What is the relationship between (a) the frequency with which high-school teachers criticize students in class and (b) the frequency of students voluntarily participating in class discussions?
Are teachers who have had more formal education held in greater respect by their students than are teachers with less formal schooling?
Do kindergarteners from one-parent families display more fear in new situations than ones from two-parent families?

Expressing the degree of relationship: The extent of correlation between variables can be expressed verbally, graphically, or statistically. Here are five typical ways that people verbally describe their impression of the strength of relationship between two variables:

- Students who are good at math are also always good at science.
- Pretty girls usually have more self-confidence than do plain-looking girls.
- I'd say that about half the time the taller kids get more respect from their peers than the shorter kids do.
- Only rarely will teachers give homework on Friday—that is, homework they expect students to complete over the weekend.
- Knowing how fast different teenagers can run is never any help in estimating how well they can sing, because running speed and singing ability are completely independent of each other.

Verbal descriptions have the advantage of being easy to create, and they can also be understood without any technical preparation on the part of the listener. However, verbal descriptions can suffer from two weaknesses. They are generally imprecise and they are typically based on casual observation rather than careful study. Imprecision in the above verbal expressions is reflected in such vague estimates of relationship as usually, about half the time, and rarely. Although the word always in the first example and never in the last are precise, such extreme relationships as always and never are themselves extremely rare or nonexistent in real-life correlations.

It is also the case that judicious listeners may suspect that verbal descriptions are founded on the speaker's casual, inexact impressions or on hearsay evidence rather than on the conscientious study of a broad range of incidents that involve the pair of variables.

Thus, in order to express correlations in a more precise and convincing form, researchers design controlled studies of correlations and express the results in graphic or statistical form. To illustrate how this may be done, let's imagine that we wish to create a written test that enables us to predict the degree of preadolescents' peer sociality. By sociality we mean how adolescents' peer sociality. By sociality we mean how adolescents are interacting with agemates. We are acquainted with three different theories of social behavior on which a test might be built, and we are not sure which of the three is the most accurate. Therefore, we intend to create three tests, each based on a different theory, and try

<table>
<thead>
<tr>
<th>Name</th>
<th>Observation Score</th>
<th>Test A Score</th>
<th>Test B Score</th>
<th>Test C Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann</td>
<td>54</td>
<td>33</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>Bart</td>
<td>37</td>
<td>26</td>
<td>18</td>
<td>24</td>
</tr>
<tr>
<td>Chelsea</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>Darrell</td>
<td>43</td>
<td>19</td>
<td>23</td>
<td>20</td>
</tr>
<tr>
<td>Eve</td>
<td>32</td>
<td>22</td>
<td>19</td>
<td>22</td>
</tr>
<tr>
<td>Freddie</td>
<td>46</td>
<td>17</td>
<td>31</td>
<td>15</td>
</tr>
<tr>
<td>Gwen</td>
<td>27</td>
<td>28</td>
<td>18</td>
<td>26</td>
</tr>
<tr>
<td>Harold</td>
<td>19</td>
<td>14</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Ivy</td>
<td>26</td>
<td>18</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>Jack</td>
<td>38</td>
<td>16</td>
<td>27</td>
<td>19</td>
</tr>
<tr>
<td>Kelly</td>
<td>33</td>
<td>21</td>
<td>28</td>
<td>21</td>
</tr>
<tr>
<td>Lance</td>
<td>25</td>
<td>35</td>
<td>16</td>
<td>33</td>
</tr>
<tr>
<td>Mavis</td>
<td>46</td>
<td>20</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>Ned</td>
<td>35</td>
<td>14</td>
<td>25</td>
<td>34</td>
</tr>
<tr>
<td>Oprah</td>
<td>39</td>
<td>23</td>
<td>31</td>
<td>24</td>
</tr>
<tr>
<td>Paul</td>
<td>41</td>
<td>20</td>
<td>23</td>
<td>25</td>
</tr>
</tbody>
</table>
scores on the 60-point observation schedule and their scores on the three 36-item tests. That information, along with the fictitious names by which we identify the participants, is shown in Table 5-1. It is very difficult to estimate the relationship between the observation scores and any of the three tests' scores when the data are simply listed as they are in the table. But if we recast the data graphically as a scatter diagram, the nature of the correlations becomes immediately evident. To prepare a scatter diagram, we arrange the test scores on the vertical axis and the observation scores on the horizontal axis. Then we plot the points at which each pupil's score on a test intersects with that pupil's score on the observation schedule. Scatter diagrams for the three tests, in relation to the observations, are displayed in Figures 5-1, 5-2, and 5-3.

**Figure 5-1**
The Relationship Between Observations and Test A

![Graph showing the relationship between observations and Test A](image1)

**Figure 5-2**
The Relationship Between Observations and Test B

![Graph showing the relationship between observations and Test B](image2)

**Figure 5-3**
The Relationship Between Observations and Test C

![Graph showing the relationship between observations and Test C](image3)

Viewing the three scatter diagrams enables us to judge which test comes closest to measuring the same characteristics as does the observation schedule. We make this judgment by noticing how closely the array of dots assume a straight line extending from the lower-left corner to the upper-right corner of the diagram. If the dots formed a straight diagonal line, it would mean that the students who scored highest on the test also scored highest on the observation schedule and vice versa. But the more the dots spread all over the diagram, the lower the degree of relationship between the test and the observations.
Thus, by inspecting the three diagrams, we recognize that Test B is the one that best predicts students' observation scores, since the dots quite obviously extend from lower left to upper right, although not in a precise line. So, the extent of correlation between Test B and the observation schedule is high, although not perfect. Our inspection of Figures 5-1 and 5-3 shows that the dots in both tend very slightly to spread along the diagonal, but it's impossible to tell whether one of those two patterns represents a greater relationship than the other between test scores and observation scores. To settle our puzzlement, we can apply statistical analysis to the data. Among the several computational techniques that we might use, we choose the most popular one—Pearson's product-moment correlation technique (symbolized by the letter $r$). Calculating $r$ for the three scatter diagrams results in a correlation coefficient of +.30 for Test A, of +.89 for Test B, and of +.34 for Test C. The closer a correlation comes to +1.0, the stronger the relationship between the two variables. Hence, from the vantage point of our statistics, the correlation between the observation scores and Test B is very high, and it is quite low for Tests A and C, with the relationship for Test C just slightly greater than for Test A.

Therefore, as our examples have illustrated, statistical descriptions of relationships are more precise than graphic displays, whereas graphic displays are typically more precise than verbal descriptions.

**Purpose and illustrative studies:** Correlation studies are usually conducted for one or more of three reasons: (a) to show how closely changes in one variable are related to changes in another, (b) to permit predictions of one variable by knowing how closely the two variables are related, or (c) to suggest how the condition of one variable has caused the condition of another variable.

An error easily committed by people who interpret correlations is to assume that every described relationship between variables—such as the relation between height and self-confidence—implies that one of the variables is responsible for the other. This is the assumption that demonstrating a degree of relationship between variables inevitably implies that one variable caused the other. The fallacy of this assumption can be illustrated with an anecdote about the relationship between the frequency of storks in Dutch communities and the incidence of births in such places. It was commonly observed that the greater the number of storks nesting in a community, the larger the number of children born there. If we were now to assume that one of these variables (quantity of storks) was the cause of the other variable (quantity of infants), then we have the evidence necessary to support the age-old canard about storks delivering human babies. However, the relationship between the quantity of storks and the quantity of babies in a community is—from both scientific and commonsense viewpoints—no more than coincidental. Thus, a demonstrated degree of relationship between two variables cannot be taken as evidence that one variable's condition caused the other's condition—either entirely or only partially. What is required for demonstrating cause is a line of logic demonstrating that a change in one variable was necessary for producing change in the other. In the stork/baby instance, there is a host of evidence suggesting that the cause of the production of human babies has nothing to do with the correlation between the stork population and the newborn-human population. In effect, it is impossible to convincingly argue a causal relationship between (a) the frequency of human parents' cohabitation (along with information about the current condition of the parents' sperm and ovum supplies, and the lack of obstruction to sperms reaching ovum) and (b) the frequency of human births. In addition, the stork/human-babies correlation can be explained without any assumption that storks bring babies. For instance, observations of storks in the Netherlands had shown that the birds preferred to nest in quieter, unpolluted villages rather than in bustling, smoky cities. And just by coincidence, villagers produced more infants per family than did city folk. In sum, a demonstrated correlation between variables can be either coincidental (usual) or causal. Any proposal that a relationship is causal must be supported by empirical evidence (facts) and a persuasive line of reasoning.

In school settings, demonstrating a likely causal relationship between variables can be particularly valuable in suggesting how teachers might promote students' progress and personal welfare. For instance, a study that included a wide range of fourth-grade classrooms in the United States showed a positive correlation between pupils' reading skills and the number of books in classroom libraries (Postlethwaite & Ross, 1992, pp. 36-37). Thus, it is not difficult to argue that increasing the availability of books in classroom libraries could be one reason for children's improving their reading competence. The correlation and its causal interpretation can then serve as a guide to action, that is, to enlarging classroom library holdings. Furthermore, in the same study, the schools whose pupils were the most competent readers had teachers who ensure that their students read a great deal in class, who have their students visit the school library on a regular basis, and who, to a lesser extent, set more reading homework, ask questions about the homework the next day, and devote more time per school week to the teaching and practice of reading. (Postlethwaite & Ross, 1992, pp. 45-46)

However, even when a convincing case cannot be made for a correlation being of a causal variety, that correlation may still be of practical value for predicting an outcome. For example, imagine that we create a *Business Decisions and Attitudes Inventory* composed of 30 items, with
each item requiring students to select which among several decisions they would consider the wisest in a business operation. We now administer the inventory to the 98 students in the four sections of our high school's course titled Business Practices. Next, we compute the degree of correlation between (a) students' scores on the inventory and (b) their total scores on classroom assignments (tests, homework, in-class projects). Let's assume that our computation reveals a very strong relationship between students' inventory scores and their total class-progress scores. Students who ranked high on the inventory also ranked high in academic performance, and vice versa. This information suggests that if we know a student's score on the inventory, we can make a pretty good estimate of how well that student will succeed in class work. If we can assume that students' ability to answer the items on the inventory is, to a great extent, the result of knowledge and intellectual ability that they possessed before entering the Business Practices class, then at the beginning of the semester—when students first enter the class—we can administer the inventory and estimate with some confidence how each student will probably succeed in the course.

In summary, when the correlation between two variables is very strong, our knowing an individual's position on one variable equip us to offer a rather accurate guess about the person's position on the other variable. We can do this without needing to assume anything about the causes behind the individual's status on either of the variables.

Procedure: There are numerous statistical ways to calculate the magnitude of the relationship between variables. Each way is designed to suit the particular sort of information that has been gathered in a research project—such as data in the form of test scores, of gender, of age, of answers on a study-habits inventory, of parents' educational levels, and more. Such statistics textbooks as the following describe which correlation methods are appropriate for which kinds of data, what steps to take in calculating the degree of correlation, and the advantages and disadvantages of each method.


Ex-post-facto. This is the simplest of the designs, one that fits easily into regular classroom procedures since it involves no alteration of usual teaching procedures. It consists of (a) applying a treatment to an individual student or a group, (b) evaluating the student's or group's performance following the treatment, and (c) estimating how much the treatment contributed to the final performance.
This is obviously the typical pattern of teaching-and-evaluating that is routinely practiced in classrooms. But the procedure can also qualify as research if the teacher (a) is seeking to discover the effectiveness of a newly attempted instructional procedure and (b) intends to base future teaching methods on the results of the experiment. The project particularly qualifies as "applied research" if the teacher (c) writes up the findings and circulates them for the edification of other teachers.

To illustrate, a fifth-grade teacher's social studies unit focused on the history and culture of the American Indians who populated the local region before the arrival of European settlers. In pursuit of this objective, she introduced a new class activity—an excursion to the city's natural-history museum so her pupils could view dioramas and artifacts from the past. When the class returned to school, she administered a test covering information the pupils were expected to have learned from viewing the museum exhibits and from hearing the museum guide's description of the events portrayed in the dioramas.

When interpreting the outcome of the excursion, the teacher accepted the pupils' test scores as indicators of how much they had gained from the excursion. However, her research methodology failed to account for how much the pupils may already have known about the topics on the test before they went to the museum. Some pupils may have visited similar museums on earlier occasions. Or they may have read books or seen television programs that included background material useful in answering the test questions. Therefore, the teacher could not be confident that pupils' success with the test was an accurate indicator of what they had learned on the field trip, so the problem of pupils' prior knowledge reduced her faith in conclusions she might draw from the test results.

Pretest-treatment-posttest. The next time the fifth-grade teacher planned to assess the value of a new teaching procedure, she could—with a bit of extra work—go a long way toward resolving the problem of the learners' prior knowledge. Before the pupils experienced the new approach (such as the museum excursion), the teacher could prepare a test (pretest) covering the information that pupils would be expected to learn from the trip. The pretest would be taken by the pupils before leaving on the excursion. Then, after their return to school, they would take a similar test—a posttest that focused on the same topics as the pretest but whose items were phrased in a different manner so that children could not answer the posttest questions simply by rote-remembering pretest answers. In other words, the posttest would an equivalent form of the pretest, assessing the same content but with test items worded differently.

When the teacher drew conclusions about the educational success of the field trip, she wouldn't base her conclusions solely on the students' scores on the posttest. Instead, she would subtract each pupil's pretest score from his or her posttest score to find a change score, which would be the difference between what the child knew before the trip and what he or she knew after the trip. The change score would be interpreted as showing how much the pupil profited from the trip. Totaled all pupils' change scores and dividing by the number of pupils who took the trip would produce an average of how much the class in general had learned.

This same research design could be extended to provide even more information about the value of the field trip if the teacher added another feature—a test or two administered on later occasions to reveal how well the pupils' learning was retained over time.

Multiple treatments. Whereas the pretest-treatment-posttest design could furnish a fairly trustworthy estimate of how much the children gained from the museum field trip, it wouldn't tell how effective such excursions might be in comparison to other ways the pupils could pursue the same learning objectives—such ways as viewing videotapes or reading a textbook. The following is a design intended to provide the desired comparative information.

In this case, the teacher begins by administering the pretest to the entire class. She then divides her class of 30 pupils into two groups, assigning half of the pupils to Group A and half to Group B. The intent is to have the two groups be as much alike as possible at the outset of the unit in terms of the pupils' knowledge of Indian culture. One way to form comparable groups would be to assign children to the groups by random means (writing each child's name on a slip of paper, mixing up the slips in a bowl or hat, drawing the slips out one at a time, and alternately assigning them to Group A and to Group B). Another way would be to place children in pairs on the basis of their similarity in qualities that the teacher estimates will contribute to their learning abilities (such qualities as academic aptitude or past experience with museums and
videotapes). A third way would be to pair up children on the basis of their similar pretest scores. One child from each matched pair would be assigned to Group A and the other child to Group B.

Next, each group receives one of the two treatments. While members of Group A are on the excursion, members of Group B view the videotapes. Subsequently, members of both groups take the same posttest which is an equivalent version of the pretest, focusing on the same objectives as the pretest but with the test items worded differently. To determine which method—excursion or videotapes—was the more effective instructional procedure, the teacher computes change scores for each pupil, sums the change scores, and calculates the average-change-score for each group. The teacher will conclude that the more effective teaching procedure was the one experienced by the pupils in the group that earned the higher average change score. And the greater the difference between Group A’s and Group B’s change scores, the greater the confidence the teacher can place in her conclusion that the higher change score is a valid indicator of the superiority of one teaching method over the other—a least with this year’s fifth-grade class and with the teacher’s particular instructional style.

More on sampling: The remarks about sampling on pages 70-73 apply not only to surveys but also to experiments. If the fifth-grade teacher did, indeed, perform her research project with care (well-constructed tests, well-conducted excursion, well-presented videotapes), she could rather confidently draw conclusions about which of her two teaching methods was the more effective. However, she would be on shaky ground if she proposed that the same results she derived with her class would be true for other teachers in other classrooms. First, we should recognize that when she divided her class of 30 pupils into two groups, each group contained only 15 members. It would require a great leap of faith to suggest that (a) what was learned during an excursion by those particular 15 fifth-graders accurately reflected what children in other elementary schools would learn on a museum trip and (b) what was learned from videotapes by those particular 15 pupils accurately represented what children in other elementary schools would gain from viewing videotapes. Therefore, the best this teacher could probably do—in addressing the issue of how broadly the results of her research should be applied—would be to add a paragraph at the end of her report that said:

In the present study, the technique of having fifth-graders view videotapes about the history and culture of an American-Indian nation proved more effective than did having pupils visit a natural-history museum. But we cannot conclude with any degree of confidence that videotapes would, in general, be superior to excursions for the study of other subject-matter by other pupils in other elementary-schools. Further research is necessary for settling the question of the comparative value of excursions and videotapes.

The simple research designs described above represent only a few of the forms that experiments can take. Additional designs are available to fit the demands of other research aims and contexts. The nature of various designs, situations for which they are appropriate, and the steps followed in applying them are described in such publications as the following:


Planning Guide

1. Carry out the following series of steps in relation to a question—or a series of questions—that you might like to answer by conducting a research project that uses a quantitative method.

2.1 State your question or questions.

2.2 Describe in some detail the method—or combination of methods—you possibly could use to answer your research question.

2.3 State the criteria you will apply in deciding which of several methods might be the most appropriate. Your criteria might include such considerations as the time, expertise, and funds you are willing to dedicate to the project.

2.4 Apply your criteria to your potential methods and decide which method would likely be most suitable. Explain the line of reasoning you followed to arrive at your final choice. For example, which of your criteria weighed most heavily in the selection process? And why did you give high priority to that criterion?
2. On the line in front of each of the following questions, write the identification letter of the research method that you believe would be most suitable for answering the question. If you believe that a combination of more than one method would be appropriate, then write more than one identification letter on the line.

**Research Questions**
- In Martinvale elementary schools, are girls more adept than boys at searching the Internet?
- Over the past two decades, what research methods have been most often used for studying the teaching of beginning reading?
- Do high school students learn American history more adequately from reading a textbook than from witnessing illustrated lectures?
- In what pattern did the city elementary schools' math curriculum evolve over the period 1930-2000?
- What percent of classrooms in the county's public schools have access to the World Wide Web?
- What rituals in boy-girl relations are reflected in the behavior of students in Elk Grove Middle School?
- In Franklin Elementary School, to what extent are families' socioeconomic positions related to pupils' academic success?
- In Larkspur School, does students' behavior differ when a teacher's aide is in the classroom from their behavior when no aide is present?
- How did different students act during the knifing incident at Morgan High, and how did those actions influence students' subsequent interactions with each other?

**Research Methods**
- A. Direct-data survey
- B. Literature survey
- C. Correlation investigation
- D. Ex-post-facto experiment
- E. Single-group pretest-treatment-posttest experiment
- F. Two-group pretest-treatment-posttest experiment
- G. Ethnography
- H. Experience narrative
- I. History

### 6 Techniques for Gathering Data

Each of the research methods described in Chapters 4 and 5 depends for its information on one or more of five data-collecting techniques—observations, interviews, questionnaires, content analyses, and tests. For example, the substance of a biography may be drawn both from interviews with people who knew the biographee and from the content analysis of letters and newspaper articles about the biographee. A study of sexual harassment in high school classrooms may be based on students' questionnaire responses and on observations of students' classroom behavior.

An experiment comparing textbooks with class discussions as devices for altering students' racial attitudes may use both questionnaires and interviews. Conclusions drawn from a science-class experiment about the influence of group-study versus individual-study of botanical terms can be based on achievement-test scores.

The purpose of this chapter is to define the five data-gathering techniques and to suggest effective ways to employ them in teachers' research.

### Observations

Observation involves collecting information by seeing and/or hearing events, then recording the results in a form suited to the needs of the research project. The following discussion focuses on three tasks—witnessing events, selecting an observation procedure, and recording what has been observed.

### Witnessing events

Observations can be either (a) immediate and direct or (b) postponed and mediated. The immediate/direct variety involves the observer—such as a teacher or a student—witnessing events as they occur. The