

CHAPTER 2

DATA COLLECTION AND PREPARATION

Instruments to Collect Data

Research usually starts with the collection of information related to the field of study by reviewing the existing literature (see Vol. 1, Chapter 4). The way the rest of the data are collected depends on the research design (see Vol. 1, Chapter 5). For instance, in survey research, the researcher generally relies on the results of questionnaires, interviews, or observations aside from reliable written documents. In experimental research, however, the researcher relies mainly on the results of the given treatment to the experimental group in order to investigate the cause and effect relationship.

The research design requires a specific method of data collection that would enable the researcher to arrive at an answer to the related research question (see Vol. 1, Chapter 3), which recapitulates the core of the research. Thus, depending on the research question and the hypothesis formulated, the researcher decides on a suitable type of analysis that would lead him or her to find an answer to the formulated research question.

The main focus of this chapter will be on the **instruments** used in descriptive research (see also Vol. 1, Chapter 5, Part I). As mentioned above, aside from written sources, questionnaires, interviews, observations, experiments and tests can be cited as the main types of instruments utilized in data collection. (To obtain detailed information on collecting data for experimental research refer to Vol. 1, Chapter 5, Part II.)

Before discussing data collection instruments in detail, the steps a researcher generally follows in conducting a study will be outlined with illustrated examples extracted from a single study:

- The researcher, with the research question/purpose statement, and hypothesis in mind, states his/her **dependent and independent variables** explicitly (see Vol. 1, Chapter 3).
- The stated variables will enable the researcher to determine the **scope** of the study and consequently, limit the **setting and the categorized issues to be inquired or observed**.
- The specific setting and the categorized issues to be inquired or observed help the researcher become aware of the possible **extraneous variables** that would cause a threat to the internal validity of the research (see Vol. 1, Chapter 3).
- The researcher then decides on the source of data collection (written documents, questionnaire, interview, observation). If there is a need for eliciting the public opinion or the behavior, the researcher decides which **instrument** to use (questionnaire, interview, observation).
- The researcher makes sure that his/her choice and the content of the instrument would fit into the specific framework of the research. The researcher then decides on the **method of analysis and the statistical techniques** to be employed.

Written Documents

- The data collected from documents enable the researcher to examine
 - what has been done in the field up to present; what needs to be done for further contribution;

- what type of research design has to be implemented for this purpose;
- what type of data is needed to arrive at the ultimate goal; and
- what procedure needs to be applied to analyze the data.

After the research design is planned out, the researcher uses the gathered information noted down on the content cards in order to write the "Review of Literature" section of the research (see Vol. 1, Chapter 4). As Glidden (1964) indicates, data gathered from other sources, especially from quantitative data, need to be analyzed and evaluated very carefully (see Vol. 1 Chapters 3 to 5) because

good technical writing is always the product of an orderly procedure involving analysis and synthesis. First, the subject matter is taken apart to allow detailed sorting of the components for purposes of classification. Second, the components are evaluated into language, and synthesized into a piece of writing by following an organized outline. Analysis, classification, and organization are distinct steps but are so closely interrelated that they are frequently carried out to some degree simultaneously. (p. 125)

Written documents are also good sources for content analysis. In descriptive and historical studies, researchers are involved in the analysis of the written and visual contents of documents. Textbooks, essays, newspapers, novels, magazine articles, political speeches, letters, diaries, advertisements, and cartoons are among the common types of documents used for this purpose. This is because individual's or people's beliefs, ideas, traditions, attitudes, moral values and many other physical, emotional, religious, social or personal characteristics of individuals can be derived from these documents.

Questionnaires

Aims of a Questionnaire

Questionnaires are widely used in survey research with the aim of

- **eliciting information** about the situation and behavior of the respondents by posing descriptive questions,
- **investigating respondent's experiences** on a specific topic by asking exploratory questions,
- **interpreting and explaining the existing or known situation** by means of inferential or explanatory questions.

The items in a questionnaire also helps the researcher to

- identify a problem related to the respondents,
- find out the priorities of the respondents in relation to a specific topic,
- investigate the abilities of the respondents in relation to a skill,
- find out a solution to an existing problem,

- find out bibliographic information about the respondents for making generalization or arriving at some criteria for future purposes,
- help the respondents to reflect on their own action, and consequently improve themselves by means of self-evaluation,
- find out the respondents' judgments and opinion on a specific topic.

Types of Questionnaires

Questionnaires could be in open-ended or closed formats. Open-ended questions elicit unstructured responses while closed questions elicit structured ones. Both these question types are used in surveys depending on the aim of the research and the number of sample to be questioned.

Open-ended questions. Open-ended questions require a respondent to write an answer either in the form of a statement, a phrase, or a word. In such a case, the respondent uses his/her own words in providing the information. Since the response is given in a free form, it is usually open to interpretation.

Open-ended questions are generally formulated in the form of a question with space to write the answer in. Another form of open-ended question is a statement including a blank that is filled in by the respondent. If the response possibilities are clearly limited and specific, according to Trochim (1997), this format is more efficient because the responses can be entered into computer data sets. He claims that written out answers concerning age, state of residence, income, or credit-hours earned would facilitate the inferential statistical procedures. These numerical mode of responses can later be categorized into groups when needed.

Close questions. Close questions are easy to use and score because all subjects make their choices from the given alternatives, the researcher receives standardized data, which can easily be prepared for computer analysis. These questions are, however, difficult to develop for it is not easy to include all the answers that would possibly be given by the respondents. For that reason, in most of the items in a questionnaire, an extra option is provided as "other" so that the respondent can write it out if his/her answer does not exist in the provided options. Some scholars such as Arreola (1997) oppose to the idea of giving the "other" option especially if a long list of choices has already been provided. If the word "other" is accompanied with expressions such as "specify and explain," the respondents may write out coherent explanations for choosing the "other" option. There are, however, few cases, where the use of "other" seems to be appropriate as in the case of inquiring about the race of an individual.

Close questions can be formulated in several different forms depending on the way the question is posed (e.g. contingency questions, matrix questions) or on the way the responses are provided in the options (e.g. Likert scale, rating scale) :

Contingency questions. In order not to frustrate respondents with questions that are of no relevance to them, researchers prefer to formulate two-part questions called contingency questions. The answer to the first part of the question helps the respondent choose the next consequent question (Trochim, 1997). This way, the respondent either skips the second part of the question and goes to the following one or continues with the second part of the same question.

Matrix questions. Sometimes, researchers combine several questions of the same set into one category for a specific aim in mind. In such a case, a matrix of items and answers is constructed under one heading comprising the instruction. It is advantageous to use a matrix format for two main reasons. First of all, it is easier for the respondents to answer a set of questions of the same nature presented in the same table-like format. Secondly, this type of format makes it easier for the researcher to compare and contrast the given responses.

Scaled responses. The design for scaled response format varies depending on the options to be considered by the researcher. The Likert scale, which is one of the most common scale-response formats, has been developed by Rensis Likert in an attempt to improve levels of measurement in social research through the use of standardized response categories in survey questionnaires (Trochim, 1997). It is a five-point scale comprising of the following options:

Strongly agree, Agree, Neutral, Disagree, Strongly disagree

Rating scale is another type of scale requiring respondents to rate some characteristic or quality on a specific scale. For instance, the quality of a product may be rated on a scale ranging from 1 (worst) to 5 (best).

Designing a Questionnaire

If there is not a ready made standard questionnaire to suit the purpose of the study, then the researcher develops his/her own. First of all, the type of questionnaire to be utilized should be considered. In general, in order to anticipate certain answers from a respondent, it is desirable to include a closed type of questionnaire consisting of categorized or pre-coded types of questions with options as answers. The purpose of using this type of a questionnaire is to reduce ambiguity and misinterpretations. On the other hand, however, an open-ended questionnaire might be helpful for the respondents to produce spontaneous, unanticipated answers that might shed light on the research.

Depending on the size of the sample, a decision can be made as to what type to use. It is most probable that a researcher conducting a large scale survey utilizing an open-ended questionnaire would encounter physical as well as financial difficulties. According to Fraenkel and Wallen (1990), for a survey to be conducted on 50 respondents or less, an open-ended question format might be designed if unexpected information is desired to be captured.

The format and the content of the questionnaire may encourage the respondent to fill out the questionnaire quickly and in a more reliable manner. Therefore, the following issues are considered important in preparing a questionnaire (Isaac & Michael, 1977; Selinger & Shohamy, 1989; Fraenkel & Wallen, 1990):

Format wise

- The format of the questionnaire should be appealing to the respondent. The spacing, sequencing, and printing have great impact on the respondent and may change his decision to respond positively or negatively depending on how well or how poorly it has been prepared.
- Questionnaires should be as short as possible. If they are too long, respondents could be reluctant to fill it out.
- No space should be provided for respondents to write their names. If they are not asked to write their names, they will be more willing to participate.

Content wise

- Include questions that are relevant to the research problem. In other words, questions that do not directly relate to the research problem should be excluded from the questionnaire.
- Exclude questions that have no relevance to the respondent.
- Avoid ambiguity and vagueness. In other words, the terminology should be familiar to the people who are responding the questionnaire. Try to use words with precise meanings whenever possible.
- Avoid complex and awkward word arrangements.
- Avoid asking questions including terms that are beyond respondents' limits. This would confuse them. Try to express the items as clearly as possible.
- Avoid "offensive and emotive" questions.
- Avoid prestige bias. If the question or the response contains a word or a statement that is associated with a prestigious person or group, the respondents may be biased in their answers.
- Avoid double purpose questions. Each question should focus only on one issue. If two or more inquiries are included in the same question, respondents would have a hard time in answering, especially if they have opposite opinions about each of these inquiries in that item.
- Try to maintain the validity of the answers, by formulating questions that would make the same inquiry about the same topic but from different perspectives.
- Avoid leading questions. The question should not be formulated in a way that would impose a particular answer.

Questionnaire Scales

While developing a questionnaire, the response categories included within the scaling system need to be chosen with great caution. This is because carelessly inserted alternatives in the scale would either confuse the respondents and lead them to mark the respond category that does not really indicate their intention, or make them reluctant to answer the related question in return. Consequently, the obtained results do not reflect the reality.

Moore (1983) describes four scales :

- Ranking scale
 - Numerical rating scale (see Example 1)
 - Graphic rating scale (see Example 2)
 - Likert scale (see Example 3)
- Semantic differential (see Example 4)
- Checklist (see Example 5)
- Nominations scale (see Example 6)

Example 1:

Directions: Circle the number that best indicates the degree to which the researcher considered the various threats to internal and external validity. The numbers represent the following values:

5-outstanding; 4-good or above average; 3-average; 2-poor or below average; and 1 not satisfactory.

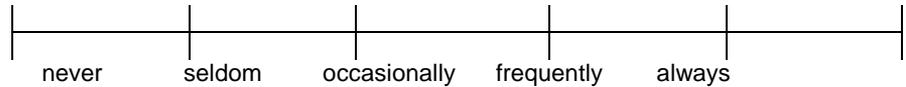
1. To what extent does the research article consider threats to its internal validity?

1 2 3 4 5

Example 2:

Directions: Place an X along the horizontal line where it indicates the degree to which the researcher considered the various threats to internal and external validity.

1. To what extent did the research article we evaluated consider threats to internal validity affecting its design?

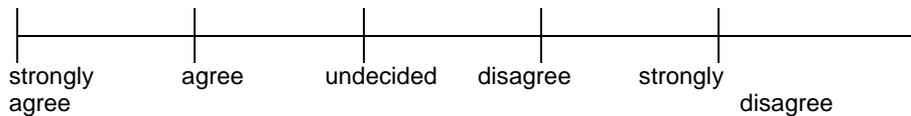


Example 3:

Directions : Please circle the letter below to indicate your perception about each statement. If you *strongly agree* circle SA; if you *agree* with the statement, circle A; if you *neither agree nor disagree*, circle U for *undecided*; if you *disagree*; circle D; and if you *strongly disagree*, circle SD.

1. The instructor was organized for class. SA A U D SD

or



(Moore, 1983, p. 203)

Example 4:

Directions: Place an X in one of the 7 spaces between each adjective pair that best indicates your perception of the teacher's behavior. The closer you place your X toward one adjective or the other, the more you think that adjective better describes the teacher than the other.

(1) ORIGINAL	_____	CONVENTIONAL
(2) PATIENT	_____	IMPATIENT
(3) COLD	_____	WARM

(Tuckerman, 1975 in Moore, 1983, p. 204)

Example 5:

Directions: Check all those which apply.

1. This research study was _____ informative
 _____ well-constructed
 _____ worthwhile
 _____ incomprehensible

Example 6:

1. Name the three students who are best in math.

(A) _____, (B) _____, (C) _____

(Moore, 1983, p. 205)

Considerations in Developing Questionnaire Scales

In order not to mislead the respondents and cause any misinterpretations of the obtained results, the researcher tries to consider the following points (Swenson, 1995; Arreola, 1997; Trochim, 1997). (After you read the consideration, try to do Exercise J to be able to see the related examples.)

Category proliferation. It is suggested to keep the categories (e.g. in relation to age and marital status) as small as possible unless they play a crucial role. This type of an approach makes the respondents feel comfortable in answering the questionnaire since it does not pry into their personal matters.

Scale point proliferation. Providing too many response categories within the scale could cause difficulty for the respondents in making their choices. According to the results of psychometric research, respondents cannot reliably distinguish more than six or seven levels of response. Offering four to five scale points is usually quite sufficient to stimulate the respondents in reflecting the direction of their response (Swenson, 1995).

Order of categories. If response categories are representing a progression from the lower level to the higher, it is suggested that they should be listed with the lower first and then move to the right as the levels get higher (e.g. 1. Never 2. Seldom 3. Occasionally 4. Frequently).

Responses at the scale midpoint. Associating scale values of 1 through 5 to response categories expressing the degree of performance or agreement/disagreement may yield misleading results because people usually express the degree of performance as average, below average and above average. When the slight variances of degrees are included into the scale, the interpretation of the results may not reflect the reality. It is the same case in expressing one's agreement or disagreement. Instead of including five different response categories for agreement, if only two categories (1. Agree 2. Disagree) are given, more reliable results are obtained because it would be easy to apply the statistical procedures, and interpret the outcome of the analysis.

Ranking questions. Instead of asking the respondents to rank the items on their own, they can be asked to make their choices among three or four categories.

Sensitive questions. Respondents do not like to answer questions eliciting information about their personal life especially when they appear at the beginning of the questionnaire. After the questionnaire is developed, the reliability and the validity should be tested before administering it to the subjects. This is also important in administering other self-

developed instruments. Therefore, detailed information will be given on reliability and validity at the end of this chapter.

Administering a Questionnaire

Questionnaires are generally printed out and mailed to the respondents for the purpose of receiving the opinion of the public on a specific issue. Nowadays, there are e-mail/Web based questionnaires, which are sent around through the use of the Internet.

It is not possible to achieve a high rate of response with questionnaires. In fact, in these self-administered questionnaires, the response depends on the interest, attitude, and background of the people who are filling out the questionnaire. The stamped and self-addressed envelope usually encourages people to respond better. The response rate is very high if the questionnaire is given to students in class as part of their course requirements. These group-administered questionnaires seem to be more advantageous because, the researcher

- gets the answers immediately,
- has an opportunity to clarify the points that may be confusing to the respondents,
- observe the situation under which the respondents fill in the questionnaire.

In such a circumstance, if the questions are about the subject matter taught, they may not reflect the truth in their response with the fear of getting a lower grade for having hurt the feelings of teachers or administrators.

Applying a coding system on the response sheet to identify the respondents is considered a violation of ethics. Therefore, researchers should not adopt a strategy of this kind even if it is for the sake of checking on the returned sheets.

In administering a questionnaire on a large population, the researcher is advised to send twice as many of the questionnaire sheets to a random sample taking into consideration the possibility of the nonreturnable questionnaires (Trochim, 1997). In other words, if the sample is planned to comprise of 250 respondents, the questionnaire should be sent to 500 randomly selected respondents.

Interviews

An interview is conducted to elicit personal opinions of the subjects about the issue in concern. Interviews can be formal (structured) or informal (unstructured). A formal interview is governed by a set of questions prepared beforehand and all the subjects that are interviewed are asked the same questions. Interviews, if conducted properly, would give more reliable results. The interviewer plays a crucial role in conducting the interview since there are important tasks that need to be fulfilled. Either before or during the interview the interviewer should

- make a list of the respondents to be interviewed.
- find their addresses.
- go out for interviews at times available for the respondents although they might be inconvenient for him/her.
- motivate respondents in order to obtain sincere answers to the posed questions.
- try to be very alert in order to clarify any confusion or concern coming from the respondent.
- note the behavior of the respondent during the interview and thus makes judgments accordingly about the quality of the information being received.

- try not to reflect all his/her personal worries and concerns in order not to distract or disorient the respondents (adopted from Trochim, 1997).

Training Interviewers

In the case of a small scale study, the researcher can manage to conduct all the interviews by himself/herself. In large scale studies, however, there is a need of more interviewers. This requires the training of the interviewers. Trochim (1997) suggests the following in training interviewers:

- Describe the entire study to the interviewers so that they would have some background knowledge about the basics of the research.
- Indicated to the interviewer whether there is a company or an agency responsible for conducting or sponsoring the research.
- Try to help the interviewers to understand the rationale of how the instrument was conducted by explaining to them the basics of survey research.
- Explain your reasons for selecting the sample very carefully.
- Make the interviewers aware that they can inadvertently bias the results especially dealing with political or moral issues. Therefore, tell them to be careful in that respect.
- Go through the interview with them so that they see how each part of the interview is related to each other.
- Explain to the interviewer the selection procedure. If it was not possible to identify the addresses, provide them with city maps and orient them in reading maps. If you want to interview people living in a specially designed house, make sure that the interviewer can make the distinction you have in mind. If you want the respondent to be at a certain age or a sex, make sure that the interviewer chooses that member in the family as a respondent.
- Rehearse the interview so that any problems that might come up are detected in advance. Video taping and discussing the difficult situations from the video would yield better results.
- Discuss the scheduling of the interview so that they would be prepared to conduct interviews at times that might not seem convenient for them.

The interviewer needs to be equipped with certain documents to look professional and also to ensure the respondent that he/she is authorized to conduct the interview:

- A cover letter from the Institute sponsoring or supporting the interview
- An identification to prove that the person whose name is mentioned in the letter is himself/herself
- A folder holding the papers with the questions to be posed to the respondent

Conducting an Interview

The appearance of the interviewers is very important as well. Therefore, they must dress professionally and be polite to the respondents but at the same time look self-confident. As soon as the door opens, they must smile and tell the respondents their intention. Trochim (1997) suggests that interviewers approach the respondent with a statement such as "I'd like to take a few minutes of your time to interview you for a very important study."

If they get accepted in, they should introduce themselves and start explaining the study before asking the questions. While interviewing, it is suggested that they read each question from the paper in the folder without changing the order.

In eliciting responses, interviewers wait for the answers with patience, try to encourage the respondent by indicating that they are listening to them with interest. They can ask for elaboration, clarification for the response that seems to be ambiguous or not to the point. If they do not understand or hear the answer properly, they can ask the respondent to repeat it. The responses can also be recorded. In that case, the probes that are used need to be indicated on a separate paper.

After the interview, the interviewers thank the respondent and ask whether they would be interested in learning about the results of the study. As soon as they leave, they might want to write their impression of the interview and the respondent that would have some effect on the result of the study.

Advantages and Disadvantages of an Interview

In direct interviews, one gets higher response rates than from questionnaires because questionnaires that are mailed are usually ignored, or are completed carelessly. Moreover, an item in a questionnaire may sound incomprehensible to the respondent, and for that reason, he/she leaves it out. In an interview, however, all the misconceptions are resolved because the interviewee has the opportunity to ask for elaboration or clarification for questions that seem ambiguous.

The disadvantages of conducting an interview:

- It might be difficult to get an appointment with everybody to be interviewed.
- Too much time is spent for interviewing as compared to giving questionnaires.
- If the people to be interviewed are located far from where the researcher lives, he has to do a lot of traveling.
- If there are other people helping with the interviewing, then there is the danger of interviewer bias. People's reactions may vary according to the behavior and attitude of the interviewer.

The last drawback can be controlled by

- conducting a very formal interview so that interviewers do not have to use their own words.
- training the interviewers as to how to ask the questions and how to behave during the interview.

Observation

The use of observation as an instrument for data collection derives from the belief that the researcher can gain more accurate and deeper understanding of people's values, structures, and conflicts from their observed actions rather than from their statements obtained from questionnaires or interviews.

The development of an individual, an issue, or an event involved in a particular case may be observed in order to investigate the procedure or the progress. Ecologists conduct a great deal of observations about living organisms and wild life in nature by taking zoom-lens pictures. Certain habits of children, the behavior of psychologically disturbed people, or the

language development of children or second language learners are observed in the same manner by taking notes about the behavior.

The observer can take different roles depending on the degree of his/her participation within that particular situation to be investigated:

1. Complete participant (covert observation)
2. Participant as observer (overt observation)
3. Observer as participant
4. Complete observer

The observer, as a **complete participant**, fully participates in the situation without identifying himself/herself. Since the identity of the observer is covered up, this type of observation is called covert observation. Observing people without their knowledge is usually rejected for ethical purposes although there is a better chance for the researcher to get a true picture of the situation.

When the observer **participates in the activity as an observer**, then the subjects become aware of his/her existence as an observer although he or she participates in the activities. For that reason, this type of observation is called overt observation. In such a case, the subjects may be reluctant to participate in the activities as they normally do at other times. This would lead to misleading results.

When the researcher takes the role of **observer as participant**, he/she tries to talk to the subjects, sits in as a participant during the whole activity but does not participate in the activities.

If the researcher takes the role of **complete observer**, he/she does the observation as an outsider without getting involved in the activities.

Depending on the participation of the observer, two techniques are commonly employed in observing the behavior and progress of the human beings/living organism or any worthwhile object or organization:

- Participant observation(see Vol. 1, Chapter 5, Part I)
- Non-participant observation

Participant Observation

Participant observation is an excellent method to derive information directly from the source because people may withdraw in reflecting the reality while answering the questionnaire. In the case of observation, this is eliminated because the researcher relies on what is seen rather than what is told. The participant observer, having a specific issue in mind, immerses himself or herself in the life of the individuals in order to understand how that issue is dealt with and records his or her impressions at site.

Hill, Howarth and McNamara (1996) also provide strong grounds for the employment of participant observation in the assessment procedure of the Competency-Based Learning Model.

Non-participant Observation

In non-participant observation, the observer either observes the individuals **naturally** without interfering with their activities or, in order to elicit certain information, designs a special situation for the subjects and makes them act accordingly. The best example of this kind is the **simulation** where subjects are involved in role-playing. Sometimes, instead of

observing a group of subjects, the observer tries to derive information from one single person or a place as in case studies.

Ethnographic research is a combination of the two, which tries to portray the system focusing on the physical, social, emotional, and intellectual aspects of the individuals as well as the setting. For instance, in the case of a school, the researcher observes the relation between students- teachers, administrators-students, administrators-teachers, and pays attention to classroom settings, teacher meetings, teacher-parent meetings, etc. in order to perceive a global picture of the school.

Disadvantages of Observation

Although observation is a good technique in obtaining information from the real situation, it has some drawbacks and for that reason the researcher has to proceed with extreme caution.

- **Subjectivism:** Observation technique is the least objective way of obtaining information. The researcher tends to be biased. It is very difficult to exclude one's impressions while reflecting one's observation.
- **Ethics** of documentation. During the observation, the notes taken by the researcher may be too confidential to be circulated.
- **Difficulty in application.** Usually, the observer, being an outsider, may not be familiar with the area and the environment and consequently may have difficulty in eliciting or deriving the desired information. When the setting is completely new to the researcher, he or she could easily be carried away by what is happening around and forget focusing on the issue to be searched. Moreover, observation is a time-consuming activity. Since most researchers work under time constraints, observation technique seems to be less suitable for investigating the existing situations and values.
- **Effect of the observer.** The presence of the observer may negatively effect the subjects' behavior and the outcome of the study would not reflect the reality.

Test Results

In order to measure the level of people's performance or their physical situation, different types of tests can be administered. The results of these tests help professionals to give the right job or task, or the right treatment, to these people. The most important issue is to give a reliable and a valid test to arrive at an appropriate solution.

Experiments

Sometimes, to investigate the results of a particular treatment, two different subject groups are chosen randomly. Keeping all the variables the same, one of the groups is given a particular treatment. At the end of the treatment, the difference between these subject groups is measured and evaluated. This is a very common method of collecting data for most scientists. This type of experiment is excellent for finding a solution to a problem or finding a cure to a disease. It also serves to develop a better tool or strategy in conducting or accomplishing a task, or producing a new product (see Vol. 1, Chapter 5, Part II for details).

Evaluating the Instrument to Be Used

in Data Collection

Both reliability and validity are crucial in considering a good measurement instrument because they are the researcher's "first line of defence against forming inaccurate conclusions (i.e., inaccurately accepting or rejecting your research hypotheses) (Trochim, 1997). The proof on the reliability and the validity of the instrument enables the researcher to proceed with the study with confidence because he/she knows that the instrument chosen is dependable and will measure what is intended to measure.

Anyone designing research must keep in mind that the measurement instruments to be used for the study must both be reliable and valid. If they are not reliable and valid, they are unlikely to measure what is intended to measure. In other words, the chance for accurate measurement decreases.

"The relationship between reliability and validity is a simple one to understand: a measurement can be reliable, but not valid. However, a measurement must first be reliable before it can be valid. ... In other words, a measurement may consistently assess a phenomena (or outcome), but unless that measurement tests what you want it to, it is not valid" (Trochim, 1997).

Reliability

Reliability is the extent to which a measurement instrument produces consistent results when administered under similar conditions. However, there are some factors such as measurement error, student fatigue, test setting problems that may contribute to unreliability.

Reliability is the ratio of three variance components. Therefore, in computing the degree to which any test is reliable, there are three important terms:

- total variance (= 1)
- true score variance (V_t)
- observed score variance (V_x)

$$\text{Reliability} = \frac{\text{true score variance}}{\text{observed score variance}} = \frac{V_t}{V_x}$$

Observed score variance includes both true score variance and error variance, so observed score variance is equal to true score variance plus error variance (V_E).

$$\text{Reliability} = \frac{V_t}{V_t + V_E}$$

The reliability coefficient is almost always less than 1. If there were no error, the test would be perfectly reliable and reliability would equal 1. If the error variance is very large, the reliability figure will be very low. Reliability ranges from 0 to 1. We cannot observe true variance. We must get at it through the observed variance. The observed variance is that which we have found in our test. Part of it is true variance due to students' performance and part of it is error variance.

$$\text{Reliability} = 1 - V_E$$

Ways to Estimate ReliabilityTest-retest

When we correlate test-retest scores, we are interested in stability of results over time. Reliability is obtained by administering the test to the same students twice and computing the correlation between the two administrations. The correlation coefficient is the reliability coefficient.

Parallel Tests

The correlation between two parallel tests is called a coefficient of equivalence. The correlation is the reliability coefficient.

Internal Consistency Methods

Split-half method. Split the test into two similar parts. Then correlate the scores on the two halves of the test as if they were two separate tests. The correlation between the two halves gives us the reliability for half the test. We then use the Spearman Brown prophecy formula to determine the reliability of the full test.

$$\begin{aligned} r_K &= \text{reliability of full test} \\ r_1 &= \text{reliability of half the test} \\ r_K &= \frac{2r_1}{1 + r_1} \end{aligned}$$

Suppose we split a test and correlate the students' scores from the two halves and get a correlation of .87. In order to determine the reliability, we apply the r_K formula.

$$\begin{aligned} r_K &= \frac{2r_1}{1 + r_1} \\ &= \frac{(2) (.87)}{1 + .87} \\ &= \frac{1.74}{1.87} \\ &= .93 \end{aligned}$$

Kuder-Richardson formula 20. In Kuder-Richardson formula 20, the average correlation between pairs of items on the test is calculated. Then this correlation, which is equal to average item reliability, is adjusted for the number of items in the test.

K means number of items (50-item test).

r_{ij} is the mean item correlation (.30).

$$\begin{aligned}
 KR - 20 r_K &= \frac{K r_{ii}}{1 + (K - 1) r_{ii}} \\
 &= \frac{(50) (.30)}{1 + (50-1) .30} \\
 &= \frac{15}{1 + 14.7} \\
 &= .96
 \end{aligned}$$

Kuder-Richard Formula 21

$$KR-21 r_K = \frac{K}{K-1} \left(1 - \frac{X(K-X)}{K s^2} \right)$$

K = number of items in a test

X = mean of the sample

s^2 = variance of the sample

Suppose we have a 100-item test, for which the mean is calculated as 65 and the sample is 10.

$$\begin{aligned}
 &= \frac{100}{99} \left(1 - \frac{65(100-65)}{(100)(10)^2} \right) \\
 &= \frac{100}{99} \left(1 - \frac{2275}{10000} \right) \\
 &= \frac{100}{99} (1 - .23)
 \end{aligned}$$

$$= \frac{100}{99} \quad (.77)$$

$$= .78$$

When the items are to measure the same ability, one of the interval consistency methods is used. If the test measures different variables, a test-retest method is applied.

The standard error of measurement (SEM) will show us the averaged differences between the observed scores and the true scores.

$$SEM = s_x \sqrt{1 - r_{xx'}}$$

There are two major kinds of tests: norm referenced and criterion-referenced. Norm-referenced tests allow us to make comparisons among individual students because their scores spread out in the distribution. In such tests, the more differentiation there is among students, the larger the variance gets.

Factors Affecting Reliability of a Test

Criterion-referenced tests compare all the students with a predetermined criterion (not in relationship to each other). In such cases, we are not likely to find a large variation. If we set 80 out of 100 as a criterion for passing our course, then all the students who have mastered the objectives will have passed; and consequently, there will not be much variance.

Following are among the important factors affecting reliability which will increase the reliability of a test:

- Length of the test: the longer the test, the more reliable it is.
- Homogeneity of items: if the items are testing the same trait(s), the reliability will be higher.
- Discriminatory power of items: items which discriminate well among students will increase the reliability.
- Variability of group ability: if the students have a wide range of ability, the test reliability will increase.
- Sufficient test-taking time: reliability formulas give better estimates of reliability on power tests than on speed tests. If speed is a factor, the split-half method is not used.

Validity

Validity is defined as the extent to which the results of the procedures to be applied serve the intended purpose. It is the extent to which a test measures what it is intended to measure. There are four types of validity: content validity, criterion-related validity, concurrent validity, construct validity.

- A. **Content validity** is the extent to which a test measures a representative sample of the content matter. If the test is indicated to have content validity it we can be sure that the test we use comprises of items covering the subject matter we are interested in testing.
- B. **Criterion related validity** is considered when test scores are to be used to predict future performance or to estimate current performance on some valued measure other than the test itself.
- C. **Concurrent validity** is taken into consideration when two tests are desired to measure the same variable.
- D. **Construct validity** tells us if the test performance can be described psychologically.

There are several factors influencing validity:

- Unclear directions within the test
- Inclusion of too difficult vocabulary items or sentence structures within the test items
- Inappropriate level of difficulty of test items
- Poorly constructed test items
- Ambiguity
- Test items inappropriate for the purpose of the test
- Insufficient number of items for objectives being tested (for detailed information language testing or statistics for applied linguistics see Bachman, 1990; Brown, 1990; Hatch & Farhady, 1982).

Data Preparation

Ways of Entering Data into the Computer

Data collection is not completed unless the collected information is entered into the computer after it has been checked for accuracy. As suggested by Trochim (1997), this can be done by **database programs** such as Microsoft Access and Claris Filemaker or **statistical programs** such as SPSS, SAS could be used to run simple descriptive statistical programs. Although database programs are more complex and difficult to learn, they give better access to manipulating the data.

Trochim (1997) suggests that in entering the data into the computer, the researcher can make use of the "double entry" procedure.

In this procedure you enter the data once. Then, you use a special program that allows you to enter the data a second time and checks each second entry against the first. If there is discrepancy, the program notifies the user and reduces entry errors.

Since this procedure is not available everywhere and requires training, it is advised that spot checks are made. Another important issue is to look at the limits or boundaries of the data. For instance, in entering the age of the subjects, if a number with two digits are observed, this would automatically tell the researcher that there is an error in the entry.

Issues to Consider in
Transferring Data to the Computer

In converting the raw data into variables in the computer, there are several important issues to be taken into consideration. The first one has to do with blanks since a blank is treated as a missing value. Therefore, whether the value is missing or there is no value given should be somehow indicated in the program in order not to alter the results of the study.

The second issue is related with the scaling system used for the survey. Scores in the scaling system should all be in the same direction when they are transferred into the computer. In other words, if the scale for a question is given in such a way that higher values indicate negative values, then the values for this item should be reversed. For instance a person strongly agreeing with the item "I buy two newspapers everyday." would mark 5 from the scale of 1 to 5. Another person who does not care much about reading newspapers, would also agree with the statement "I do not read newspapers." In these cases, while the strong agreement indicates a positive value for one, it indicates a negative value for the other. Therefore, there should be an item reversal made for the second item.

The third issue has to do with variables such as age, time, and income that may require grouping so that the analysis would yield better results. For instance, instead of looking for the values of people at different ages, it might be better to consider them within specific age groups (e.g. 20 - 30, 31- 40 etc.) .

The important issue is related with the transfer of the that data into the computer properly. The design of the research (Vol. 1, Chapter 5) will determine the way of storing the data into the computer and the analyzing them accordingly.

EXERCISES

A. The following questionnaire items lack some main points to be considered in designing a questionnaire. Read each item and see why it has been found inappropriate. Then try to modify it accordingly.

1. Which university have you graduated from?
(aim: effectiveness of the ELT curriculum)
2. Do you oppose copyright protection for the look and feel of ELT software's?
3. "According to the TESOL (Teaching English to Speakers of Other Languages) *Newsletter* poll, 80 percent of English teachers oppose copyright protection"
What is your opinion on this issue?
4. Don't you agree that smoking should be prohibited in the student cafeterias?
5. Considering the long string of litigation, e.g. Apple versus IBM, what do you think are the long term prospects for cooperation among the computer industry?

B. Read each questionnaire item and tell the type.

1. What characteristics of a person would lead you to rate her or him a good administrator?
2. What is your opinion about the policy of "no smoking" on the campus ?
 - a) I support the policy
 - b) I am opposed to the policy
 - c) The policy does not interest me.
 - d) I am undecided about the policy
 - e) Other
3. A. Have you ever used ELT software?
 - No (Go to question B)
 - Yes

If yes, about how many hours per week?

 - Less than one
 - one
 - 2 to 5 hours
 - 6 to 10 hours
 - 11 to 20 hours
- B. Have you ever used Multimedia Software?
4. We should support copyright protection for source and object code.
 - Strongly agree Agree Neutral Disagree
 - Strongly disagree

5. How do you rate the quality of the computers in your self-access center?
Circle one of them. Worst -1 2 3 4 5 6 7 8 9 10 - Best

6. Put a check mark in the appropriate place.

Statement	Disagree Strongly	Disagree Somewhat	Agree Somewhat	Agree Strongly
Current method should be considered carefully before trying something new	[]	[]	[]	[]
Diverse experience at several companies is preferable to in-depth expertise at one company for several years	[]	[]	[]	[]
It is usually worth the risk to try a technological innovation if it has any potential of giving you a competitive advantage over proven technology	[]	[]	[]	[]

(Adapted from Trochim, 1997)

- C. Considering the criteria in developing a questionnaire, design a set of questions to assess the following traits of members of a group consisting of English teachers:
- Their teaching style
 - Their evaluation style
 - Their knowledge of ELT methods
 - Their teaching experience at different levels
 - Their educational background
- D. When might an interview survey be preferable to a questionnaire survey or vice versa? Give examples discussing the advantages and disadvantages for each case.
- E. Suppose an administrator at an English language preparatory school would like to find out how the instructors feel about the new textbook in use. Write a series of six contingency questions for an interview or a questionnaire.
- E. Look at the items of the questionnaire stated below and answer the following questions:
- What might be the purpose of a research study using such a questionnaire?
 - Keeping the purpose in mind, discuss the validity of the items.
 - Is the form of the questionnaire user-friendly? If no, how would you redesign it?
 - Try to restate two of these close-ended items in a contingency format.
 - Restate three of the items in an open-ended form.
 - Correct grammar error if there are any.

A. Please check the most appropriate choice considering the gender roles in society.

1. Considering the language they use, men () women () are more polite.
2. Women () men () emphasize on the details more when talking on a topic.
3. Men () women () use more abusive language.
4. Confirmation questions such as "isn't it? really?" are used by men () women ().
5. Expressions like "well, err, I mean" are used more frequently by women () men ().

f. What suggestions can you offer for improving the rate of responses of such a survey?

G. Which method(s) of data collection would be the most appropriate for the following studies?

- a. The error correction techniques used by ELT teachers in teaching young learners.
- b. The ideas of the subscribers of HOME & GARDEN about a recent article appeared in the magazine on the imported African decorations.
- c. Features of the new motivation program implemented in a foreign based company operating in Turkey and the long-term expectations of the executive staff in the development of the productivity.
- d. The feelings of elementary school teachers about special classes for the gifted students.

H. What could be the advantages and disadvantages of designing your own instrument or using a ready made one for data collection in your research study?

I. Study the following questionnaire items developed in different scales. Afterwards, indicate the type of drawback for each, and then try to modify it. Indicate a case (if there is one) where the utilized scale does not seem to require any modification.

a. Marital status:

- | | |
|---------------------------|--------------|
| 1) Single (never married) | 4) Divorced |
| 2) Married | 5) Separated |
| 3) Widowed | |

b. How often do you go to the movies?

- 1) Never 2) Rarely 3) Occasionally 4) Fairly often
5) Often 6) Very often 7) Almost always 8) Always

c. How often do you watch football games on TV?

- 1) Frequently 2) Occasionally 3) Seldom 4) Never

d. Mr. X's emotional intelligence is

- 1) Much below average 4) Above average
2) Below average 5) Much above average
3) Average

e. Mr. X has good class management skills.

- 1) Agree 4) Tend to disagree
2) Tend to agree 5) Disagree
3) Undecided

f. How often do you visit your parents?

- 1) Seldom 2) Occasionally 4) Frequently

g. Circle the item that you agree with.

- 1) The teacher grades our compositions fairly.
2) The teacher does not grade fairly.
3) The teacher corrects our mistakes.
4. The teacher does not correct our mistakes.

J. Would you agree that it is not ethical to observe/record people without their information and permission? Discuss your reasoning with your friends.