survey is to be administered by telephone or in person, additional time needs to be allotted for training of the surveyors to guarantee consistent application of questionnaire items. Consequently, surveys should not be chosen solely for perceived efficiency.

By design, surveys limit data acquisition. Qualitative researchers are not limited to set questions or item responses, and qualitative methods are said to produce "rich and thick" information. For example, respondents may provide information the interviewer did not request. "In comparison, surveys are limited in their ability to capture unsolicited but meaningful information, because the items limit the response choices. Although there are ways to extend the amount and type of information you can acquire (for example, by using open-ended items and allotting space for comments), establishing the questions in advance does limit the information you will obtain. Careful and thoughtful design of the questionnaire can

increase the likelihood of capturing the information you are interested in, although it can never completely overcome the limitation posed by defining your questions in advance.

Surveys are subject to misinterpretation. One goal of instrument development is to reduce ambiguity. Ambiguity occurs when respondents misunderstand the meaning of an item and respond to it from their interpretation rather than its intended meaning. Additionally, if the questionnaire is administered as a telephone survey, the surveyor's inflection and tone in reading items may influence the respondent's choice. Misinterpretation can be reduced by pretesting the instrument. Additionally, training can increase consistent presentation of survey questions by telephone surveyors and reduce the impact of tonal quality on the respondent's understanding of the question.

In summary, survey research methods provide a unique process for gathering data. Considering the benefits and limitations of surveys can help you decide whether a survey will provide you with the information needed to support your theory or answer your questions. Careful and thoughtful design can also help you overcome some of the inherent limitations associated with surveys.

Psychometric instrument is a broad term used to describe an array of instruments designed to assess cognitive, affective, and physical functioning and personality traits. Consequently, some psychometric instruments can be just as easily categorized as behavior rating instruments or inventories, such as the Vineland Adaptive Behavior Scales. Examples include instruments designed to assess depression or psychosis, intelligence, self-esteem, and hyperactivity. Psychometric instruments developed to assess vocational abilities and aptitudes are used to predict an individual's suitability for an occupation or a specific job. Some psychometric instruments are very specialized and make use of pictorial response sets. For example, the Children's Apperception Test (Bellak & Bellak, 1993) and the Rorschach inkblot test (Exner & Weiner, 1994) use drawings and abstract designs to solicit responses. Psychometric instruments can be completed through self-report or by an independent observer.

A unique subcategory of psychometric instruments consists of those designed for behavior analysis. Behaviors are measured before, during, and after treatment to determine if an intervention is producing the desired effect as evidenced by changes in the frequency or duration, or both, of a targeted behavior. The number of times the behavior occurs may be counted, or a behaviors may be classified into discrete categories such as *occurred* or *did not occur*, or *correct* or *incorrect* (Kazdin, 1982). Frequency counts are typically tallied for predetermined blocks of time (such as fifteen-minute intervals) or by providing a list of frequency ranges for each observation (such as "Number of Occurrences: 0 to 1, 2 to 4, 5 to 7, 8 or more"), or the observer may just count the occurrences.

Considerations in the Use of Observation Instruments

Many instruments are designed to support the observation and recording of human activity. One example is the ubiquitous job appraisal, used to support personnel decisions such as granting pay raises and promoting, demoting, or retaining employees. Observation instruments are also often used to provide feedback to individuals; such as checklists to observe a teacher's instructional and classroom management skills. In this situation the information can be used by a master teacher to assist with mentoring and coaching. In many human service fields, including mental health, rehabilitation, and education, checklists and inventories employing observational rating are used to identify deficits that may then be the target of treatment efforts.

Observational instruments present many of the same challenges to the developer as do questionnaires. For example, like questionnaire respondents, observers may misunderstand the meaning of an item and rate it based on their personal interpretation rather than the meaning that you had intended.

Perhaps the greatest challenge in the development of observational instruments is ensuring that different raters provide similar ratings when observing the same event at the same time. The extent to which different raters agree on their ratings is referred to as *interrater reliability,* which we'll examine in more depth in Chapter Four. Here it suffices to say that enhancing rater consistency will require additional time in the development and training stages.

As you read these descriptions you probably realized that although researchers would like to be able to make clear distinctions between these various instruments, one of the difficulties in categorizing instruments by purpose is that the terminology often has multiple, overlapping meanings. For example, *rating scale* may describe an entire instrument or any item within an instrument that is developed along a measurable continuum. So a checklist, inventory, or poll can be composed of rating scale items. And as we shall examine later in this book, the word *scale* can have a definition that is distinct from the term *rating scale* even though these terms are often used synonymously. Similarly, the term *survey* is used to describe a type of instrument as well as the method of administration (as in "conducting a survey"). Finally, an instrument might fit the definition of more than one instrument type; for example, an inventory might be constructed as a checklist; a behavior rating scale might fit the definition of a psychometric instrument. The sample instrument at the end of this chapter was designed to evaluate instruction. On the one hand it functions as a survey by soliciting opinions about the course, instructor, and learning environment. On the other hand it functions as a

performance measure to judge the merit or value of different aspects of training. We use these terms to help us communicate our intentions to others, an in, "I am developing a questionnaire, which will be administered as a telephone survey." However, as instrument designer, your primary task is to develop an instrument (which may be a questionnaire, checklist, inventory, and so on) that measures what you intend to measure.

Components of an Instrument

Typically, there are six parts to an instrument, and all or most will be included regardless of the intended purpose or the process used to collect the data. Although we explain the purpose of each component individually, in reality the components function together to create an integrated document. The ability of a respondent or a rater to complete an instrument is based on the quality of each component as well as the instrument as a whole. You may find it helpful to refer to Instrument 1.A at the end of this chapter (the University of Virginia Continuing Education Workshop Evaluation) as we present each section.

Title. The title helps to convey the purpose of the instrument. It is placed at the top center (usually) of the instrument and is the first thing that someone will see when handling the document. The title should be consistent with the instrument's intended purpose.

Introduction. The purpose of an introduction is to explain why you constructed the instrument, how you will use it, and the type of information you want to obtain. It may also contain information about how the instrument and its contents will be managed—for example, how you will ensure the respondent's confidentiality as a participant. The introduction may be conveyed on a separate page or as a separate section at the beginning of the instrument. A 1998 survey (Instrument 7.A) administered to more than ninety thousand public employees in Virginia, included a cover letter from the state governor saying: "I am keenly interested in your concerns about our State government in general and your personal experience as an employee." The purpose of this survey was also described in the instructions for completing the instrument: "This survey will be used to better understand the views of the State's work force" (Office of the Governor, Commonwealth of Virginia, 1998). Notice how these statements convey that the instrument will be used to obtain data reflecting opinions and beliefs.

Directions or instructions. Directions should be given at the beginning of the instrument and within the instrument where respondents need guidance to complete items—for example, when there is a change in item format. External raters or observers will also need comprehensive directions, to ensure interrater consistency.

The purpose of directions is to guide the respondent through the instrument and to assist in obtaining the type of information sought. They should be brief and clear; the directions on the workshop evaluation are concisely stated in one sentence.

Items. At the heart of the instrument are the items, and as we explain in this book, a multitude of item formats can be used to solicit information. *Selection items* provide the respondent with potential choices (that is, choices given in advance) from which the respondent makes a selection. A rating item consists of a *stem,* which is a phrase, sentence, or question that elicits information, and the *response set,* from which the respondent makes the selection. The response choices may form a graded continuum, or scale, or they may be alternatives. An example of a selection item with a scaled response set is being asked to *rate* your experience with a product as *satisfactory, somewhat satisfactory, somewhat unsatisfactory,* or *unsatisfactory*. An example of an item with alternative choices is being asked to *select* your age from a list of age ranges. Another example of a selection item is being asked to *rank* your choices. In this case you must consider, compare, and order all the response alternatives, rather than selecting just one of them.

Supply items, such as open-ended questions, require respondents to provide the answer themselves. Such responses tend to be more descriptive and provide opportunities for more elaboration. However, they may also measure more than the respondent's knowledge; for example, they may also reflect writing skills and language development. Consider the following question, written first as a selection item and then as a supply item:

EXAMPLES

1. I would recommend this organization as a place of employment (circle one):

Strongly Disagree   Disagree   Agree   Strongly Agree

2. What would you recommend about this organization as a place of employment?

In the first example, the selection item, the responses may be easier to score and categorize. This might be a factor if the question is to appear in an instrument administered to hundreds of people. However, the second example, the supply item, may elicit a better sample of the respondent's range of knowledge or attitudes, producing more information than the first example would provide. However, analysis of hundreds of lengthy narratives would be more labor intensive.

Demographics. The demographic section gathers such information about the respondent as age, gender, occupation, and marital status. It should solicit only information that is vital to the project. An instrument used to obtain information about an object will solicit descriptive information about that object. For example,

the demographic section of a checklist to audit medical records might include space to record the record number, the date the audit was completed, and the name of the auditor. The demographic section can be placed at either the beginning or end; Chapter Twelve examines the advantages and disadvantages of each placement.

Closing section. Depending on the purpose of the instrument, it may include a closing section. A survey, for example, might use this section to thank the individual for responding and also to repeat certain information, such as the return address.

Selecting an Appropriate Instrument

As the creator of an instrument, you have some latitude as to the type of instrument to develop, such as a questionnaire, checklist, or inventory; the mode of administration to use; and ultimately the item format(s) to use, such as open-ended questions, rating scales, or ranking items. The decision about the type of instrument and item format(s) will typically be based on the following considerations.

The purpose of the study. The type of instrument to use is usually determined by identifying the type of information you need. In turn, the type of information you require should emerge from defining the purpose of your study. For example, a marketing questionnaire would be suggested if the purpose of your study is to identify consumer characteristics. We'll examine this process in more detail in Chapter Five.

The research design. Identifying the purpose of the study will also suggest some research designs and information-gathering strategies, such as an experiment using random assignment, survey research, or naturalistic inquiry (Chapter Two). The design may therefore suggest instrumentation. For example, a research design based on naturalistic inquiry and fieldwork would imply the use of an observational recording form.

Object of measurement. Closely related to the purpose of the study is the question of who or what will be the focus of inquiry. If the answer to this question is a what, then you will need to develop an instrument that will be completed by an independent or external rater. If the answer to the question is a who, then the options will include self-report as well as an independent rater.

Data collection methodology. In the design phase of instrument construction, it is important to consider the type of data that might be produced and how those data will be collected. This information can help in determining the item format to use. For example, if data are to be collected by multiple independent observers, that might suggest using items with predetermined response choices.

Resources. The design of the instrument and consequently the choice of instrument can also be influenced by the resources available during all phases of your study. For example, will you be the sole instrument designer, or will it be a group effort? What are the time constraints: how much time do you have to design, test, and administer the instrument? Who will do the data collection and the data analysis? These questions should help as you consider the instrument and item format(s) to use.

Taken together these considerations should help you decide on the type of instrument to develop. For example, in order to respond to a school board request, an elementary school principal needs to know for how many hours during the school day children are engaged in reading. The principal could ask her teachers to guesstimate the amount of time, or she could use a more systematic approach, such as having teachers keep a written record. To make data collection more efficient, the principal decides to develop a checklist, so that the teachers need only circle the times that students are engaged in reading activities. In developing this instrument the principal takes into account a number of factors, such as the type of information needed, how data will be collected, and available resources (that is, she is sensitive to the limited amount of time that teachers have to collect the data).

The Process of Instrument Construction

As we discussed earlier, artists typically go through a number of steps in the creation of a painting. First, they may draft a number of preliminary sketches of their subject, drawing the subject from different angles, varying the lighting, and if painting the human figure, trying different poses. Some artists like to make color sketches in pastel chalks or watercolors before committing to a final product in oil or acrylic paint. And even after they have begun applying paint to canvas, they may make changes during the process of actually completing the painting.

This process of constantly revising a composition is an *iterative* process, and instrument construction, like painting a picture, should be viewed as a systematic yet creative activity that requires continual refinement and revision (Figure 1.2). Although there are clearly steps that must be completed in this process, instrument construction may not always progress sequentially. For example, after you receive feedback from a friend, colleague, or potential user of the instrument, you may find you need to rewrite specific items or reorganize the instrument itself. Consequently, the following activities should be viewed as part of a creative cycle.

Articulate the purpose and focus of the study. This is perhaps the most important activity, and yet it is often overlooked or minimized. Specifying the purpose helps you identify the themes or concepts you want to understand, the methodology you might use, the type of instrument to develop, and the questions or items that

FIGURE 1.2: STEPS IN THE INSTRUMENT CONSTRUCTION PROCESS.

Identify the purpose and focus of the study (questions you want answered).

Obtain feedback from stakeholders to clarify the purpose and focus.

Administer instrument; analyze and report results.

Identify the research methodology and type of instrument to use for data collection and measurement.

Pilot test and revise prior to final administration.

Begin to formulate questions or items.

Revise instrument based on feedback; prepare for pilot testing.

Pretest items and preliminary draft (with content experts, stakeholders, potential respondents or raters).

you might include in the instrument to obtain the information you seek. Part of this process involves reviewing literature (books, journal articles, and so on) relevant to the topic. Discussing your project with others who are familiar with its topic can help you identify aspects to focus on or exclude from the study. This is important if the study has been commissioned, as you may need to negotiate the purpose and content of the study. Additionally, you may want to speak with potential respondents, as they too can provide information and their personal perspective about the topic. In considering the approach to use to collect and analyze the data (that is, the methodology), ask whether the topic lends itself to qualitative approaches such as interviews or focus groups, to examination of archival data, to a survey, or to an experimental design.

Failure to complete this activity could result in obtaining information that does not answer your question or that does not provide useful and accurate

information for decision making. Drafting a purpose statement can also assist in specifying how the data will be collected and who will complete the instrument.

Activity review and check. It is important to share your purpose statement with others and to obtain feedback that can clarify and focus the project.

Formulate items. This is both a technical and creative process. It is based on your knowledge of the topic, the information you have described in your purpose statement, and your review of supportive information to help in the creation of specific items. For example, if you are going to conduct an assessment of employee satisfaction, you will probably want to know, prior to designing the questionnaire, what the organization does and how it is organized. Additionally, you will want to understand some of the management theories and concepts that apply to the understanding of employee satisfaction. This information will assist you in brainstorming and writing out questions or items. This process may also involve examining similar instruments to see how others have carried out this measurement. (You may even decide that rather than create a new instrument you will use or adapt an existing instrument, being sure to cite the original source[s], respect copyright, and pay a fee if the instrument is proprietary.)

This creative process may be a solitary or a group endeavor. At this stage, items may be written as questions or statements, and no attempt is made to format them into, for example, open-ended questions or selected response items. Additionally, during this activity decisions can be made about the specific information to obtain. For example, if you have decided to create a survey that will be administered only to women, there is no need to include a demographic question asking for the respondent's gender.

Activity review and check. Obtain feedback from others on item validity: do the items appear to ask for information that will answer your question or obtain the information you need for decision making? Modify your items based on this feedback.

Structure and format the items. The next activity is to decide on the item format or formats that will provide the information you need efficiently and effectively. This decision will depend in part on the purpose of your study. For example, an instrument designed to support observations of another's behaviors might suggest the use of a behavior rating scale, whereas an instrument with items that ask for opinions will likely suggest the use of a scale measuring the respondent's level of agreement with a statement. And depending on what you want to measure, you might incorporate a variety of item formats in the instrument.

Activity review and check. Obtain feedback from content experts (individuals knowledgeable in the subject you are studying) as well as from potential respondents. This feedback is used to determine whether the items make sense and are unambiguous and whether the information you obtain from the items will provide you with the information you want. This feedback can also assist in identifying problems

with instrument administration. Do individuals who will use an observational instrument agree on the meaning of the items and what they are supposed to observe and report?

Organize and format the instrument. At this stage you are ready to organize items so that they flow in a logical order, and depending on the purpose of the instrument, you may need to consider the content of the instructions and the demographic section.

Activity review and check. Typically, you will want to pretest the instrument as a complete document rather than item by item. This pretest determines how long it takes to complete the instrument and identifies places where respondents or raters had difficulty completing the instrument due to bottlenecks or confusing items or instructions. This is an important activity, as an incomprehensible or awkwardly formatted instrument may result in a low response rate (for a survey questionnaire) or unreliable data (for a checklist or behavior rating instrument).

Administer and revise the instrument. Even though you have taken steps to correct problems during instrument design and construction, the responses from the first administration may suggest improvements, and you may find that the instrument requires further revision if you intend to use it again. Such revision is a common activity with instruments used to measure performance, as repeated administrations will certainly highlight shortcomings or portions of the instrument that need to be corrected. You may find problems in the administration of the instrument that can be improved as well.

Summary

In this chapter we provided a definition of an instrument, described the parts of an instrument, and listed the steps in the instrument construction process. We also described some ways that instruments might be categorized and pointed out that the terminology researchers and evaluators use to describe these categories can be somewhat indefinite. A major theme is that instrument construction is an iterative process— as you will find yourself repeatedly revising and refining the instrument in response to feedback—and therefore it is an activity that is both technical and creative.

Instrument 1.A: Illustrating the Parts of a Questionnaire

Throughout this book we present sample instruments or items to illustrate our themes and main points. These samples can provide guidance as you design your own instrument and when you need to evaluate instruments designed by others. We use our first example to illustrate the components of an instrument.

The Continuing Education Workshop Evaluation Form (Instrument 1.A) was developed a number of years ago to evaluate a continuing education workshop and its instructors. The results are provided to the instructors so they can use this information to revise and improve the curriculum and classroom instruction. The title of the form is printed in bold type at the top of the page. The title—Continuing Education Workshop Evaluation Form—clearly connotes the instrument's purpose. The demographic section consists of three items: course title, course number, and schedule number. The questionnaire is typically completed at the conclusion of the training, and participants are given the demographic information to fill in these blanks at that time. Further down the page is a line for the instructor's name. The form provides for the rater's anonymity, as the individual completing the form does not provide identifying information. The instructions are concisely stated in one sentence and are printed near the top of the page.

Through experience, the department using this form has identified those few questions that produce meaningful information, resulting in a very concise instrument. For example, the instrument makes use of both selection items and open-ended statements. Each item is preceded by a number. Items 1 through 7 are to be rated on a *Likert* response scale, from *strongly agree* to *strongly disagree*. *Not applicable* (NA) and *no opinion* or *neutral* (N) options have been provided in the response set. The items address instructor skills, the physical environment, and course content. Items 8, 9, and 10 provide space for written comments. In order to be concise, this instrument does not include an introductory or a closing section.

The instrument is organized coherently and logically. Large black boxes are printed in the response matrix corresponding to item 8 to indicate that this is an open-ended item; for clarity and continuity, black boxes should also have been printed after items 9 and 10. The instrument is designed for automated data collection, using an optical mark recognition device (Scantron). For each of the first seven items, the respondent darkens an oval corresponding with the appropriate description on the scale printed near the top of the page. This facilitates data collection. Alternatively, respondents might have checked a box or circled a number.

The Workshop Evaluation Form also highlights the difficulty of categorizing instruments and the need to clearly define an instrument's intended purpose. The instrument's primary function appears to be to work as a survey questionnaire, by soliciting opinions about the course, instructor, and learning environment. When it is used for this purpose, an instructor can use the findings to improve the content and process of a workshop. However, items 1 through 5 of this form could also be used by a supervisor to assess student opinions about the instructor's skills and abilities. When used for this purpose, the instrument functions as a performance appraisal.

INSTRUMENT 1.A: WORKSHOP EVALUATION.

Continuing Education Workshop Evaluation Form

	Instructor #						Schedule #			
	⓪	⓪	⓪	⓪		⓪	⓪	⓪	⓪	⓪
	①	①	①	①		①	①	①	①	①
	②	②	②	②		②	②	②	②	②
	③	③	③	③		③	③	③	③	③
	④	④	④	④		④	④	④	④	④
	⑤	⑤	⑤	⑤		⑤	⑤	⑤	⑤	⑤
	⑥	⑥	⑥	⑥		⑥	⑥	⑥	⑥	⑥
	⑦	⑦	⑦	⑦		⑦	⑦	⑦	⑦	⑦
	⑧	⑧	⑧	⑧		⑧	⑧	⑧	⑧	⑧
	⑨	⑨	⑨	⑨		⑨	⑨	⑨	⑨	⑨
	⑩	⑩	⑩	⑩		⑩	⑩	⑩	⑩	⑩

Course Title: _____

Course Number: _____

Schedule Number: _____

Name of Instructor: _____

Please assess each of the following statements based on the key by filling in the bubble in the column that best represents your opinion.

SA - Strongly Agree	**A** - Agree	**N** - No Opinion or Neutral	**D** - Disagree	**SD** - Strongly Disagree	**NA** - Not Applicable

	SA	A	N	D	SD	NA
1. The objectives for the course were clearly stated.	☐	☐	☐	☐	☐	☐
2. The instructor effectively taught the stated objectives.	☐	☐	☐	☐	☐	☐
3. The intructor used a variety of teaching strategies (two or more of the following: lecture, discussion, small group activity, visual/audio aids, individual assistance).	☐	☐	☐	☐	☐	☐
4. The workshop demonstrated how to apply the strategy or process presented.	☐	☐	☐	☐	☐	☐
5. The instructor demonstrated openness and receptivity to student needs and opinions.	☐	☐	☐	☐	☐	☐
6. Facilities and equipment (e.g., audiovisual equipment) were adequate. If not, comment below.	☐	☐	☐	☐	☐	☐
7. The subject matter was relevant to my professional needs.	☐	☐	☐	☐	☐	☐
8. What I liked best about this course was:	■	■	■	■	■	■
_____	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐
9. To strengthen the course I would suggest:	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐
10. Additional workshops, courses, or programs I would like the University Continuing Education to offer:	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐	☐

Source: University of Virginia Department of Continuing Education. Reprinted with permission.

Instrument 1.B: Medical History Questionnaire

Instrument 1.B, a medical history questionnaire, was created by extracting items from several different medical history questionnaires that were readably available from sites on the Internet and is provided as an example of an instrument based solely on items that solicit objective information. The instrument illustrates various item formats as it uses both open (supply) and closed (selection) items such as checklists and dichotomous (*yes* or *no*) response sets.

INSTRUMENT 1.B: SAMPLE MEDICAL HISTORY.

Name: _____ Date of Birth: _____ Current Age: _____

Gender: ☐ Male ☐ Female Health Care Insurance: ☐Yes ☐No

If Yes, name of insurer: _____ Policy No.: _____

Do you have a present or past history of: (*check all that apply*)

☐ ADHD	☐ Diabetes	☐ Pneumonia
☐ Alcohol Abuse	☐ Eating Disorder	☐ Skin Problems
☐ Anemia	☐ Heart Disease	☐ Single Cell Anemia
☐ Arthritis	☐ Hepatitis	☐ Sleep Problems
☐ Asthma	☐ Hernia	☐ Smoking
☐ Back Problems	☐ High Blood Pressure	☐ Thyroid Disorder
☐ Cancer	☐ High Cholesterol	☐ Tuberculosis
☐ Chronic Cough	☐ HIV/AIDS	☐ Urinary Tract Infection
☐ Dental Problems	☐ Measles	☐ Whooping Cough
☐ Depression/anxiety	☐ Mumps	☐ Other (please specify):

Are you currently on any medications: ☐ Yes ☐ No
If Yes, please list the medications here:

Do you have any allergies: ☐ Yes ☐ No
If Yes, please list your allergies here:

Exercise History:

In what kind of exercises or sports do you currently participate?
How often do you participate?
How many years have you participated in these activities?

To be completed by nurse:

Height: _____ Weight: _____ Blood Pressure: _____ Pulse: _____
Vision: Without Glasses: Right ____ / Left ____ With Glasses: Right ____ / Left _____

Thank you for completing this form.

This medical history questionnaire also requires two modes of administration. Although it is primarily designed to be a self-report instrument, with the client providing the information, data to complete the last section are obtained by a nurse. Note that there are no instructions at the beginning of the instrument; instead, directions are provided as necessary in the body of the questionnaire. Additionally, a purpose statement is probably not needed as the instrument is designed to be used in a specific setting, such as a doctor's office, or during the hospital admissions process.

Instrument 1.C: Example of a Checklist

The Research Evaluation Checklist (Lutz, 2006) was developed to assist stakeholders evaluate research projects and reports. For example, this checklist might be used by an agency institutional review board (IRB) to assess a research proposal. Consequently, it is also a useful instrument for someone (perhaps you) who is planning a research project or is documenting the results of a project.

This is a simple *yes* or *no* checklist, with the reviewer literally checking a box when the document being examined contains the required information or leaving the box blank when the information is not present. Items are clustered by themes and concepts. This example illustrates the kind of checklist described in this chapter. Although the checklist itself offers no instructions, it appears in a manual that describes how it might be used, as follows:

> The second checklist is longer and will be useful when completing a thorough evaluation of a research report. The questions are organized according to the components of a research report. (Part III [of the manual], Getting the Most Out of a Research Report, provides a definition of a research report and describes its typical contents.) People who have little or no experience in evaluating research may wish to select questions from the second checklist rather than using all of them. Some questions include terms that may be unfamiliar to you. Be sure to check for their definitions in the Glossary in Part IV, Understanding the Language of Research. In Part II and Part III, certain terms will appear in underlined *italics* the first time they are used. Their definitions can be found in the Glossary [Lutz, 2006, p. 11].

INSTRUMENT 1.C: RESEARCH EVALUATION CHECKLIST.

Significance of the Study
❑ What makes the study useful, important or of interest to you?
❑ What do you want to know from the *research*?
❑ Did the researchers clearly explain why they did the study?
❑ Have the researchers convinced you of the importance of the study?
❑ Does the study address a *gap in knowledge* or provide new information about the topic or issue?
❑ Were *constructs, variables* and terms clearly defined?
❑ Was a clear *rationale* presented for the constructs or variables examined in the study?
❑ Were consumers or family members involved in helping to *design* the study?
❑ Were board members, administrators or agency staff involved in helping to design the study?

Research Questions and Hypotheses
❑ Were the research questions and hypotheses clearly stated?
❑ Did the research questions and hypotheses accurately forecast what would take place in the study?

Methodology
❑ Were the procedures clearly described?
❑ Could someone repeat the study after reading the *methods* section?
❑ Were appropriate criteria used to select the *sample*?
❑ Were methods used to prevent *bias* in the study?
❑ Were the procedures and instruments *reliable* and *valid*?
❑ Were the procedures and instruments free of potential bias (for example, age, gender, racial or ethnic)?
❑ Was a *control group* needed to address the main question of the study?
❑ Was an appropriate control or *comparison group* used in the study?
❑ What steps did the researchers take to prevent harm or distress to the research participants?
❑ Was the study approved by an institutional review board (IRB) or a similar committee on ethics?

Sample, Representativeness, and Generalizability
❑ Did the researcher justify the size of the sample?
❑ Did the researcher describe the methods used to determine sample size?
❑ Was the sample size sufficient to find significant results?
❑ Do the characteristics of the sample match those of the *population* of interest?
❑ Did the *study setting* match the location where the results will be applied?

Statistical Methods and Results

❑ Did the researchers use appropriate statistical tests to evaluate hypotheses or answer the research questions?

❑ Did the researchers use the appropriate statistical tests for *small sample sizes*?

❑ Did the researchers clearly identify the *major findings* of the study?

Discussion, Limitations, and Implications

❑ Did the researchers identify the limitations and biases of the study?

❑ Did the researchers discuss how the limitations and biases influence the results?

❑ Did the researchers discuss recommendations from or practical implications of the findings?

❑ Did the researchers discuss what future studies could be done on this topic or issue?

❑ Did the researchers state how they will use the findings?

❑ Did the researchers recommend how others can use the findings? For example, did they indicate that the findings might be used to:

    ❑ improve a current practice or service?

    ❑ develop or revise a policy?

    ❑ change a mental health law?

    ❑ support a request for funding?

    ❑ support implementation of a new program or service?

Source: Adapted from Heacock, Koehoorn, & Tan, 1997; Kazdin, 1998; Krathwolh, 1988; Centers for Disease Control and Prevention, 2006; National Alliance on Mental Illness, 2006; Lutz, 2006. Reprinted with permission.

Endnotes

1. Despite our best efforts, we humans "filter" what we observe, based on prior experience. Consequently, two people may see the same event, but interpret it differently, and if asked to measure the event using some form of observational instrument, they may come up with decidedly different ratings; consider, for example, the scores different judges give to an ice-skating performance.

2. Scriven (2000, p. 1) defines checklist as, "a list of factors, properties, aspects, components, criteria, tasks, or dimensions, the presence or amount of which is to be separately considered, in order to perform a task."

Key Concepts and Terms

attitude scale	objective	selection item
behavior rating	observation	self-report
checklist	poll	stem
instrument	performance rating	subjective
inventory	psychometric instrument	supply item
item	questionnaire	survey
Likert response scale	rating scale	test
mode of administration	response set	