LEARNING QUESTIONS

3.1 What is the scientific method and how is it used to study development?

3.2 What are the different types of research designs used to study development?

3.3 What must we consider when interpreting and communicating the results of a study?

3.4 What special precautions must be used when our research participants are children?

Master these objectives using an online action plan at edge.sagepub.com/levine3e
Chapter 3: How We Study Development

TEST YOUR KNOWLEDGE

Test your knowledge of child development by deciding whether each of the following statements is true or false, and then check your answers as you read the chapter.

1. T □ F □: We know that watching television causes children to have shorter attention spans because many teachers report that this is what they see in the children in their classrooms.

2. T □ F □: If we conduct research by interviewing students in a large high school, we can assume that our findings will apply to any adolescent who is the same age as the ones we have studied.

3. T □ F □: When conducting research by doing an observation, it is important that the person who is doing the observation does not know the purpose of the research.

4. T □ F □: If observations are carefully done, you will be able to determine the causes of the behavior you are observing.

5. T □ F □: Children's memories are good enough to allow them to give reliable eyewitness testimony.

6. T □ F □: Studying a single individual intensively is a valid scientific methodology.

7. T □ F □: An experiment always consists of a situation set up by researchers to test specific hypotheses.

8. T □ F □: Even if research consistently finds that mothers who talk to their children a great deal have children with high self-esteem, we should not conclude that frequent conversations with parents build self-esteem in children.

9. T □ F □: Research has found that boys who watch a lot of violence on TV are more aggressive, but you know someone who watches a lot of violence and you see that he is not at all aggressive. This disproves the research.

10. T □ F □: Once we have established that the results of our research are statistically significant, we can be confident they will have an impact on real-world situations.


In this chapter, we look at how researchers study children and adolescents to add to our understanding of growth and development. As we said in Chapter 2, we all have some intuitive beliefs about development, often based on our own life experiences, but theories must be subjected to rigorous scientific testing to determine whether or not they are valid. This can be done in a number of ways, which means that there isn’t one best way to study development. Rather, we look for the most appropriate method to investigate the particular topic we want to examine.

We take you through the steps of the scientific method: developing a hypothesis, choosing who to study, and figuring out how to measure the concepts you want to explore. For each step you will find an Active Learning feature that allows you to apply what you are learning and check your understanding of the concepts as they are presented. After we have taken you through the scientific research process, we discuss how to make sense of the results of a research project and then describe the important process of communicating results to a wider audience so that they become part of the ongoing scientific conversation. Finally, we focus on the importance of protecting the welfare of children and adolescents who take part in this research.
Part I: Understanding Development: Why and How We Study Children and Adolescents

THE SCIENTIFIC METHOD

What is the scientific method and how is it used to study development?

Child development is one of many disciplines that uses the scientific method to add to its body of knowledge. This approach helps us organize the information that we currently have and generate new ideas that extend our understanding. It also is a self-correcting process because the findings from our current research guide our future efforts. We begin by asking a question (often based on our observations or theoretical ideas), next we identify the factors or elements that need to be examined to answer that question, and then we put our question to the test. Based on what we find, we can accept or reject the premise on which the original question was based (Salkind, 2004). Each of these steps is described in more detail in the remainder of this chapter.

BASIC AND APPLIED RESEARCH

The primary goal of much research is to add to our understanding of the phenomena we are interested in and to help us refine our theories. This type of research is called basic research because whether the results have any immediate application is not the primary concern. On the other hand, there is research that clearly intends to help us make changes that will affect people’s lives. This is called applied research because its goal is to solve immediate problems or improve the human condition. Applied research might look at different parenting styles, classroom practices, or health care policies with the goal of improving what we do. Of course, this is not a black-or-white situation. We often find that basic research lays a foundation for later applied research.

DEVELOPING HYPOTHESES

Because the field of child development is grounded in science, we rely on the scientific method to build our understanding of it. Figure 3.1 illustrates this process. This process often starts with our observations. If you spend any amount of time watching children, whether you are looking at their moment-to-moment behavior or at how they grow, change, and develop over time, you probably will have some questions about what you see. Theories, such as those you read about in the last chapter, are developed to answer the questions that arise from our observations. From these theories we develop predictions about what children will think, feel, and do in certain situations. These predictions are called hypotheses.

No matter how much sense a particular hypothesis seems to make, it still needs to be tested. The methods described later in this chapter are some of the ways that we test hypotheses. If we cannot find a way to subject a hypothesis to a test, it has little or no scientific value and is little more than speculation. For example, you may believe that watching too much television can cause children to have a short attention span, but until this belief is tested and supported by research, it remains just a hypothesis. We use an example of research that was designed to test this hypothesis in the set of activities called Active Learning: The Scientific Method that appear throughout this chapter. You can begin these activities by trying Active Learning: The Scientific Method—Forming a Hypothesis.

---

**Scientific method** The process of formulating and testing hypotheses in a rigorous and objective manner.

**Basic research** Research that has the primary goal of adding to our body of knowledge rather than having immediate direct application.

**Applied research** Research that has the primary goal of solving problems or improving the human condition.

**Hypothesis** A prediction, often based on theoretical ideas or observations, that is tested by the scientific method.

---

© by Dr. J. A. Beekman, 2013

We know that watching television causes children to have shorter attention spans because many teachers report that this is what they see in the children in their classrooms. False

---

Testing a hypothesis. Does watching television affect children’s ability to stay on task and to stay focused when they need to? Research by Levine and Waite (2000) has tried to answer this question.

Applied research Research that has the primary goal of solving problems or improving the human condition.

Hypothesis A prediction, often based on theoretical ideas or observations, that is tested by the scientific method.

---

Copyright ©2018 by SAGE Publications, Inc.

This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
The Scientific Method—Forming a Hypothesis

Research articles most often begin with an abstract that briefly summarizes the entire article.

1. First read the following abstract of an article by Levine and Waite published in 2000.

Abstract

To evaluate the common assumption that television viewing is related to attentional difficulties in school, 70 fourth and fifth grade students recorded a “television diary” for one week and reported their preferred television shows. Parents estimated their child’s television viewing time and reported their child’s preferred shows. Assessment of attentional difficulties included teacher ratings, parent ratings, standardized tests, and classroom observations. It was found that the amount of television a child viewed was significantly related to teacher ratings of attentional difficulties, but not to parent ratings, classroom observations, or a standardized test. Type of shows viewed did not relate to any attentional outcome variable. There was a clear relationship between fourth and fifth grade children’s ability to pay attention in school, as assessed by their teacher, and the amount of time they spent watching television.

(Continued)
2. After reading this abstract, look again at the first sentence. Write down what you believe to be the major hypothesis, or prediction, of this study.

Hypothesis: ________________________________________________________________

In the first section of their paper, in which the authors review what other researchers have found that leads to their own hypothesis, they state: “The major hypothesis of this study is that for children in later elementary school the amount of television viewed will be related to ‘ADHD behaviors,’ especially as shown in the classroom” (Levine & Waite, 2000, p. 667). In other words, the authors are predicting that the more television children watch, the more difficulty they will have maintaining their attention. Note for now that the hypothesis says “will be related to” and not “will cause.” We will return to this point when we discuss experimental and correlational research designs later in the chapter.

OPERATIONALIZING CONCEPTS

To test a hypothesis, we first need to find some way to turn the concept into something we can see and measure. In scientific terms, we operationalize the concepts we want to study. These concepts are then referred to as variables, that is, concepts or characteristics that can be measured and can have different values. However, this process is not as simple as it may sound. Suppose you are interested in studying aggression. You probably think that aggression is one of those things that you would recognize if you saw it, but would you? It might be fairly
straightforward if we only considered hitting, kicking, or pinching someone as the way we operationalize aggression, but that would only tap into physical aggression. Maybe we should add verbal aggression, such as name-calling, swearing, and screaming to our list. But isn’t it also pretty hurtful when someone spreads rumors about you or tries to turn other people against you? These are examples of a type of indirect aggression called relational aggression.

Measures of any or all of these types of aggression might be included in our operationalization of the concept. The point is that you must make clear what you are including in your definition of aggression and what you are not. Active Learning: The Scientific Method—Operationalizing Concepts gives you a chance to see for yourself how difficult this step of the scientific process can be.

**Active LEARNING**

**The Scientific Method—Operationalizing Concepts**

Continuing with our example, Levine and Waite (2000) were looking at the relationship between television viewing and children’s attentional behavior. To operationalize “television viewing,” they had the children keep diaries of their television viewing for one week and asked the children’s parents to estimate how much television their children watched. Thus, they operationalized television viewing as the amount of television watched in a week. This is a pretty straightforward way to operationalize “television viewing,” but it is more difficult to decide how to operationalize “attentional behavior.” These authors operationalized this concept by observing children in their classrooms for 15-minute periods during which they recorded behaviors such as the child being “off task, fidgeting, [making] inappropriate vocalizations, playing with an object, and being out of seat” (Levine & Waite, 2000, p. 672, adapted from Barkley, 1991).

Think about how you would operationalize each of the following concepts if you wanted to measure them as a part of your research. Be very clear in your descriptions of each type of behavior you include so someone else who is looking at the same behaviors could place the behaviors into the same categories. For example, if you wanted to observe aggression, you might first choose a specific aspect of aggression, such as physical aggression. You would then need to define this in behavioral terms. You might define physical aggression as any action that causes physical pain to another child. Then you would describe the specific behaviors you would record during your observation, such as hitting, biting, and pinching.

Following these steps, operationalize the following concepts, as you might observe them in adolescents:

- Self-esteem
- Stress
- Affection

Compare your list with those of other students in your class and discuss the following:

- How much overlap is there in the behaviors each student identified as part of the operationalization of the concept?
- Is the description of each behavior clear and unambiguous?
- Do the categories of behavior described cover the entire range of the concept or only a portion of that range?
- How would you measure each of your indicators? For example, will you use observation, self-report of the participants, the reports of others, or physiological measurement?
Reliability and Validity

Two essential characteristics of any measure used in scientific research are reliability and validity. A measure is reliable when it produces the same or similar results each time it is used. There might be some variation from one occasion to another, but a measure would not be very useful in research if it gave you widely different values each time you used it.

We also need to be sure that our measures accurately reflect the construct or characteristic in which we are interested. A measure is considered valid if it measures what it says it is going to measure. For example, in Active Learning: The Scientific Method—Operationalizing Concepts, you might have defined self-esteem for an adolescent as a feeling of being proud of one’s self for one’s achievements. You might then have tried to measure this concept by asking questions such as “Are you proud of your accomplishments?” and “Do you ever feel bad about yourself when you are not able to do something well?” It certainly appears that these questions would be a valid measure of the concept you are trying to assess, so this is called face validity. However, there are many other ways to see whether your questionnaire is a valid measure of self-esteem, such as seeing whether it gives similar results to another valid test of the same concept.

It is important to note how reliability and validity are related to each other. Just because a test is reliable does not necessarily mean that it is also valid. Suppose we told you that we had developed a new way to measure intelligence based upon your shoe size. This likely would be a very reliable measure because each time we measured your feet, we would get the same value. The size of your feet doesn’t change a lot from day to day. However, do you think this would be a valid way to measure intelligence? Certainly not, because there isn’t any reason to think that foot size is related in any meaningful way to your cognitive abilities. Measures that are used in scientific research need to be both reliable and valid.

SAMPLING AND REPRESENTATIVE SAMPLES

We want our research to do more than say something about the particular children or adolescents who take part in our research. We want to be able to generalize our results to larger populations, but this step must be done carefully. For example, if we conducted research at kindergartens in suburban public schools, we should not mistakenly assume that our findings apply equally to children from social, economic, or ethnic backgrounds that are very different from those of the children we studied. We might not find the same results if our research involved children from low-income families enrolled in a Head Start program or children from higher-income families enrolled in an expensive private school.

To understand the process of sampling, we need to distinguish between a population and a sample. A population includes everyone in the category we are interested in studying or learning more about. All toddlers, all elementary school children with dyslexia, or all adolescent females—each represents a population. But thinking about these groups, it is clear that including an entire population in any research study is impossible, so we need to select a sample from that population. Because we want to generalize the findings from a particular study to the population, a sample should reflect the characteristics of the population of interest and this is called a representative sample. There are at least two ways to do this. First, you can select your participants randomly from the population you want to study. If the selection is truly random, by chance the group of people you select should have the characteristics of the group as a whole. Second, if you know the characteristics of the population, you can choose your participants in a way that matches that set of characteristics. If you want to generalize your findings to all toddlers, for instance, you need to include boys and girls in the age range you are interested in and include demographic characteristics (for example, socioeconomic characteristics, ethnic and racial diversity, types of family structure) in proportions that are similar to the general population.

If our sample does not mirror the characteristics of the population (for example, if it primarily includes Anglo children from middle-class two-parent families), we would need to be careful to generalize our results only to children with the same characteristics as the sample.
When you read research that is published in professional journals, you'll see how careful researchers are to specify how widely their conclusions can be applied, and they often call for additional research that will extend their work to a broader cross section of participants. Representative sampling is used in all types of research, whether the researchers conduct surveys, questionnaires, observations, or experiments.

In Chapter 1, we warned against generalizing from your own experiences to the experience of other people. You cannot rely on your personal experience or an isolated example from someone else’s experience in place of scientific evidence based on research using representative sampling. If a friend were to tell you that his grandfather had begun smoking at age 16 and lived to be 97 years old, would you take that as evidence that smoking is not harmful to your health? In light of the overwhelming evidence of the harmful effects of smoking on health, this type of anecdotal evidence should not carry much weight in your decision making.

An example of what can happen when population sampling is not done correctly comes from research on air bags in automobiles. Research clearly shows that air bags save lives, but it also shows that they are not equally effective for all drivers and passengers. Although air bags can help prevent crash fatalities, they also can cause a range of injuries that include corneal abrasions, aortic ruptures, abdominal injuries, and fractures of the forearms (Segui-Gomez, 2000). The problem comes from the fact that the force of the deployment of the airbag was initially determined using crash test dummies that were the size of the average American male. Generalizing from findings of what made air bags most effective for the average man to the population as a whole resulted in air bag systems that actually placed women and children at greater risk for nonfatal injuries. Fortunately, research is now being conducted with female-size and child-size crash dummies to collect new data that will help manufacturers develop air bags that are more effective for everyone (Shaver, 2012).

**Active Learning: The Scientific Method—Sampling** allows you to look closely at the sample used by Levine and Waite (2000), think about the characteristics of the sample used in this research, and decide to which groups of children the results of this study can be generalized.

---

*One size does not fit all.* Automotive engineers originally designed air bags to protect the “average” American man. Today they realize that passengers come in all shapes and sizes and what works for an average-size man doesn’t work well for others.
The Scientific Method—Sampling

Returning again to the research on television and attention, we can look at how Levine and Waite (2000) selected the sample for their research. Read the following description of the children who took part in the study:

Seventy fourth and fifth grade students participated in the study. There were 33 girls and 37 boys. Children’s ages ranged from 8 years 6 months to 11 years 9 months. . . . The children came from four public schools: two schools were urban, one was suburban, and one was rural. Fifty children were White, non-Hispanic; three were Hispanic, two were African American, one was Asian, one was identified as “other” and not specified. Thirteen parents chose not to report the child’s ethnicity. The sample was primarily middle and working class. (p. 670)

From this description of the sample, describe the population to which the results of this study might apply. To which populations should we not generalize the results from this sample?

Methods and Measures

Once researchers have developed their hypothesis, operationalized the concepts within it, and chosen their sample, they must decide how they will gather the data for their study. We will describe some of the ways commonly used to study children and adolescents, including observations, self-report measures, standardized tests, physiological measures, archival records, case studies, and ethnography. Each method has both advantages and disadvantages (see Table 3.1), and that is one of the reasons why we can’t say that any one method is better than the others. Researchers choose the method that is most likely to answer the questions that guide the research.

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>Rich source of information</td>
<td>Can be confused with interpretation</td>
</tr>
<tr>
<td></td>
<td>Can observe behavior as it naturally occurs</td>
<td>Potential observer bias</td>
</tr>
<tr>
<td></td>
<td>Can be conducted in a laboratory to gain</td>
<td>Can produce large amounts of raw data that</td>
</tr>
<tr>
<td></td>
<td>control in the situation</td>
<td>must be coded and analyzed</td>
</tr>
<tr>
<td></td>
<td>Can lead to new hypotheses</td>
<td>The presence of an observer may change the behavior being</td>
</tr>
<tr>
<td></td>
<td></td>
<td>observed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cannot identify the causes of behavior</td>
</tr>
<tr>
<td>Method</td>
<td>Advantages</td>
<td>Disadvantages</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Surveys, Questionnaires</td>
<td>Gathers information quickly and efficiently</td>
<td>Must precisely word questions</td>
</tr>
<tr>
<td></td>
<td>Can be used to gather information on many different topics</td>
<td>Can include misleading or biased questions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Respondent may not answer honestly</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Respondent may not be able to accurately recall or report on behavior</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Respondent may provide a socially desirable answer rather than a truthful one</td>
</tr>
<tr>
<td>Interviews (structured and clinical)</td>
<td>Can be a first-person or a third-person account</td>
<td>No second observer to verify the information</td>
</tr>
<tr>
<td>Standardized Tests</td>
<td>Can assess many qualities or characteristics</td>
<td>Must periodically update norms</td>
</tr>
<tr>
<td></td>
<td>Allow an individual to be compared to the average performance of a group</td>
<td>Performance tests must be scored and interpreted by trained examiner</td>
</tr>
<tr>
<td></td>
<td></td>
<td>May be biased against certain groups</td>
</tr>
<tr>
<td>Physiological Measures</td>
<td>Can gather data that don’t require language or an active response from participants</td>
<td>Requires expensive equipment that can be difficult to maintain</td>
</tr>
<tr>
<td></td>
<td>Responses are difficult to fake</td>
<td>Interpretation of data is not always clear</td>
</tr>
<tr>
<td>Archival Data</td>
<td>Large amounts of data may be available</td>
<td>No control over the variables collected or the sample characteristics</td>
</tr>
<tr>
<td>Case Studies</td>
<td>Source of rich information and hypotheses for future research</td>
<td>Cannot generalize information to a larger population</td>
</tr>
<tr>
<td></td>
<td>Can utilize multiple methods</td>
<td>Time intensive</td>
</tr>
<tr>
<td></td>
<td>Can investigate situations it would be unethical to intentionally create</td>
<td>Possible observer bias</td>
</tr>
<tr>
<td>Ethnography</td>
<td>Can provide a rich, detailed look at cultural groups</td>
<td>May change the behavior of group members</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time intensive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Risk of observer bias</td>
</tr>
</tbody>
</table>

**Observations**

We can learn a great deal about anything we are interested in—including development—by making careful observations. Scientific observations differ from our casual, everyday observations of the world because they need to be both systematic and objective and must be carefully planned and executed if they are going to be valid.

To understand more about the difference between objective observation and subjective interpretation, try **Active Learning: Observation or Interpretation?**

---

**Active Learning: Observation or Interpretation?**

A narrative description is a complete, step-by-step description of a child’s behavior. It is important to learn to separate this objective description from interpretations that unintentionally read a meaning or cause into the behavior. To better understand this distinction, observe a child for about 15 to 30 minutes (or you can observe anyone if a child is not readily available). Divide the pages you use into two columns. In the left column, write down what you see the child do. For example:

(Continued)
(Continued)

**Observation:** Annie stands in the doorway and looks into the room. She buries her head in her father’s leg and holds on.

Later, in the right column, give your interpretation of the behaviors. For example:

**Interpretation:** Annie seemed shy about entering the room and turned to her father for comfort.

Your interpretations should be about the child’s behavior and not a summary statement about what the child “is like,” because the child may appear to be different at another time. For example, don’t write that Annie is a shy child. Maybe she doesn’t feel well that day but usually bounces into the room with no fear. It takes many observations to make any kind of general statement about a child.

Try the following exercise to practice. Write down what you see in this set of photos under Observation and then write down how you would interpret what you see in the column for Interpretation. In each case you will describe what the child is actually doing and what their face looks like in the Observation section. Only in the Interpretation section will you write what you think is the meaning of what you see. For example, is the boy in the third photo looking angry, sad, or devious? Each of these would be a different interpretation of the expression on the child’s face.

<table>
<thead>
<tr>
<th>Observation</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.jpg" alt="Observation 1" /></td>
<td></td>
</tr>
<tr>
<td><img src="image2.jpg" alt="Observation 2" /></td>
<td></td>
</tr>
<tr>
<td><img src="image3.jpg" alt="Observation 3" /></td>
<td></td>
</tr>
</tbody>
</table>
If the researchers conducting the observations are testing their own hypotheses, there is a risk that they might see or pay more attention to things that tend to support those hypotheses and overlook things that don’t. This tendency is called **observer bias**. Having more than one observer code the observations helps assure us that the observations are objective rather than subjective. Another safeguard against observer bias is to use observers who don’t know the specific hypothesis that is being tested (that is, observers who are “blind” to the hypothesis), so that it cannot influence their perception of the events they are observing.

Observations can be made in a setting where the behavior of interest naturally occurs, or they can be made in a setting that offers more control over circumstances, such as a research laboratory. One advantage of doing observations in naturally occurring settings is that we get to see children behaving as they normally do within the social relationships that are a part of their everyday lives, and we see them in situations that have real emotional significance for them (Dunn, 2005). However, moving observations into a laboratory gives researchers greater control over the situation and allows them to create a specific set of conditions in which to conduct their observations. You will read about some of the early influential work in the field of child development that used observation as its methodology in *Journey of Research: Doing Observational Research*.

---

**Journey of Research: Doing Observational Research**

The scientific study of child development is relatively recent, but parents have been keeping records of their children’s growth and development for their own enjoyment for a very long time. In the 1800s, educated parents often kept what were called *baby biographies* of their children. In 1877 Charles Darwin, a renowned scientist of the day, took the notable step of publishing some observations he had made of his own son in the scientific journal *Mind*. In this article, Darwin described his observations regarding early reflexive behaviors in the infant, as well as later voluntary movement. He also noted incidents of anger, fear, pleasure, and affection in his son. The fact that a noted scientist found these observations worthy of publication in a journal gave the study of children’s development a legitimacy that it hadn’t had before.

Years later, one of the most influential developmental theorists, Jean Piaget, made extensive, detailed observations of his three children, and these observations formed the basis of a number of his writings on cognitive development (see, for example, Piaget, 1952/1963). Although Piaget did not conduct his observations using the guidelines we have described (for instance, there was only one observer, he was testing his own hypotheses, and was observing children in whom he was emotionally involved), his observations helped him to develop his theory that has subsequently been tested by other researchers with more modern methods.

At about this same time, Roger G. Barker and Herbert F. Wright were studying 119 children living in a Midwestern town using a methodology they called *ecological psychology*. In one of the publications resulting from their work, it took the authors 435 pages to record and interpret a single 14-hour observation of one 7-year-old boy (Barker & Wright, 1951). It took a team of eight observers, each making 1-minute observations, to record this information—a very labor-intensive process, to say the least. Today video cameras have largely replaced paper-and-pencil recordings, and are used by both child development researchers and proud parents to record children’s behavior. However, the goal of observational research remains the same: to capture the full, rich range of behavior as it naturally occurs.

---

**Observer bias** The tendency for an observer to notice and report events that the observer is expecting to see.

---

When conducting research by doing an observation, it is important that the person who is doing the observation does not know the purpose of the research. **True**

Copyright ©2018 by SAGE Publications, Inc. This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
Making a detailed record of everything that happens in a stream of behavior can make researchers aware of aspects of behavior that they have not noticed before and can be a good source of new hypotheses for future research. Both of these are advantages of doing observational studies. However, recording everything that happens even in a fairly short period of time produces a tremendous amount of raw data that need to be analyzed and reduced before useful information emerges.

For this reason, researchers often structure their observations in ways that allow them to be more focused. Checklists can be used to assess many aspects of children’s development, including social skills, physical skills, language development, and problem behaviors. When observing the child, the researcher simply marks the presence or absence of each item on the checklist. For example, a checklist of gross motor skills might include the behaviors walking, jumping, and hopping. The observer would ask the child to do each of these and check it off if he can accomplish it or the observer can simply observe the child’s behavior and record what she sees. A checklist provides a quick way to look at a child’s development in relation to what other children of the same age can do. Also, the observer may use the checklist to keep track of children’s progress as they grow and develop. As a result, checklists are often used in educational, medical, and other settings that require quick, efficient assessment of a child’s level of functioning.

Although observations are very useful sources of information about behavior, there are some limitations when using this method. First, the goal is to capture behavior as it naturally occurs, but the mere presence of an observer might change the way that people behave. Fortunately, children usually adapt to the presence of an observer without too much difficulty. Although they are initially curious and might ask questions about what the observer is doing, the lure of getting back to what they were doing (such as playing with their friends) is usually far stronger. Another limitation of observational research is that it does not tell us directly about the causes of behavior because the observer does not have enough control over the situation to make this determination. For example, if you observe a child who stays on the sidelines while other children are playing and refuses to interact with the other children, there are many possible explanations for this behavior. From the observation alone, it is impossible to tell whether this is the behavior of a child who simply is not very social, a temporary reaction to something that occurred earlier in the day, or an indication of an adjustment problem for the child. Based on these observations, the researcher might formulate a hypothesis to explain the child’s behavior, but additional research would need to be conducted to determine the cause of the observed behavior.

**Self-Report Measures**

Another way to gather information relatively quickly and efficiently is to use self-report measures such as surveys, questionnaires, and interviews. However, the usefulness of the information gathered from self-report measures largely depends on the accuracy of the answers they produce. Having questions that are precise, well-written, and understandable is essential to the validity of this research. You may have had the experience of trying to complete a survey in which the questions were unclear or difficult to answer. In that case, it wouldn’t matter how much you wanted to give accurate information; you wouldn’t be able to do so. Of course, another possibility is that the person taking the survey is unwilling or unable to give complete or accurate responses. And sometimes people give the answer they think the researcher is looking for or give answers that they think make them look good in the eyes of the researcher (a problem that is called social desirability). For example, students being surveyed about their drug and alcohol use may downplay the amount they use so they will not look deviant to the researcher. You can see how social desirability becomes a challenge for researchers who are investigating a sensitive topic, such as sexuality, drugs, or prejudice.

Typically, each survey participant responds to the same questions, presented in the same order. These questions might be open-ended (for example, “What do you do when you are spending time with your friends?” or “What is your favorite subject in school?”) or may use...
a forced-choice format (for example, “How many hours a week do you watch television? (a) less than 1 hour, (b) between 1 and 3 hours, (c) between 3 and 6 hours, or (d) 6 hours or more”). To better understand how poorly worded or otherwise misleading questions can keep a participant from providing complete and truthful answers, read Journey of Research: Children’s Eyewitness Testimony.

Children’s Eyewitness Testimony

An important illustration of how the way a question is asked can affect the response that you get comes from research on children’s eyewitness testimony. Think for a minute about the subtle difference between asking someone “Did you see that?” and asking her “Didn’t you see that?” The first alternative suggests that there can be one of two legitimate answers (“Yes, I saw that” or “No, I didn’t see that”), but the second alternative implies that you may have missed something that someone else saw. The pressure is to respond to the second question by saying “Of course I saw that.” Although you may feel that you would respond to such a question by simply saying what you did or didn’t see, regardless of how the question was phrased, a child is more likely to be swayed by the question itself.

Prior to the 1990s, there had been relatively little research on children’s ability to accurately recall events so that they could serve as eyewitnesses, but in the 1990s there were some high-profile cases of alleged child abuse that placed children in the witness seat. Under relentless and often suggestive questioning, the children described horrible abuse they had allegedly suffered at the hands of adults who were caring for them. Based on this testimony, several of the defendants received jail sentences. However, these charges were later dismissed or the plaintiffs were released from prison because of the improper way that evidence had been gathered. Here is an example of the type of improper questioning that was used in the notorious McMartin Preschool case in which seven teachers were accused of sexually abusing several hundred young children:

Interviewer: Can you remember the naked pictures?
Child: (Shakes head “no”)
Interviewer: Can’t remember that part?
Child: (Shakes head “no”)
Interviewer: Why don’t you think about that for a while, okay? Your memory might come back to you.

(Interview Number 111, p. 29, as cited in Garven, Wood, Malpass, & Shaw, 1998)

We now know that even young children are able to accurately recall events and can give reliable eyewitness testimony (Odegard & Toglia, 2013). However, when the questions they are asked are misleading, they are subjected to repeated questioning, or the interviewer makes overt suggestions about what has happened, we cannot trust children’s answers (Ceci & Bruck, 1995; Krähenbühl & Blades, 2006). It is clear from the way the interviewer in the McMartin case kept repeating the question and refused to take the child’s denial that there had been naked pictures that this interviewer had a particular answer in mind and wanted the child to give that answer. Suggesting that a memory “might come back” implies that the event is something that happened but has been forgotten, rather than allowing the possibility that it never happened at all. Because children are limited in their ability to understand and interpret language, we need to be particularly careful about the wording of questions when designing surveys, questionnaires, and interviews for them.

Children’s memories are good enough to allow them to give reliable eyewitness testimony. True
Usually an interviewer asks everyone who is interviewed the same set of questions, but sometimes the interviewer might want to ask additional follow-up questions or ask the respondent to expand on her original answers or provide examples. In this case, a clinical interview allows the researcher greater flexibility.

In addition to his naturalistic observations of children which we described earlier, Piaget used clinical interviews to refine his theoretical ideas. In the following example, Piaget used a clinical interview to learn about children's sense of morality. In one pair of stories, he asked the children he was interviewing to compare the actions of two girls. The first girl, Marie, wants to surprise her mother by sewing her a nice present. She doesn't really know how to use the scissors and ends up cutting a big hole in her own dress. The second girl, Margaret, takes her mother's scissors while her mother is out. She doesn't really know how to use them and makes a little hole in her own dress. Piaget would then ask a series of questions to determine which girl the children considered to be naughtier:

P: Which one is the naughtiest?
C: The one who made the big hole.
P: Why did she make this hole?
C: She wanted to give her mother a surprise.
P: That's right. Then which of the little girls was nicest?
C: (hesitation)
P: Say what you think.
C: The one who made the little hole is the nicest.
P: If you were the mother . . . which would you have punished most?
C: The one who made a big hole.

(Piaget, 1965, p. 127)

Although this interview has some standard questions like “Which one is the naughtiest?”, many of the other questions are unique, based on a particular child’s responses.

A variation on the self-report approach is to rely on a second party, like a parent, teacher, or care provider, to give the information on the target child. This is a particularly important way to collect data on infants and children who are too young to respond to an interviewer’s questions. The more time the people being interviewed spend with the child and the more familiar they are with the child’s behavior, the more likely it is that they will be able to provide high-quality information. For example, parents have been asked to describe the antisocial and acting-out behavior of their adopted adolescents (Klahr, Rueter, McGue, Iacono, & Burt, 2011), and teachers have been asked to rate the social-emotional competence of their students (Merrell, Cohn, & Tom, 2011).

Standardized Tests
Tests can provide information on a wide range of topics relevant to understanding development. You are probably familiar with standardized tests such as IQ tests and achievement tests. In Chapter 8, we have a great deal of...
information about the use of standardized tests in educational settings. Researchers standardize a test by administering it to large groups of children to establish norms for the test. This provides an expected range of scores, and individuals who later take the same test will have their results compared to this norming group. By comparing an individual child’s performance to the appropriate age norms, we can determine whether that child is performing at the same level as the average child of the same age or is performing above or below average. Many of these tests are paper-and-pencil tests that can be administered to groups of children all at the same time, but some of them are called performance tests because they require the child to do something (for example, assemble a puzzle, build a tower of blocks). Someone who has been trained to administer and interpret the test, such as a school psychologist, works one-on-one with the child to administer a performance test.

Standardized testing has been a controversial topic for many years. Much of this controversy has centered on the validity of the tests, that is, whether they are actually measuring what they say they are measuring. A related controversy has surrounded how standardized test results are interpreted and used. Early wide-scale use of intelligence tests during World War I and World War II found that native-born Americans scored better on the tests than immigrants, immigrants from Northern and Western Europe scored better than ones from Southern and Eastern Europe, and Black Americans received the lowest scores of all (Glaser, 1993). The controversy centered on whether the differences in test scores reflected inherent differences in mental abilities between these groups, or whether the tests were biased in some way that put certain groups at more of a disadvantage than other groups. This became known as the cultural test bias hypothesis (Reynolds, 2000).

College entrance tests such as the Scholastic Assessment Test (SAT), formerly known as the Scholastic Aptitude Test, or the American College Testing (ACT) program have received the same types of criticisms. Critics have claimed that they are biased in a way that discriminates against certain groups of students. On average, men score higher than women on these tests, Asians/Asian Americans/Pacific Islanders and Whites score higher than Mexican Americans/Latinos or African Americans, and students from families with higher incomes score higher than students from families with lower incomes (College Board, 2008; FairTest, 2010).

Careful selection of norming samples that include all the segments of the population who will use the test helps to reduce testing bias. It is also important that the test developers be culturally sensitive so they do not use language or describe specific situations in the test that will be familiar to some test takers but not to others. In recent years a number of psychometric reviews have come to the conclusion that “well-constructed, reliable, well-standardized psychological tests are not biased against native-born American racial or ethnic minorities” (Reynolds, 2000, p. 145; see also Balkin, Heard, Lee, & Wines, 2014) and that any content bias is usually quite small. However, in response to concerns about possible bias in SAT/ACT results, over 850 colleges have stopped using them in admission decisions or have made them optional (FairTest, 2016; Hiss & Franks, 2014). Because educational decisions that are made based on standardized test scores have such important consequences for individuals, organizations such as the American Psychological Association and the American Educational Research Association recommend that decisions never depend upon a single test score.

In addition to using standardized tests to determine whether children are “on track” regarding their development, tests are also useful for assessing the effectiveness of programs and interventions. For example, the effectiveness of a summer enrichment program that paired talented young adolescents from limited-opportunity backgrounds with high school students or college-aged students was assessed using standardized tests of mathematics and reading skills (Laird & Feldman, 2004). In another program evaluation, a standardized test of math ability was used to assess the effectiveness of a 6-week classroom intervention conducted in Head Start classrooms designed to promote emergent math skills in preschool children (Arnold, Fisher, & Doctoroff, 2002).
Physiological Measures

It is not always easy to gather the types of data we need to answer important questions about development, but today a number of physiological measures are helping us get those answers. Electromyograms (EMGs) measure electrical activity in the brain (Dupler, 2014). Functional magnetic resonance imaging (fMRI) measures blood flow in the brain to show which parts of the brain are active (UC San Diego Center for Functional MRI, 2016). The photos on the facing page show the information we get from these measures. Event-related potentials (ERPs) measure the brain’s electrical response to meaningful sensory stimuli. Other physiological responses that have been measured include the activity of sweat glands to measure arousal, the dilation of the eye to indicate heightened interest, an increased heart rate to reflect information processing demands, and changes in breathing patterns to indicate anxiety or stress (Sowden & Barrett, 2006).

These measures have been in the forefront of research on information processing, as well as in studies of brain dysfunction in conditions such as ADHD and autism. They also provide important information on facets of human functioning related to the central nervous system, emotions, stress, cognitive processes, personality, and intelligence (Sowden & Barrett, 2006). Because physiological measures enable us to measure what is going on inside an individual without needing to rely on verbal communication, they have been particularly important in research involving infants and young children who can’t answer questions because they either are not speaking at all yet, or their language ability is so limited that they can’t understand and follow complex instructions or provide complex verbal responses to questions.

One advantage of using physiological measures in research is that the responses are not under the conscious control of the individual, so they cannot easily be faked or presented in a way intended to make them socially desirable (Blascovich, 2004). Most of these measures are noninvasive, meaning that nothing needs to be placed inside the body to collect the data. For instance, the photo on this page shows a type of cap with electrodes embedded in it that is used to record brain activity in an infant. A cap is placed snugly on the child’s head, but is not at all painful for the child. A weak electrical current then passes from the cap through the bones of the skull to record brain activity.

As an example of how physiological measures have been used, researchers looked at how 12-month-old infants reacted to stimuli in the environment when their caregivers reacted to an event in a positive, negative, or neutral way (de Haan, 2007). By measuring event-related potentials, they found that the infants paid more attention to the stimuli when their caregivers reacted in a negative way than when the caregivers reacted in a neutral or positive way. In an evolutionary sense, it makes sense that infants would pay more attention to things that their caregivers found unpleasant or distressing.

Research using fMRIs to scan the brains of children diagnosed with autism spectrum disorder has found that the part of the brain that controls face recognition is underactive in these children (Kalb, 2005). As you may know, children with autism often do not make eye contact with others and show little or no interest in social relationships. By using physiological measures, researchers have been able to understand more about the processes that underlie certain behaviors. Building on these findings, they are now developing interventions that can teach face recognition skills to autistic children. In one such program, children use computerized games to choose a face appearing on the screen that matches a target face or to connect three different views of the same face (Tanaka 2006).
et al., 2010). After 20 hours of training, the participants were better able to recognize mouth features and eye features in faces.

**Archival Records**

Researchers don’t always collect their own data. They may use *archival records* (or secondary data) that others have collected, sometimes for a different reason. For instance, a researcher might use historical diaries, letters, or photographs to gain insight into what childhood was like in the past. Reports and statistics collected by the U.S. government provide a historical snapshot of many topics relevant for a developmental researcher. In this book, you will read about research on child abuse, adolescent pregnancy, and many health conditions affecting children and adolescents that are based on government reports. Medical records and school records are other sources of archival data that are relevant to understanding a child’s development.

One of the most ambitious sources of archival data related to child development is the Panel Study of Income Dynamics (PSID). Data collection from a nationally representative sample of over 18,000 individuals began in 1968 and has continued on a yearly basis since then (PSID, 2015a). Through in-person interviews with parents and children younger than age 18 and telephone interviews with those over the age of 18, information was gathered on a number of family, school, and neighborhood characteristics that are linked to the physical health, emotional well-being, cognitive development, and social relationships of the children (PSID, 2015b). Because data collection of this scope requires a great deal of time, money, and effort, making archival data available to many researchers is more cost-effective than having each researcher collect data individually. More than 3,000 peer-reviewed publications are based on the data in this archive (PSID, 2015a). However, a limitation of using archival data is that the researcher has no control over the variables that are available for analysis, nor any choice in the characteristics of the sample, although in large and diverse data sets, researchers can often create a sample that is relevant to their research interests.

**Case Studies**

A *case study* takes a comprehensive and intensive look at a single individual or a small group of individuals. This intense focus allows the researcher to look at a topic in much greater depth than would be possible in a study involving a large number of participants. As part of the case study, the researcher can use a variety of the methods we have already discussed, including observations, clinical interviews, psychological tests and assessments, or archival data. Case studies have been used to investigate widely different research questions, such as the psychological challenges faced by a child with a severe mental disorder called psychosis (Green, Fazio-Griffith, & Parson, 2015), the problems encountered by a blended family...
Part I: Understanding Development: Why and How We Study Children and Adolescents

(Zeleznikow & Zeleznikow, 2015), and the coping mechanisms of an adolescent with HIV-positive parents (Lowe, 2007).

Individuals who are subjects of case studies are often exceptional in some way, which is what makes them interesting subjects for this type of examination. Their life experiences often include situations that we would never intentionally create for a child just so we could study them. For example, one well-known case study in child development involved Genie, a young girl who had been raised in conditions of horrible deprivation. After Genie was removed from her abusive family, a group of researchers worked to rehabilitate her and in the process kept detailed case notes on her attempts at recovery. You will learn more about the story of Genie in Chapter 9 when we discuss language development. Sigmund Freud also published several papers based on case studies of his patients, including a young boy with a pathological fear of horses and a young man suffering from an obsessive fantasy (Kahn, 2002).

Earlier in this chapter, we talked about the need to guard against observer bias when doing observational research. Because of the close relationship that can develop between the researcher and the subject of a case study, it is particularly important that the researcher strive to remain objective in his observations and interpretation of data. Despite this challenge, case studies offer us an in-depth picture of development because they can bring together information from multiple sources using multiple methods. Although the findings from case studies do not generalize to a wide segment of the population, they can be a rich source of new hypotheses that can be explored by future research with other more typical or representative groups of individuals.

Ethnography

Ethnography is a research technique adapted from the field of anthropology in which a researcher lives with a group of people as a participant observer, taking part in the group’s everyday life while also observing and interviewing the people in the group. Ethnography is a type of qualitative research that begins with observations and then uses those observations to generate hypotheses to help explain the behavior observed (Morgan, 2014). It does not manipulate variables, nor is it designed to test hypotheses. The primary goal of this approach is to use observations to understand how people make sense of their world and the experiences in it (Merriam, 2009). This technique is especially helpful when studying children in different cultures as it allows the researcher to see the whole context of the children’s development. For example, Samantha Punch (2012) lived in an area of Bolivia in order to study children’s development in a poor, rural community, and when Sudhir Venkatesh (2008) wanted to administer questionnaires to gang members in inner-city Chicago, he ended up living among them for almost 10 years. This resulted in a rich account of their lives in a book titled Gang Leader for a Day: A Rogue Sociologist Takes to the Streets.

Although ethnographies give us a rich picture of a cultural group by someone who is embedded in that group for an extended period of time, researchers must consider the fact that their presence may change the behavior of the individuals in the group that they are observing. This can become a threat to the validity of the conclusions drawn from the study. Another threat to the validity of the study is that the researchers’ own biases may affect how they interpret what they see, so they must always guard against observer bias. A researcher’s personal safety could be placed in jeopardy in some research settings, and ethnographic researchers must be willing to commit to the considerable amount of time needed to carry out this type of study.

Active Learning: The Scientific Method—Measures will help you review some of the types of measures we have just discussed by asking you to recognize examples.
The Scientific Method—Measures

Now let’s examine the types of measures used in the study on television viewing and attention problems by Levine and Waite (2000). In the abstract of the article, the authors state that “70 fourth and fifth grade students recorded a ‘television diary’ for one week and reported their preferred television shows. Parents estimated their child’s television viewing time and reported their child’s preferred shows. The assessment of attentional difficulties included teacher ratings [the ADD-H Comprehensive Teachers Rating Scale], parent ratings [the Distractibility/Hyperactivity scale (DI) of the Parenting Stress Index], standardized tests [the Stroop color and word test], and classroom observations” (Levine & Waite, 2000, p. 667).

For each type of measure listed below, put an X in front of the measure if it was used in this research (leave blank if it wasn’t used). If a method was used, describe what type of data was collected using this measure.

- observations: classroom observations of attention difficulties
- self-report measures: child’s television diary and parent’s report of the child’s preferred shows
- standardized tests: the Stroop color and word test, the ADD-H Comprehensive Teachers Rating Scale, and the Distractibility/Hyperactivity scale (DI) of the Parenting Stress Index
- physiological measures
- archival records
- case studies
- ethnographic studies
- other types of measures

REPLICATION OF RESULTS

Our confidence that our conclusions are valid is strengthened if results are replicated or repeated in subsequent research. We can do this by repeating the research using other groups that are the same or similar to the group we originally studied, or by extending the research to new groups to determine whether the conclusions can be generalized to new situations. We also expect that other researchers will be able to replicate our results by conducting their own independent research and coming to the same or similar conclusions (Makel, Plucker, & Hegarty, 2012).
Replication of results is at the core of the scientific method. As we discussed in Chapter 1, if results cannot be reproduced by others the finding is just a fluke. Yet most research is published because it shows something new, so replications are often not accepted for publication. In the past several years, there has been a movement to change this. The Association for Psychological Science (n.d.) has provided an outlet for publications that is dedicated to replications in order to “emphasize findings that are robust, replicable and generalizable” (para. 2).

CHECK YOUR UNDERSTANDING

1. Explain what it means to say that a measure is reliable and valid.
2. Explain the importance of having a representative sample in research.
3. What are the advantages and disadvantages of using observations, standardized tests, and case studies to collect data?

HOW RESEARCH IS DESIGNED

3.2 What are the different types of research designs used to study development?

We have talked about a variety of methods used to gather information about development. You can think of research methods and measures as the tools we use to collect the data we are interested in. In this next section, we describe research designs, which are the roadmap or blueprint we use to structure our research. The design we choose tells us what will be done and when and how we will do it. You will learn the differences between experimental and correlational research designs and the strengths and limitations of each. We also describe the types of developmental research designs used in research with children and adolescents.

EXPERIMENTAL RESEARCH DESIGNS: IDENTIFYING THE CAUSES OF BEHAVIOR

Experimental research design occupies a central place in the study of child development because it allows us to identify the causes of behavior. With research that is not based on an experimental design, we can speculate about the causes, but we do not have enough control over the situation to make a firm determination. However, when a researcher designs an experiment, the goal is to control as many aspects of the experimental situation as possible in order to draw conclusions about the causes of the outcome.

Experiments can take different forms and can include one, two, or more groups of participants, but these are the essential features you will find in an experimental research design:

- The experimental group is the group that receives the special treatment of interest to the researcher.
- The control group does not receive the special treatment and provides a baseline against which the experimental group can be compared.
- The participants are randomly assigned to either the experimental or the control group. Because this assignment is made by chance, the two groups will likely start out being very similar to each other, without any systematic differences that could affect the outcome of the experiment. To get a random assignment to groups, you could simply flip a coin for each participant, with all “heads” going into one group and all “tails” into the other; or you could put all the names in a hat and pull them out one by one, alternately assigning them to one group or the other.
• The **independent variable** is the special treatment the researcher hypothesizes will make a difference between the experimental and control groups at the end of the experiment. The assumption is that the independent variable is the cause of any change we observe following the experiment.

• The **dependent variable** is the outcome of interest to the researcher. We measure it at the end of the experiment to see whether manipulating the independent variable has produced the expected effect.

If we look at an example of experimental research, this terminology will have more meaning for you. A recent study looked at the impact of attending Head Start on cognitive development and several social-behavioral outcomes for preschool children, including social skills, positive approaches to learning, and aggressive behavior (Zhai, Brooks-Gunn, & Waldfogel, 2014). All children were eligible to enroll in Head Start but there were not enough spaces for all of them. The children who were randomly chosen to participate in the Head Start program became the experimental group and the other children who could not be enrolled became the control group. Remember, the experimental group is the group receiving the special treatment that the researchers are interested in studying, and the control group does not receive the treatment. What is different between the two groups is the independent variable (in this case, whether the child attended Head Start or not). At the beginning of the experiment, the researchers made sure that the children in the two groups were matched on a number of characteristics to be sure that the groups were as similar as possible. This gives us confidence that any difference found at the end of the experiment is due to what the researchers did, rather than to an initial difference between the groups. The dependent variable is the outcome of interest to the researchers and what is measured and compared between the groups at the end of the experiment. In this study, the outcome measures were the children's cognitive and social-behavioral development.

By the end of first grade, the researchers found that the children who attended Head Start had better cognitive scores and parent-reported behavioral development than the children who did not attend Head Start. Because the two groups of children were similar at the start of the experiment, and because the only relevant difference between the groups during the intervention was whether the children attended Head Start or not, the researchers could conclude that participating in Head Start was the *cause* of the difference they observed at the end of the experiment.

Table 3.2 shows the steps in this process and can help you understand how they relate to each other.

### Table 3.2

**The experimental process.** This table shows how an experiment (in this case an experiment to study the effect of attending Head Start) is conducted, starting with a sample of the population of interest to the researcher and ending with results that can be interpreted.

<table>
<thead>
<tr>
<th>Step 1: Select a Sample</th>
<th>Step 2: Random Assignment</th>
<th>Step 3: Pretest</th>
<th>Step 4: Treatment</th>
<th>Step 5: Posttest</th>
<th>Step 6: Compare Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>A representative group of children is chosen for the study.</td>
<td>Children are randomly assigned to groups.</td>
<td>Pretest establishes groups are the same.</td>
<td>Independent variable is administered to the experimental group.</td>
<td>Dependent variable is measured for both groups.</td>
<td>Researcher determines whether results support hypothesis.</td>
</tr>
<tr>
<td><strong>Sample</strong></td>
<td><strong>Experimental Group</strong></td>
<td>Cognitive and social-behavioral development test scores</td>
<td>Attend Head Start</td>
<td>Cognitive and social-behavioral development test scores</td>
<td>Experimental group scores &gt; control group scores</td>
</tr>
<tr>
<td></td>
<td><strong>Control Group</strong></td>
<td>Cognitive and social-behavioral development test scores</td>
<td>Do not attend Head Start</td>
<td>Cognitive and social-behavioral development test scores</td>
<td></td>
</tr>
</tbody>
</table>

**Independent variable** The variable in an experiment that the researcher manipulates.

**Dependent variable** The outcome of interest to the researcher that is measured at the end of an experiment.
Active Learning: Experimental Research Design provides an opportunity to review the terminology used in experiments and to check that you can recognize each element when you see it in the description of an experiment.

**Active Learning**

**Experimental Research Design**

You can test your understanding of experimental research design by identifying the components of an experiment in this example taken from a study by Beth Hennessey (2007) that was designed to build social competence in a sample of school-age children.

A total of 154 fourth graders in eight classrooms participated in the study. Students completed the Social Skills Rating System in the fall and again in the spring and the teachers rated their students’ social competence at both times. During the year, half of the teachers in each school implemented the Open Circle Program, a social skills training program that “encourages students, teachers and administrators to learn and practice communication, self-control and social problem-solving skills” (Hennessey, 2007, p. 349), while the other half did not use the Open Circle Program. Based on the teachers’ reports, Hennessey concluded that the students who were in the classrooms that used the Open Circle Program training showed greater improvement in their social skills and problem-solving behavior than students who did not receive this training.

From the description of this experiment, identify the following:

- **Experimental group**
- **Control group**
- **Independent variable**
- **Dependent variable**

**Answers:**

- Experimental group: The group that received the Open Circle Program training
- Control group: The group that did not receive the Open Circle Program training
- Independent variable: Whether the group received the social skills training or not
- Dependent variable: The measure of social competence and problem solving

Because we can control many aspects of a situation that might affect the outcome of an experiment, we presume that we understand the cause of any changes that we find in the results. However, it is still possible that some other variable or condition that we haven’t taken into account is responsible for the outcome. For this reason, it is essential that experiments be carefully planned and carefully executed. It also may have occurred to you by now that, as appealing as using the experimental method might be to researchers, it cannot be used to answer many of the questions that are of great interest to us as developmentalists. There are many situations that we could never ethically create as an experiment. For example, if we want to study the effect of peers on children’s social development, we couldn’t intentionally keep some children from having friends just so we could see how their development differs from children who do have friends.
Natural or “Quasi” Experiments

Sometimes a situation occurs without a researcher creating it, and we can use that situation as a natural or “quasi” experiment to test a hypothesis. In West Africa, administrators in private schools can decide which discipline techniques they will use in their schools. Some choose to use corporal punishment, such as slapping children on the head or pinching them when they misbehave. Other administrators choose to use nonphysical child management techniques (Talwar & Lee, 2011). Researchers used this naturally occurring situation to examine the effect of these two different school environments on children’s willingness to lie about their misbehavior.

Children from both types of schools listened to a researcher play with a toy while they had their back turned to her. The researcher then said she needed to leave the room for a while and told the child not to turn around to look at the toy while she was gone. When she returned, she asked the child if he or she had looked at the toy in her absence. Although the majority of the children from both types of schools peeked, many more of the children from the schools that used physical punishment denied what they had done. In other words, they lied about the fact that they had peeked. The researchers concluded that the punitive school environment gave children the motivation to lie as a means of self-protection.

Similar to experiments conducted in a laboratory, researchers in this study controlled as many variables as possible with the exception of the variable they were interested in studying. Both schools were located in the same city and enrolled students from similar socioeconomic backgrounds. Students from the two schools also scored at a similar level on some standardized tests of cognitive ability. The relevant variable that differed between them was the discipline practices of the school they attended, so this was the independent variable in this study. The likelihood that they would lie about their misbehavior was the dependent variable measured at the end of the experiment.

One drawback in a natural experiment is that it is more difficult to rule out other factors that may affect the results. For example, in a true experiment, some teachers would be randomly assigned to use physical punishment and a comparable sample of teachers would be assigned to use nonphysical punishment. You can clearly see why this research could only be done as a natural experiment. What teacher would agree to hit a child when that teacher did not believe in using physical punishment in a classroom? However, because this was a natural experiment, it is possible that teachers who were drawn to the school that used physical punishment differed from those who went to the other school in other ways that made them less approachable to students. Therefore, it is possible that it is the teacher’s personality or some other difference in the school environment that is responsible for the outcome rather than simply the use of physical punishment.

CORRELATIONAL DESIGNS

The second way we can test hypotheses is to examine the relationship between two or more variables using a correlational research design. This design is used to look at
how two or more naturally occurring variables relate to each other; researchers do not create an independent variable as they would in an experimental research design. When we look at correlations, we are interested in two things: the strength of the relationship and the direction of the relationship. Figure 3.2 illustrates these aspects of correlations. The first is the direction of the relationship. Correlations can be positive or negative. In a positive correlation, the value of one variable increases as the value of the second variable increases. For example, lifetime earnings are positively correlated with the number of years in school. As years completed in school go up, so do lifetime earnings. (We hope you find that to be good news.) In a negative correlation, as the value of one variable increases, the value of the second variable decreases. For example, the more often people brush their teeth, the fewer cavities they are likely to have.

The second characteristic of correlations is the strength of the relationship between two variables. This can range from a correlation of +1.0 (a perfect positive correlation) to a correlation of −1.0 (a perfect negative correlation). You would realistically never expect to find a perfect correlation between variables related to development. Many correlations in development research are in the moderate range of +.15 to +.40. As a correlation moves down from +1 or −1, the relationship gets weaker and weaker. When a correlation reaches zero, it means that there is no relationship between the two variables. For example, the correlation between people's shoe size and their IQ would probably be close to zero because there is no reason to think that these characteristics are related to each other in any systematic way. You can test your understanding of the direction and magnitude of correlations by completing Active Learning: Positive and Negative Correlations.

**Active Learning**

**Positive and Negative Correlations**

A school counselor uses three different tests to assess children's aptitude for math so she can advise the school about the best placement for each child. After testing 400 children, she measures each child's actual performance in math classes and correlates it with the results of the three tests. She finds the following correlations between test scores and the children's math performance:

- Test 1: −.85
- Test 2: +.35
- Test 3: +.65

Which test should she keep as the best predictor of how a child is going to do in math, and why is that test the best predictor?

**SOURCE:** Adapted from Chew (2006).

---

**Positive correlation**

A correlation in which increases in one variable are associated with increases in another variable.

**Negative correlation**

A correlation in which increases in one variable are associated with decreases in another variable.
Examples of correlations. In these graphs, each dot represents one individual’s scores on Variables 1 and 2. Lines that slope upward indicate a positive correlation, and ones that slope downward indicate a negative correlation. The spread of the data points around each line shows how strong the correlation is. The closer the points are to falling on a straight line, the stronger the correlation.

Let’s think about how you can use information from correlational data. If you read an article in your local newspaper that said, “Study finds that mothers who talk to their children a great deal have children with high self-esteem,” could we correctly conclude that frequent conversations with their children will build self-esteem? We could not. Correlational research tells us that there is a relationship between two variables (in this case, mothers’ conversations and children’s self-esteem), but we do not know what, if anything, was controlled in the study. A third variable that wasn’t even measured by the correlational research might be responsible for the relationship we observed. For instance, it may be that the mothers who talked a lot to their children were also ones who gave their children a lot of praise and positive feedback. In this case, it wouldn’t be the amount of what they said that mattered as much as the nature of what they said. There are a lot of correlational studies in the developmental literature, and we know a good deal about a topic when we understand the relationship between two variables, but we must use caution when interpreting correlational findings because the fact that two things occur together does not mean that one of them necessarily caused the other. In Chapter 1, you were advised to be a careful consumer of information about development. If you know that when there is a correlation between two variables it is not the same thing as saying that one causes the other, it will make you a better consumer of information you might hear on television or read on the Internet.

Why would a researcher choose to use a correlational design rather than to conduct an experiment? One reason would be that doing a correlation is a much simpler procedure. You simply need to collect the data for the two variables you are interested in. You could even use archival data collected by someone else. That is why correlational information can be used as a starting point for research with an experimental design. For instance, based on the correlational finding described previously, we could develop an intervention that trains mothers to use more positive statements in their conversations with their children. Within an experimental research design, we could then compare their children to other children whose...
mothers talk to them the same amount but who use fewer positive statements. If the amount of conversation is the same for both groups of mothers, but the mothers who use more positive statements have children with higher self-esteem, we can then correctly conclude that positive statements (not just the amount of conversation) are a cause of higher self-esteem in children.

Before moving on to our next topic, check your understanding of the different types of methods used in child development research by completing Active Learning: The Scientific Method—Research Designs.

**Active Learning: The Scientific Method—Research Designs**

Look again at the abstract of the article by Levine and Waite (2000) found in Active Learning: The Scientific Method—Forming a Hypothesis near the beginning of this chapter. In it you have enough information to answer these questions:

1. Does this study use an experimental or a correlational research design? Here is a hint: Did the experimenters randomly assign children to different groups who then received different treatments?
   
   **Answers:**
   
   1. The children were not randomly assigned to groups. The researchers looked at the relationship between the amount of television they watched and their attentional difficulties (that is, they correlated these two measures). Therefore, this is a correlational research design, not an experimental one. It is also possible that the researchers might have measured other factors such as the amount of attention paid to the television or the amount of time spent watching television, which could be related to attention difficulties.

2. Can the researchers conclude that television viewing causes children to have attention problems? Why or why not?

   **Answers:**
   
   2. The answer again is no. Because this is a correlational design, the researchers cannot conclude anything about causation. They can only state that television viewing is related to attention difficulties. However, it is possible that children who have attention difficulties are more likely to watch a lot of television because it holds their attention more effectively than other activities, so we cannot be sure which variable is the cause and which is the effect. It is also possible that a third variable, such as parental depression or neglect, might cause both increased television viewing and attention problems. Therefore, we cannot make a direct cause-and-effect statement.

**Developmental Designs**

If we define development as “change with age,” there are several ways to examine the changes that occur as children grow and develop. As we look at these designs, we will find that each has its own advantages and disadvantages, but that each also has a place in answering the complex questions we have about development.

**Longitudinal Research**

A longitudinal design follows one group of individuals and gathers data from them at several points in time. The biggest advantage that comes with doing a longitudinal study is that it gives us the clearest picture of how the variables we are interested in change as a function of age.

To illustrate, Côté, Tremblay, Nagin, Zoccolillo, and Vitaro (2002) conducted a longitudinal study of 1,569 children and adolescents from kindergarten through sixth grade. Teachers rated each child on several behaviors once a year over this period. The results from this longitudinal study found that the risk for developing a conduct disorder in adolescence was highest among the boys who maintained a high level of hyperactivity over time. Figure 3.3 shows the results from this study.
Although longitudinal research provides unique and valuable information, it can be a difficult method to use. Stop for a moment and decide for yourself what would be some of the challenges of following and testing more than 1,500 students each year for 7 years. It should be clear to you that this was a very ambitious research effort.

Because you need to track your study participants across a period of time, it is inevitable that some of the participants who begin a longitudinal study will not complete all the waves of data collection. The loss of participants from a sample over the course of a longitudinal study is called attrition. If all participants had an equal risk of dropping out of the study before its completion, this would reduce the size of your sample and reduce your ability to detect differences among the participants, but it might not otherwise jeopardize the validity of your research. However, all students are not at an equal risk of dropping out. Some children and adolescents are more likely than others to move, to be unavailable when one or more of the data collections occur, or simply to decide to withdraw from the study. For instance, it would be more difficult to retain children who are poorer, less healthy, or more trouble-prone because they are more likely to be absent from school more often or may move more frequently and change schools.

To the extent that it is children with these characteristics who are more likely to drop out, the final sample is less representative than it was when the study began. Although we may have started with a representative sample, over time sample bias will creep into our longitudinal study. If the most trouble-prone children are the ones most likely to drop out, the children who remain in the study at the end may be functioning at a higher level in any number of ways compared to the children who were lost from the sample, and this would inflate our final estimate of the children’s abilities or level of functioning.

If you have ever watched the national college basketball play-offs (the NCAA “March Madness”), you have an idea of how sample bias creeps into a longitudinal study. You

---

**FIGURE 3.3**

**Longitudinal study of hyperactivity.** In this study, four patterns of hyperactivity were found for boys over time. Almost 17% never showed hyperactivity, 35.8% had decreasing levels of hyperactivity, 10.9% had slightly increasing but low levels, and 36.5% remained hyperactive from age 6 through 12. Boys who showed a consistently high level of hyperactivity were at greater risk of developing a conduct disorder than those who had lower levels.

![Graph showing the longitudinal study of hyperactivity.](source)


NOTE: The trajectory in bold purple represents the risk trajectory.

---

**Attrition** The loss of participants over the course of a longitudinal study.

**Sample bias** Changes in the makeup of the sample in a longitudinal or sequential study that make the sample less representative over time.
could measure the average basketball skill level across all the players on the 64 teams that begin the competition. However, during the play-offs, teams are defeated and drop out of the competition. If you based your final estimate of the level of basketball skill of the players on the teams that made it through to the Final Four, your estimate would be considerably higher than your original estimate because the teams with the weaker players have been eliminated. The same thing happens as more children with greater challenges in their lives drop out of longitudinal research, and the longer the period of time covered by the study, the greater this risk becomes.

To answer the question that we asked earlier about the challenges of doing longitudinal studies such as the one conducted by Côté and colleagues (2002), you probably realize that it takes a good deal of time and money and a large number of research personnel to conduct a study of a large group of individuals across multiple waves of data collection. Many researchers do not have the resources they would need to conduct this type of research.

Other challenges of doing longitudinal studies include the fact that researchers are locked into using one set of measures, even if better alternatives come along during the course of the study. If researchers change measures during the study and find changes in the level of the outcomes they are measuring, they cannot be sure whether the change was due to the fact that the participants were older or whether the new measure was actually measuring something slightly different than the original measure. In addition, if the same or similar measures are used repeatedly, it is possible that there is a practice effect and a participant’s response will be influenced by how that individual answered questions at an earlier point in the study. Despite these concerns, however, longitudinal research still provides a very powerful way to look at developmental change, and that is why it is widely used.

**Cross-Sectional Research**

A cross-sectional design is an approach that uses multiple groups of participants who represent the age span of interest to the researcher. If you were interested in developmental changes between elementary and middle school, you could use groups of participants who were 8 years old, 10 years old, and 12 years old and collect data from all the groups at about the same time. Then, by comparing the results between groups, you could construct a picture of the changes that occur over that period of development. Because all of the data collection occurs at the same time, you can efficiently collect data from a large number of participants in a relatively quick, cost-effective manner. You don’t need to wait years between your data collection points. Obviously participant dropout and sample bias is not an issue because there is only a single data collection. However, there are some drawbacks of using a cross-sectional design. Based on cross-sectional research, you will know that children of different ages show differences on the outcome you measured, but you won’t know why. We presume that it is age changes that are responsible, but we need to be careful when making these presumptions.

One of the challenges in doing cross-sectional research is that the different age groups in the study must be as similar as possible on any variable that might affect the study’s outcome. Here is an extreme example to make this point clear: Imagine you are interested in how self-esteem changes during the transition from elementary to middle school. To examine these changes, you study a group of 8-year-old students who attended a public elementary school in a disadvantaged neighborhood, a group of 10-year-old students from a private school with a religious affiliation, and a group of 12-year-old students from a suburban public school. Even if you found differences in self-esteem between the groups, could you...
correctly interpret them as age-related changes or changes associated with school transitions? Clearly you couldn’t. Because the groups came from such widely different school settings (and, therefore, it is likely that they differ from each other in a variety of ways in addition to age), you could not make any valid interpretation of these data. The differences that may exist between groups used in actual cross-sectional research would be much more subtle than those in this example, but any difference between the groups that is not recognized and accounted for by the research can be a threat to the validity of the conclusions drawn from this type of research.

Finally, cross-sectional research can be affected by what is known as the cohort effect. A cohort is a group of people born at about the same point in historical time. A cohort effect becomes a concern in cross-sectional research when the age range represented by the different groups is large enough that the participants in the groups come from different birth cohorts or when some aspect of the environment has changed significantly between the groups in the study. We use the term baby boom generation to describe people born between the mid-1940s and the mid-1960s; Gen Xers were born between the early 1960s and the late 1970s; and millennials were born between the late 1970s and about 1995. You can easily see how members of each of these cohorts would have had different experiences while growing up that would make them different in a number of ways. For example, millennials grew up with computers, but baby boomers never saw a personal computer until they were adults. If we were to measure competence in computer use, we would likely find that the millennials are more skilled than the baby boomers. However, we could not conclude that as people get older they become less skilled at using a computer. The difference we see is likely due to a different set of life experiences, not to a loss of computer skills with increasing age. This is an important point to keep in mind because researchers who conduct cross-sectional research want to conclude that differences seen between age groups in cross-sectional research are due to developmental changes, and not to something else.

Sequential Research

Finally, sequential designs bring together elements of cross-sectional research and longitudinal research. This design uses several groups of people of different ages who begin their participation in the study at the same time (just as cross-sectional research does) and follows the groups over a period of time (just as longitudinal research does). What makes this method unique is that there is overlap between the groups on their ages at one point in the testing.

For example, if we were interested in looking at children’s health over the age range from birth until age 20, we could begin by assessing four different groups: infants, 5-year-olds, 10-year-olds, and 15-year-olds. If we repeated our assessment 5 years later (when the infants were 5 years old, the 5-year-olds were 10, etc.), we then would have two different groups that had been assessed at age 5. Because we only needed to follow the groups for 5 years, we would have reduced the risk of participants dropping out of the research study to below what it would have been in a 20-year study (and therefore reduced sample bias). We also would have reduced the time, money, and personnel needed to conduct the study compared to a 20-year-long study of children’s health. Finally, if there were any cohort differences between the groups, those effects would have become apparent in the results when we compared the results of the group that ended the study at a given age to the results of the group that started the study at that age.

Although some advantages are gained through the use of sequential research, you still need to construct the cohort groups so that they are as much alike as possible at the start of...
the study, and you still need to be able to track and reassess the groups at regular intervals, so sample attrition and practice effects are still potential problems.

Microgenetic Research

A fourth type of developmental research design is a microgenetic design. In this case, the term genetic refers to the genesis or beginning of a behavior, not to the action of our genes. Micro refers to the fact that this approach looks at small, moment-to-moment changes that eventually lead to larger developmental change. This is a particularly good way to look at naturally occurring behaviors in a social context (Lavelli, Pantoja, Hsu, Messinger, & Fogel, 2005).

To use a microgenetic design, an individual or small group of individuals are observed frequently during the time that a change is expected to happen. The frequency of the observations allows researchers to see changes that otherwise would be missed by designs that rely on less frequent observations. The full picture of developmental change that results describes not only what happens and when it happens, but also the process by which the change occurs (Fogel, 2011). For example, Spencer and colleagues (Spencer, Vereijken, Diedrich, & Thelen, 2000) used this design to study the emergence of reaching behavior in infants by observing four infants on a weekly basis between the ages of 3 and 30 weeks. They found that infants could reach out and grasp a toy they wanted between 12 and 22 weeks of age. They also found that an infant needed to have certain component skills in their behavioral repertoire before they could reach for an object, but that these skills did not necessarily appear in the same order for each infant. This illustrates why microgenetic designs are a good way to study individual differences. The fact that different infants took different pathways to reaching would have been lost if the researchers used another type of design.

Microgenetic designs also are particularly good at answering one of the questions introduced in Chapter 1: the question of how much stability versus change there is in development. A microgenetic design identifies new behaviors (the change elements) that build upon earlier ones (the stable elements) to shape the trajectory of development.

As with any research design, there are advantages and disadvantages of microgenetic research. Participating in this type of research requires a strong commitment on the part of the participants. In one study, parents brought infants to a laboratory for observation every week for a full year (Lavelli et al., 2005). However, this level of commitment to a study also means that parents and children become true collaborators in the research process. We also need to recognize that because the design requires repeated testing, some of the change we observe may be due to a practice effect. Because of the intensive nature of the observations, this is a time-consuming and expensive approach. And, while it produces a great deal of data that is rich and informative (Flynn, Pine, & Lewis, 2006), it also means that all that data must be analyzed and reduced before it can yield an understandable set of conclusions. This can be a daunting task.

In summary, while other developmental research designs give us a snapshot of what is happening at different points in development, the microgenetic approach gives us insight into the process of change as it is actually happening (Flynn et al., 2006). The number of microgenetic studies has increased in recent years and its use appears to be growing.

Table 3.3 summarizes the characteristics, advantages, and disadvantages of the four developmental research designs we have described.

CHECK YOUR UNDERSTANDING

1. Explain why an experiment can identify the causes of behavior while a correlational study cannot.
2. Describe the advantages of using a longitudinal design and a cross-sectional design.
3. Describe the advantages of a sequential design over longitudinal and cross-sectional designs.
4. Describe the advantages of a microgenetic design over typical longitudinal and cross-sectional ones.
TABLE 3.3

Comparison of developmental research designs.

<table>
<thead>
<tr>
<th></th>
<th>Longitudinal Designs</th>
<th>Cross-Sectional Designs</th>
<th>Sequential Designs</th>
<th>Microgenetic Designs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample Size</strong></td>
<td>Large</td>
<td>Large</td>
<td>Large</td>
<td>Small</td>
</tr>
<tr>
<td><strong>Length of Study</strong></td>
<td>Long-term, typically from months to years</td>
<td>Short (all children participate at about the same time)</td>
<td>Relatively long</td>
<td>Typically brief, but with very frequent observations</td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
<td>Captures the continuity of development</td>
<td>Quicker, easier, less expensive than longitudinal studies</td>
<td>Can detect cohort effects</td>
<td>Captures both stability and change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Less time-consuming and expensive than longitudinal designs</td>
<td>Lower risk of attrition and threat of sample bias</td>
<td>Good for studying individual differences</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Can identify factors that promote or hinder developmental change</td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>Requires multiple points of data collection</td>
<td>Requires that groups are carefully matched on a number of characteristics</td>
<td>Requires that groups are carefully matched</td>
<td>Requires strong commitment from participants</td>
</tr>
<tr>
<td></td>
<td>Expensive, time-consuming</td>
<td>Potential cohort effects</td>
<td>Requires multiple points of data collection</td>
<td>May create a practice effect</td>
</tr>
<tr>
<td></td>
<td>Risk of attrition and threat of sample bias</td>
<td></td>
<td></td>
<td>Time intensive and expensive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Great deal of data to reduce and analyze</td>
</tr>
</tbody>
</table>

INTERPRETING AND COMMUNICATING THE RESULTS OF A STUDY

3.3 What must we consider when interpreting and communicating the results of a study?

After we conduct research, we must still interpret the results. The accuracy of our final understanding of what a study shows is greatly affected by how the data are interpreted, but two people could look at the results of a study and interpret them in a different way. In reaching our conclusions, we must be careful not to generalize beyond the characteristics of the sample that participated in the research. We also need to remember that conclusions drawn from research—even very carefully conducted research—are generalizations that apply to groups of individuals. As you learned in Chapter 1, there is a great deal of diversity among individuals within any group. Research tells us about what is average or typical for a group of people, so the fact that some individuals fall outside that range does not invalidate the general conclusion.

To help us correctly interpret our studies, results of research are tested using various statistical methods. If you were evaluating the effectiveness of a new program intended to improve social problem-solving skills, you would hypothesize that at the end of the experiment the students who participate in the program will have better social problem-solving skills than a similar group of students who don’t participate in the program. Tests of statistical significance give us confidence that results from research are not accidents or chance occurrences.

However, even if the results of a research study are statistically significant, you still might wonder whether the findings make any difference in the real world. Just because a study finds

---

Copyright ©2018 by SAGE Publications, Inc.
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.
that differences between two groups are statistically significant (that is, they did not happen by chance), it does not necessarily mean that the difference is large or important. A real-life example of this distinction comes from research on the relationship between birth order and intelligence (Kristensen & Bjerkedal, 2007). An analysis of the IQ test scores of over 250,000 young men in the Norwegian military service found that the scores of the first-born men in the sample were higher than the scores of second-born men at a statistically significant level of .05 (meaning there are only 5 chances in 100 that this is a chance finding). But before any first-born readers of this text begin celebrating their intellectual superiority over their siblings, you need to know that the difference in test scores between the two groups was only 2.82 points. Although these results are statistically significant, it is not likely that a difference of less than 3 IQ points will have any practical significance in the lives of these men. Results that are statistically significant do not always translate into something that has consequences in the real world. Increasingly, researchers are also reporting effect sizes, which are a statistical measure of how large the difference is between groups that are being compared. While the significance level tells us that there is a difference, the effect size tells us how big that difference is (Sullivan & Feinn, 2012). Sometimes a statistical procedure called meta-analysis is used to combine data from different studies to determine whether there is a consistent pattern of findings across studies. We will discuss the results from meta-analyses on several important topics throughout this book.

After researchers have confidence in the conclusions drawn from their research, they share that information with others. In Chapter 1, we described the peer review process that articles go through before they are published in journals. A similar process precedes the presentation of research information at professional meetings. This process gives others who are knowledgeable about the topic the opportunity to critique the way that the research was conducted and identify any possible flaws in the logic or problems with the methodology, analysis, or interpretation of the findings before those findings are shared with others.

Researchers also make their information available to the public, government agencies, private organizations, and policymakers who can use this information in ways that benefit children and families. In Chapter 1, you read about the importance of sharing findings with policymakers so they can be translated into public policy. Most university websites feature research currently being conducted by their faculty and students. Think of how often you see new findings from the field of child development covered as a story on the evening news or in a local newspaper or national magazine. All of these avenues help disseminate research findings to a wider audience than just the scientific community.

CHECK YOUR UNDERSTANDING

1. Why must we be careful when we generalize the results from any one study?
2. What is the difference between statistical significance and effect size?

**ETHICS IN RESEARCH WITH CHILDREN AND ADOLESCENTS**

3.4 What special precautions must be used when our research participants are children?

Any research with human participants must ensure their safety and well-being. U.S. Department of Health and Human Services (2005a) regulations provide specific protections...
for research participants. They may be exposed to only minimal risks during their participation, and any potential risk must be weighed against the anticipated benefits from the research. They must be informed of the purpose of the research and its risks and benefits and must freely agree—without coercion—to participate; they have the right to withdraw from participation at any point. Finally, the privacy of participants must be protected, and the data collected must be treated as confidential.

Because of their particular vulnerability, children are given additional protections when they participate in research. The Society for Research in Child Development (SRCD, 2012) has developed specific guidelines that include:

- the expectation that no physical or psychological harm will be done to children who participate in research.
- the expectation that the researcher will use the least stressful research procedures possible.
- the necessity to obtain consent from both the children and their parents for the child’s participation. If children are not old enough to give consent because they do not necessarily understand the full significance of the research, they still must assent if they are old enough to do that.
- the right of children and families to freely choose to participate in the research and to be able to terminate participation at any point if they want to.
- the responsibility of the researcher to inform parents of any threats to the child’s well-being that they become aware of during the course of the research and to arrange for assistance for the child.

It is only with the help and cooperation of children and families that the field of child development can continue to build its understanding. They are truly participants in this process.

CHECK YOUR UNDERSTANDING

1. What does the U.S. Department of Health and Human Services require as ethical guidelines for research with human participants?
2. What additional considerations come from the Society for Research on Child Development for research involving children?

CONCLUSION

In this chapter, you received a broad introduction to the various ways that researchers add to our knowledge base in child and adolescent development. Each approach has advantages and disadvantages, so no one approach is the best choice in all situations. Instead we strive to find an approach or method that is appropriate for the type of research we are conducting. Beyond that, as information accumulates over time, we build our confidence in our findings and conclusions. This is particularly true when different methods have been used by different researchers but the information they find fits together into a coherent picture.
3.1 What is the scientific method and how is it used to study development?

The scientific method is the way we add knowledge to our understanding of child development. It begins with observations, which generate hypotheses. After we operationalize the concepts in our hypotheses, we select a representative sample from the population that participates in the research. Any measures that we use must have good reliability and validity. Research can be conducted using observation, checklists of behaviors, or by asking children (or people who know them well) to report on behaviors. Self-report measures include surveys, questionnaires, and interviews (including clinical interviews). We also gather data by using standardized tests, physiological measures, or archival records. Sometimes a case study is conducted to intensively study a single individual who is of interest or the researcher conducts an ethnographic study to study a different culture.

3.2 What are the different types of research designs used to study development?

Research designs are the roadmap or blueprint that we use to structure our research. We can use an experimental design, or a correlational design, and we can use one of several developmental designs: longitudinal, cross-sectional, sequential, or microgenetic designs. In an experimental design, participants are randomly assigned to either the experimental group or the control group. The independent variable is the treatment given to the experimental group but not to the control group. At the conclusion of the experiment, the dependent variable is measured for both groups and compared. If there is a difference, we can conclude that the independent variable caused the change. Because the researcher has a great deal of control over what happens during the experiment, the results can help determine the causes of the behavior we observe. A correlational research design measures the strength and direction of the relationship between two variables. Although a correlation indicates that two variables are related, it cannot tell us what causes the relationship. In a longitudinal design, a single group of study participants is followed for a period of time and is tested or assessed repeatedly. In a cross-sectional design, several groups of participants of different ages are assessed at the same time and compared to get a picture of how changes occur as a function of age. A sequential design follows several groups over time, with overlap in the ages of the participants at the times of each test. In a microgenetic design, the researcher makes frequent, detailed observations of the participants around the time that a developmental transition is expected to occur.

3.3 What must we consider when interpreting and communicating the results of a study?

The results of a study must be interpreted with caution, being careful not to generalize them beyond the characteristics of the sample used in the research or to draw conclusions that go beyond the scope of the study. We need to remember that the results apply to groups, not individuals, so there will be individuals who are exceptions to the study results. Finally, we can assess the statistical significance of the findings using tests of probability, but finding statistical significance does not necessarily mean that the results will have real-world or practical significance. Effect size gives a statistical measure of how large the difference is between groups that are being compared. Through publication in research journals and presentations at professional meetings, researchers disseminate their findings. They also make their findings available to the public, governmental agencies, private organizations, and policymakers.

3.4 What special precautions must be used when research participants are children?

All research must protect the physical and psychological safety and well-being of all participants, and children in particular. Researchers must minimize any risk to them, and protect their confidentiality. If children are old enough to understand the nature of the research, they are asked to provide informed consent that is freely given. Children (or their parents) have the right to withdraw from participation in the study at any point and researchers are obligated to rectify any threats to the children’s well-being that they detect.
### KEY TERMS

Strengthen your understanding of these key terms with mobile-friendly eFlashcards at edge.sagepub.com/levine3e

<table>
<thead>
<tr>
<th>Term</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applied research</td>
<td>66</td>
</tr>
<tr>
<td>Archival records</td>
<td>81</td>
</tr>
<tr>
<td>Attrition</td>
<td>91</td>
</tr>
<tr>
<td>Basic research</td>
<td>66</td>
</tr>
<tr>
<td>Case study</td>
<td>81</td>
</tr>
<tr>
<td>Checklist</td>
<td>76</td>
</tr>
<tr>
<td>Clinical interview</td>
<td>78</td>
</tr>
<tr>
<td>Cohort effect</td>
<td>93</td>
</tr>
<tr>
<td>Control group</td>
<td>84</td>
</tr>
<tr>
<td>Correlational research design</td>
<td>87</td>
</tr>
<tr>
<td>Cross-sectional design</td>
<td>92</td>
</tr>
<tr>
<td>Dependent variable</td>
<td>85</td>
</tr>
<tr>
<td>Effect size</td>
<td>96</td>
</tr>
<tr>
<td>Ethnography</td>
<td>82</td>
</tr>
<tr>
<td>Experimental group</td>
<td>84</td>
</tr>
<tr>
<td>Experimental research design</td>
<td>84</td>
</tr>
<tr>
<td>Generalize</td>
<td>70</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>66</td>
</tr>
<tr>
<td>Independent variable</td>
<td>85</td>
</tr>
<tr>
<td>Interview</td>
<td>76</td>
</tr>
<tr>
<td>Longitudinal design</td>
<td>90</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>96</td>
</tr>
<tr>
<td>Microgenetic design</td>
<td>94</td>
</tr>
<tr>
<td>Natural or “quasi” experiment</td>
<td>87</td>
</tr>
<tr>
<td>Negative correlation</td>
<td>88</td>
</tr>
<tr>
<td>Norm</td>
<td>79</td>
</tr>
<tr>
<td>Observer bias</td>
<td>75</td>
</tr>
<tr>
<td>Operationalize</td>
<td>68</td>
</tr>
<tr>
<td>Population</td>
<td>70</td>
</tr>
<tr>
<td>Positive correlation</td>
<td>88</td>
</tr>
<tr>
<td>Questionnaire</td>
<td>76</td>
</tr>
<tr>
<td>Random assignment</td>
<td>84</td>
</tr>
<tr>
<td>Reliability</td>
<td>70</td>
</tr>
<tr>
<td>Representative sample</td>
<td>70</td>
</tr>
<tr>
<td>Sample bias</td>
<td>91</td>
</tr>
<tr>
<td>Scientific method</td>
<td>66</td>
</tr>
<tr>
<td>Sequential design</td>
<td>93</td>
</tr>
<tr>
<td>Standardized test</td>
<td>78</td>
</tr>
<tr>
<td>Survey</td>
<td>76</td>
</tr>
<tr>
<td>Validity</td>
<td>70</td>
</tr>
<tr>
<td>Variable</td>
<td>68</td>
</tr>
</tbody>
</table>