CHAPTER

Interpreting and Summarizing Published Research

Preview

This chapter provides practical suggestions to help you

- Review
- Interpret
- Annotate

to read published literature related to your topic. Included are suggestions to help you rapidly peruse reports to determine

- Topic and purpose
- Type of report
- Findings and conclusions

and to

- Skim information
- Summarize information
- Make annotated citations

Most reports, by type, are status reports, group comparison reports, and relationship reports. Interpretation of these reports often requires some knowledge of statistical analysis techniques. You may see

- For status reports: Mean, median, mode
- For group comparison reports: Chi-square
- For relationship reports: Coefficient of correlation

Targeted Learnings

In Chapter 4, you learned how to locate and use some of the most valuable secondary and primary sources of information in the research library. In this chapter, you will find suggestions on how to peruse published reports, interpret their contents, summarize them quickly, and make effective notes for bibliographic use. As you read the chapter, be sure to look for information related to these questions:

1. Why should research reports be skimmed rapidly before they are read carefully?
2. What specifically does one look for when skimming reports?
3. What entries should one make on note cards when annotating published reports?
4. What are the three general types of research reports encountered in published literature?
5. What characterizes the three different types of research reports?
6. What common analysis terminology—verbal and/or statistical—is usually associated with each of the three types of reports?
7. Why do the different types of research reports utilize different verbal and statistical procedures to analyze their data?
8. What two different meanings does the term significance have in the context of research?
9. Why is the term significant often misleading to uninformed readers of research?
You have already familiarized yourself with the major sources and locations of published research. That was the first phase in learning to use the library profitably. You are now ready to undertake the second phase, which involves accurately interpreting the published information in the sources you locate. This phase is completed in three parts, dealing with (1) how to read and summarize the material, (2) how to identify the type of report and interpret statistical analyses you might encounter, and (3) how to enter annotations on summary cards to maximize their value and ease of use.

## Reading Research Reports

Research literature comes in various forms now familiar to you, such as references, handbooks, reviews, journal articles, technical reports, and scholarly books. After becoming able to locate relevant information, how does one make effective use of it? The first step is to learn how to **skim** the material quickly to ascertain its nature and the researcher's conclusions.

### Skimming the Information

You know how to find a number of published reports related to almost any educational topic you choose. Because so much material is available, you need to be able to move through it quickly. The summaries presented in indexes such as *CIJE* and in journals such as *Psychological Abstracts* provide immeasurable help in pinpointing materials of value to your topic. However, summaries and abstracts omit many important details and can by no means replace reading the original material.

To move rather quickly through quantities of reports, you should practice skimming the materials so that you can quickly determine what the study is about and what the author concludes. By using this strategy, you can expeditiously obtain information and further identify reports that have direct bearing on your topic.

### Determining What the Material Is About

To pinpoint the focus of the article, look quickly for a **title** and an **introduction**. The title should clearly indicate the topic. Most researchers use descriptive titles for their reports, but such is not always the case. An example of a citation whose title makes clear the nature of the research being reported is


In contrast, fictitious examples of titles that would leave one guessing are

- Onward and upward in educational research
- The golden egg of bilingual education

These two titles suggest that the reports have something to do with educational research and bilingual education, but little else is indicated.

Once you have examined the title, you are ready to scan the introduction of the report, which may be labeled as such. If not labeled, the first paragraph or two usually introduces the report and tells why the topic is considered important.
Identifying the Writer’s Conclusions

To find the conclusions presented by the writer of the report, look for an abstract, summary, or a section near the end of the report entitled Findings, Conclusions, or Discussion. If there is no such heading, the conclusions are usually presented in the final paragraphs of the report or article.

Some journals—the exceptions rather than the rule—present an abstract at the beginning of the article. This is helpful to readers since it provides a brief synopsis of the article. If there is no abstract at the beginning, flip to the end of the article to see if there is a summary, which will indicate concisely what the article is about and what are its conclusions.

Cautions to Keep in Mind

As you skim documents that describe (1) ethnographic research (such as patterns of interaction among teachers), (2) historical research (such as the backgrounds of the first teachers in Elmwood schools), or (3) action research (such as development of a special spelling program for Cutter Elementary School students), remember that the conclusions presented in those reports may be correct for the specific groups and locales investigated, but that does not necessarily make them applicable elsewhere.

On the other hand, if you are reading technical reports of research funded by grants or reports published in research journals, most of the information you encounter will be intended for wide dispersion and will have been subjected to close scrutiny by experts in their respective fields who judge the research methodology, analysis of data, and appropriateness of conclusions. When reports meet these criteria, the results can nearly always be generalized to similar locales and populations. This is not to say that the conclusions are invariably correct, nor that you should never analyze critically the methods and conclusions of reports published in prestigious journals. But the likelihood is great that researchers reporting those investigations have followed established procedures, analyzed the data appropriately, and formulated conclusions that follow from the analyses and seem logically correct.

Summarizing the Reports

Suppose you have read a journal article reporting a large study on high school students’ attitudes toward the use of alcohol. You have gleaned the following information:

Lazerus surveyed 2,000 high school students in four eastern states concerning their attitude toward the use of alcohol . . . found that 42 percent approved of the use of alcohol in general, but that only 17 percent approved of their parents using alcohol.

In what form should you make your notations so they will be most accurate and useful? Good notations should not only summarize information, as the preceding entry does, but also indicate the topic and include a complete bibliographical citation. If you do a preliminary search via computer, you can print out the citations, thus ensuring completeness and accuracy. If you do a manual search, it is very important that you make your notes clear and complete. It is irritating and a frustrating waste of time—not to mention a sometimes unrealistic task—to have to go back later to find an item of information that was omitted or written illegibly.

If you have note-taking software and a laptop computer, the computer can enter, organize, and rearrange your notes in a number of ways. If you do not have that capability, it is suggested that you enter your summary notes on 5- by 8-inch index cards. Place each reference on a separate card. These cards are large enough to hold a good deal of information, and they are easy to store, organize, regroup, and retrieve. Index cards are valuable even if you have used a computer to print out citations and summaries. Simply cut and tape the
citations onto the cards. But keep in mind that the article abstracts seldom contain all the
details you need for your research. Those important notations must be made by hand.

One helpful suggestion is to write the topic of the report in the upper-right-hand corner
of the card and write the author’s name and date of publication in the left-hand corner. By
doing this, you can easily group cards by topic, publication date, or alphabetized authors’
names. This will save time when you begin organizing your references and preparing your
bibliography.

An article citation must include all of the following:

- Author’s name
- Date of publication
- Title of article
- Name of journal
- Volume of journal and issue (issue optional)
- Page numbers of the article

A book citation must include the following:

- Author’s name
- Year of publication
- Title of book
- City where published
- Name of publishing company

As previously mentioned, make sure to record the citation correctly in its entirety. Make
sure, also, that you enter notational information accurately, and if you quote material di-
rectly, indicate so clearly. Your card for a journal article might look like that shown in Fig-
ure 5.1. Note the author, date, topic, complete citation, and notes about principal findings.
(All aspects of this particular entry are fictitious.)

Figure 5.1 Summary note card for a journal article


Cruise and Fisher (2002) surveyed elementary-level teachers throughout the Midwest in an
tempt to describe their current attitudes toward the No Child Left Behind (2001) legisla-
tion. The authors collected their data through the use of a self-developed survey, adapting
many items from existing surveys into their own survey. The survey consisted primarily
of Likert-scaled items, but also included several open-ended items. Evidence of validity
included a content review of their original instrument by five individuals with expertise in
survey design and two individuals employed at state departments of education. Reliability
of their resulting data was calculated using Cronbach's alpha coefficient (α) and was equal
to .89. Five thousand five hundred teachers were randomly selected from state department
of education databases across 11 states. One thousand nine hundred eighty-three (n =
1,983) usable surveys were returned, for a return rate of 36%.

Results included the fact that the majority of teachers believe in the overall purpose of
NCLB but disagree with the ways in which it is being implemented. Specifically, teachers
agreed with statements that NCLB has forced them to change the focus of their classroom
instruction and that both teachers and students are experiencing greater levels of stress
due to various requirements of the law. Findings also seemed to indicate that teachers may
not have a clear understanding of the overall impact of the legislation. Recommendations
for teacher professional development, focusing on improving teachers’ understanding of the
law and its probable influence on classroom instruction, are discussed.
From the following fictitious information, write note cards in the manner suggested:

1. James F. Roberts did research on teacher partnering that showed that men are somewhat lazier than women when it comes to doing menial tasks. That finding was reported in his 1995 book, Unequal Teacher Partners, which was published in New York by the Domestic Press.

2. Cecilia A. González studied role acceptance in teacher partnering. In her article “Parameters of Teacher Role Acclimatization,” published in the July 1993 (volume 4, pp. 12–18) Journal of Teacher Partnering, González reported that she found no gender differences in teachers’ ability to adjust to new roles in the partnering process.

Interpreting the Statistical Information You Encounter

Researchers use words and phrases in special ways to express concepts clearly with an economy of language, but the special terms they use are usually confusing to people not schooled in research. If you are to understand the research reports you read, you must have a grasp of the fundamentals of this research language, especially having to do with statistics, significance, and conclusions.

Suppose you have read a report that said, “The difference between the groups was not significant.” Or, “The coefficient of correlation was .85.” Or, “The standard error of the mean was 1.72.” What would these statements mean to you? If you are a beginner in research, they would mean little or might even be misleading.

Chapter 8 and Appendix A present considerable information about statistical terms and procedures. However, certain terms must be introduced here if you are to read research reports—even their conclusions—with adequate understanding. You can learn the meanings of these terms more easily when they are presented in relation to types of research findings. These types of findings generally consist of one or more of the following:

1. The status of one or more groups—say, a school, a community, a peer group, the eleventh grade in Elmwood High School. This status describes people, places, events, objects, and the like, as they now exist or once did. Status reports are typically appropriate for qualitative and survey research studies.

2. Comparisons between two or more groups. Again, this might involve schools, communities, grade levels, gender, ethnic groups, and so forth. These comparisons are done to determine the existence and nature of differences that might exist between
or among groups—differences such as preferences, lifestyles, achievement, rates of learning, or access to resources. These differences may exist inherently in the groups or may have been produced by introducing a variable that caused one group to become different from the other. Reports of this types are appropriate for nonexperimental research studies (such as causal-comparative studies), as well as for experimental, quasi-experimental, and evaluation studies.

3. Covarying relationships between two or more sets of measurements (such as between scores of language ability on the one hand and reading ability on the other) obtained from the same group of individuals. Such relationships, when confirmed, enable you to predict one of the variables from the other, and the relationships may imply—although they do not convincingly demonstrate—the existence of cause and effect between the variables. Correlational reports are, of course, appropriate for correlational studies, another type of nonexperimental quantitative research.

Let us examine some of the terms you can expect to encounter in reports that present these three different types of findings.

**Status Reports**

Status reports describe things as they are or once were. Often status studies are qualitative in nature and do not make use of statistical procedures; the findings in qualitative research are ascertained through verbal logic and are presented as verbal statements. Such is often the case for ethnographic and historical research. But other status studies are quantitative in nature and involve numbers and testing. In those reports, you are likely to encounter the use of the following types of descriptive statistics and terms.

**Raw Numbers or Raw Scores**

Raw numbers or raw scores are numerals that indicate individuals counted (e.g., 35 adolescents) or scores made on tests before they are converted in any way (e.g., 78). You can expect to see entries such as these:

- Elmwood High School was opened in 1874, with 32 students attending grades 8 through 12. Today it serves 2,750 students, grades 9 through 12.
- Samuel's raw score on the math section of the Stanford Achievement Test was 43. Raw scores of other students in Samuel's class ranged from a high of 45 to a low of 6.

**Terms That Indicate Typicality or Central Tendency**

Three terms are commonly used to depict what is average or typical for a group of raw numbers or scores:

1. **Mean**, symbolized $M$ or $\bar{X}$. The mean is the arithmetic average of a group of raw scores or other measurements that are expressed numerically. Commonly thought of as synonymous with the term *average*, the mean is calculated by adding the raw scores together and then dividing the sum by the number of scores.

2. **Median**, symbolized $Mdn$ or $Md$. The median is the point halfway between the highest and lowest scores of a particular array (group of scores). In other words, it is the value that separates the bottom 50 percent of scores from the upper 50 percent of scores. The median is determined by arranging the scores in order from lowest to highest, then counting halfway through the number of scores. For example, the median for the following array of scores would be 4:

$$1, 2, 2, 3, 4, 6, 12, 13, 13$$
Writing a Review of Literature

Writing a review of literature is an integral part of any formal proposal or design of a research study. A literature review involves the "systematic identification, location, and analysis of documents containing information related to the research problem" (Gay, Mills, & Airasian, 2006). Once you have obtained documents related to your topic, you must analyze them for several reasons:

- To determine what has already been done.
- To learn about specific research strategies, procedures, and instruments that have been shown to be effective in researching your topic.
- To form the basis for interpreting your results (Gay, Mills, & Airasian, 2006).

However, you must avoid the temptation to include everything you have found in your search for published literature. Additionally, a literature review is not simply a compilation of summaries of articles, one after another. On the contrary, a literature review is an opportunity for you to demonstrate your ability to "recognize relevant information . . . and to synthesize and evaluate it" according to the overall purpose and goals of your study (Taylor, 1998).

Writing the literature review is quite honestly not an easy task. There is no "recipe" for developing the actual written review. It will likely take several iterations of written drafts, perhaps organizing and reorganizing the analysis of your related literature. Although there exists no formula for this important research activity, several websites offer many suggestions that can be especially helpful. We review a sampling of those sites here:

- Writing a Literature Review in the Health Sciences and Social Work (www.utoronto.ca/hswriting/lit-review.htm)

This site poses several sets of questions to help you look critically at your literature review; the questions focus on your review as a whole and on your assessment of each book or article you have included.

In addition to this site, Purdue University has developed an Online Writing Lab (http://owl.english.purdue.edu/oldindex.html), which includes links to several helpful pages, many of which incorporate activities for understanding and reinforcement:

- Paraphrase: Write It in Your Own Words (http://owl.english.purdue.edu/handouts/research/r_paraphr.html)

This site explains what paraphrasing is, why it is an important skill in technical writing, and includes an example containing an original passage, a legitimate paraphrase, and a plagiarized version; also provides six steps to effective paraphrasing.

- Quoting, Paraphrasing, and Summarizing (http://owl.english.purdue.edu/handouts/research/r_quotparsum.html)

Building on the previous site, this one offers explanations of the differences between quoting, paraphrasing, and summarizing published work. A brief sample essay is included, along with guiding instructions, for writing a summary of the piece—a good activity.

- Avoiding Plagiarism (http://owl.english.purdue.edu/handouts/research/r_plagiar.html)

This site defines plagiarism (the unacknowledged use of someone else’s work or ideas) and provides examples of actions that might be considered plagiarism, as well as some not considered to be such. The site is further composed of specific examples documenting when credit should be given to an author. Included is an activity that provides sample situations in which the reader must decide the degree to which there exists a risk of plagiarism.

Also affiliated with the development of a written review of literature is the associated list of references, which must follow a prescribed format. In education, the preferred guidelines are provided by the Publication Manual of the American Psychological Association (5th ed.), commonly referred to as the "APA Manual." Since the citation of electronic sources of information continues to be an evolving entity, and since citation format may change as technologies change, the American Psychological Association has provided this information on its website titled Electronic References (www.apastyle.org/elecerefhtml). Included on this site are the recommended citation formats for:

- E-mail communications
- Websites
- Specific documents on a website
- Articles and abstracts from electronic databases
- Web citations within the text of your document

The APA has also included specific examples of reference citations on this Web page.

Finally, as mentioned in the text, it is important to maintain accurate records of all of the sources that you have referenced in your literature review. This can be
If there is an even number of scores in the array, the median is the arithmetic average of the two middle scores. The median in the following array would be 7.5 (the average of 7 and 8):

4, 5, 5, 7, 8, 9, 10, 10

The median is less affected than is the mean by unusually high or low scores (also known as extreme scores) and for that reason is frequently used to report information such as average personal income or price of housing. Its usefulness in educational research is limited.

3. **Mode**, symbolized \( Mo \), is simply the most frequently occurring score. The mode has no other use in statistics.

### Terms That Indicate Spread or Diversity

Researchers are interested not only in what is average for a group but also in the dispersion of values within that group—that is, how spread out the scores or measurements are. Dispersion is often expressed in the following terms:

1. **Range**, which is the group spread from the highest through the lowest score or measurement. It is calculated by subtracting the lowest score from the highest score. For example, the range of an array of scores, in which the highest score is 45 and the lowest 15, would be 30: 45 - 15 = 30. The concept of range has no further application in research statistics.

2. **Standard deviation**, symbolized \( SD \) or \( s \). Standard deviation indicates the divergence of scores away from the mean of the group. Standard deviation is a concept not encountered in ordinary experience and is, therefore, difficult to describe empirically. It is analogous to the concept of an average deviation, which indicates how much each score, on average, differs from the mean. But average deviation is not a stable measure, statistically speaking; therefore, standard deviation, a stable and widely applicable measure, is used instead. Mathematically, standard deviation is closely associated with the normal curve. If a group of scores or measurements is normally distributed, 68.26 percent of them fall between plus one standard deviation and minus one standard deviation from the mean. The area beneath the normal curve between 11.96 and 21.96 standard deviations includes 95 percent of all scores, and the area between 12.58 and 22.58 standard deviations includes 99 percent of all scores (see Figure A.3 in Appendix A). These relationships are illustrated and further explained in Chapter 8 and Appendix A.

### Converted, or Transformed, Numbers and Scores

Raw scores are frequently converted or transformed into comparable scores to make them more understandable and more easily compared. For example, educators frequently work...
with norms, charts that show what is typical for certain ages, groups, or grade levels. It is important to remember, however, that once raw scores have been converted, they cannot be treated as if they were still raw scores. For example, converted scores cannot be averaged in an effort to obtain a composite (i.e., overall) converted score. Converted scores you are likely to encounter in research reports include the following:

1. **Grade equivalents** are shown in grade norms that accompany standardized achievement tests. They indicate the average scores made on a particular standardized test by students at different grade levels in various localities. The grade levels are further divided into months; they might, for example, show the mean score for students at seventh grade, sixth month; fourth grade, third month; and so on. Let us assume that Manuel, a second grader, has a raw score on a standardized reading test equal to 22, which, when compared to the norms accompanying the test, is seen to have a grade level equivalency of 3.3 (third grade, third month). A research report that contains this information might say that Manuel's reading level was 3.3. This does not imply that Manuel should be moved to the third grade, however. It simply indicates that Manuel's current level of reading ability is above the norm, or average, for second-grade students.

2. **Age equivalents** tell us the average performance of students at particular age levels. Age norms that show age equivalents accompany standardized intelligence tests. If Ricardo makes a raw score of 74 on his IQ test, one could look in the norms and find, perhaps, that 74 is the score made by the average person when 12 years, 6 months of age. It could then be said that Ricardo has a mental age of 12 years, 6 months, which may or may not be the same as his chronological (calendar) age.

3. **Percentile ranks** (symbolized %ile or PR) are converted scores that indicate one's relative standing in comparison to others who have taken the same test or have been included in the same measurement. Percentiles have nothing to do with percent correct but instead indicate relative position. Suppose you were informed that your raw score on a graduate aptitude test was 89. That would tell you virtually nothing. But if you were also informed that your score placed you at the 73rd percentile, you would understand that you did as well as or better than 73 percent of all people who had taken the test.

4. **Stanines** are conversions not into individual scores but into wider bands of scores. The word stanine comes from standard nine, in which the range of possible scores is divided into nine bands. The first stanine is the lowest, the ninth stanine is the highest, and the fifth stanine normally includes the mean of all scores made. Many raw scores fall into the fourth, fifth, and sixth stanines but few into the first or ninth.

### Review of Terminology Used in Status Reports

- **Raw scores**—scores made or numbers involved
- **Mean** *(M or X)*—the arithmetic average
- **Median** *(Mdn or Md)*—the midpoint between highest and lowest in an array of scores
- **Mode** *(Mo)*—the most frequently occurring score or measure in an array
- **Standard deviation** *(SD or s)*—indicator of the dispersion from the mean for a set of scores
- **Grade equivalents**—the average scores made by students at particular grade levels
- **Age equivalents**—the average scores made by students of particular age levels
- **Percentile rank** *(%ile or PR)*—indicator of a given score's standing, relative to others made on the same test or measurement
- **Stanines**—nine bands, showing relative position within which all scores are distributed
The reader is reminded that specific information concerning the calculations of measures of central tendency and diversity—accompanied by sample exercises—has been included in Appendix A.

Comparison Reports

We have considered some of the statistical and other special terms often employed in research that describe the status of individuals or groups. Now let us consider comparison reports, which typically involve the application of inferential statistics and which you are likely to encounter in evaluation, causal-comparative, and experimental research. Suppose a researcher compared the levels of self-esteem among various groups of students attending Elmwood Lincoln High School or the levels of reading progress of groups in Elmwood elementary schools. Reports of those studies would be likely to include one or more of the following:

Chi-square ($\chi^2$)

You will usually see this term spelled out as a word in the report, then symbolized afterward. Chi-square is a statistical procedure that allows one to determine, when measurements are expressed as categories in the form of frequency counts, whether a difference exists (1) between two groups, (2) between before-and-after measurements of the same group, or (3) what is expected for a group compared to what is actually observed for the group. Suppose Mr. Michael wants to improve his students' regard for algebra II. To determine their present attitude he asks them to respond anonymously by checking one of the following:

- I like algebra II.
- I don't like algebra II very much, but I don't dislike it either. I'm neutral.
- I don't like algebra II.

Mr. Michael tallies responses from his four sections of students and finds that:

- 5 like the class (that number is placed in the “like” category)
- 27 are neutral (that number is placed in the “neutral” category)
- 71 dislike the class (that number is placed in the “dislike” category)

He then organizes and presents his classes in a different way which he believes will cause his students to like the class better. Two months later he asks them to respond as before. This time he finds that:

- 27 like the class
- 42 are neutral
- 34 dislike the class

It seems to Mr. Michael that student regard for the class has improved considerably, but he is uncertain whether the change indicates a real improvement or simply a chance difference in the way students responded. Chi-square is a procedure that enables Mr. Michael to statistically answer his question. Chi-square compares what one expects to see against what one actually observes. In Mr. Michael’s case, a null hypothesis would assume that after the new approach to teaching had been used, one would still expect to find that 5 students like algebra II, 27 were neutral, and 71 disliked it. But the actual observation is now that 27 like the class, 42 are neutral, and 34 dislike it. The frequencies (or counts) of responses for students’ opinions before the instructional change (expected) and after (observed) follow:

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Like</td>
<td>27</td>
<td>5</td>
</tr>
<tr>
<td>Neutral</td>
<td>42</td>
<td>27</td>
</tr>
<tr>
<td>Dislike</td>
<td>34</td>
<td>71</td>
</tr>
</tbody>
</table>
It is from a table like this that chi-square is computed in order to determine whether the difference between the observed frequencies and expected frequencies is, or is not, attributable to chance errors made in selecting the participants. The procedure by which this is done, along with a complete example, is provided in Appendix A.

**Difference between Means (using the t test)**

You saw that if responses are made by categories, as in Mr. Michael's case, chi-square can be used to determine differences between groups. If, however, responses are obtained as scores or other numerical measurements, the difference between groups can be assessed in terms of group mean scores, through an analysis technique called **difference between means**.

Suppose that Ms. Jones, a colleague of Mr. Michael, also teaches algebra II in Elmwood High School. She has followed with interest Mr. Michael's efforts to improve student liking for his class, but while she admits he might have caused students to like his course better, she doubts his students are learning any more than hers. In fact, she bets Mr. Michael that her traditional way of teaching the class will produce higher student achievement than does Mr. Michael's new approach.

At the end of the grading period, the two teachers give the same final test to their students. They exchange classes, each administering the test to the other's students, and they arrange for impartial scoring. They find that Mr. Michael's students make a mean score of 88 on the exam, while Miss Jones's students make a mean score of 90. She gloats, but Mr. Michael contends that the difference between the groups' scores is not significant, that it is due only to chance or sampling error. Which teacher is right?

The t test enables us to answer that question. This procedure (called an **independent-samples t test**) analyzes the difference between the means of the two groups, to determine whether the difference is significant—that is, whether the difference of two points can, or cannot, be attributed to chance errors made in selecting the participants. The procedure for computing t tests is presented in Appendix A. A variation of this design (called a **repeated-measures, or dependent-measures, t test**) allows us to analyze the difference between the means of pre- and postmeasures taken on the same group of individuals, again looking for a significant difference.

**Analysis of Variance (ANOVA)**

**Analysis of variance** is used to test for differences between two groups when the sample sizes are relatively large and unequal, provided the discrepancy between sample sizes is not extreme (Gravetter & Wallnau, 2002), and to compare three or more groups of any size.

Suppose Mr. Michael's efforts to increase students' regard for algebra II stirred up interest in other Elmwood high schools. While many teachers preferred continuing with their customary teaching approaches, several began changing, some using individualized self-paced teaching, others large-group instruction followed by small-group tutorials, and still others using algebra to solve real-life problems. They decided to ask that the results of their efforts be analyzed statistically. All used the same final examination. They found that the mean score for traditionally taught students was 91; for those taught individually the mean was 89; those using small-group tutorials made a mean score of 92; and students using real-life situations made a mean score of 85.

Differences in test performance seem evident, but even assuming that no intervening variables unduly affected the results, the question remains: Do the mean scores reflect real achievement differences—possibly the result of methods of instruction—or are the observed differences simply attributable to chance? ANOVA, considered further in Appendix A, is used to answer that question. Variations of ANOVA include designs with more than one independent variable (factorial ANOVA) and designs with more than one dependent variable (multivariate ANOVA, or MANOVA). If, while reviewing published research, you encounter any of these variations and believe that they are beyond your understanding, do
not panic. You should simply remember that they all have the same underlying purpose—to compare groups.

Review of Terminology Used in Comparison Reports

*Chi-square* ($\chi^2$)—a procedure for determining the significance of differences when data are categorical in nature

*Difference between means*—assessment that uses the *t* test, a procedure for determining the significance of a difference between means obtained from two different groups of participants or from the same group measured twice

*Analysis of variance* (ANOVA)—a procedure for determining the significance of differences among means obtained from two or more groups of participants; also used to explore interactions among several variables

Correlational Reports

Researchers are always interested in discovering correlations, which are relationships between individuals' performances on two or more measures, such as between intelligence test scores and reading test scores. Such correlations between variables, where individuals' performances on one of the variables tends to accompany similar (or inverse) performances on the other, permit greater understanding of observed phenomena. They also permit us to predict either variable from knowledge of the other. Such predictions cannot be made with absolute accuracy because correlations are almost never perfect, but the predictions can be useful nonetheless. Examples of useful but far-less-than-perfect predictions are forecasting weather from humidity, temperature, and air pressure, and predicting overall cardiovascular health from the amount of cigarette smoke inhaled over time.

In education, it is known that intelligence and reading ability are correlated, which enables us to predict one from the other. If we know that Juan, in fourth grade, is an outstanding reader, we can predict with fairly good accuracy that he has above-average intelligence. Or, if we give Juan an IQ test and find that he scores above the average, we can predict with fairly good accuracy that he is, or can become, a good reader.

Correlations do not, by themselves, demonstrate cause and effect. High intelligence cannot be said to cause one to be a good reader, nor does learning to read well cause one's intelligence to increase. True cause–effect relationships are shown when one (independent) variable is manipulated such that a second (dependent) variable then changes as a result of that manipulation. In correlational studies, variables are not manipulated. However, correlations often suggest cause–effect relationships, which are then sometimes verified through experimental research.

Correlational reports that you encounter in the literature will prominently feature a term called **coefficient of correlation**. The coefficient is shown as a decimal number that indicates the degree of relationship between the two or more variables being investigated. The coefficient of correlation, symbolized by the italicized letter *r*, ranges from a possible value of $-1.00$ to $+1.00$. Both $-1.00$ and $+1.00$ indicate extremely high (actually, perfect) correlations. A coefficient of 0 indicates no correlation—the absence of a relationship.

The coefficient of correlation may be positive or negative. If it is a **negative correlation**, a minus sign is used (e.g., $-.48$); if it is a **positive correlation**, no sign is used (e.g., $.48$). Positive and negative correlations are of equal magnitude; that is, positive $.48$ (.48) is not larger or stronger than negative $.48$ ($-.48$). Positive and negative simply show the **direction** of the relationship. In a positive relationship, high scores on one variable tend to accompany high scores on the second variable, while low scores accompany low scores. In a negative relationship, high scores on one variable tend to accompany low scores on the
other variable. For example, quality of diet is positively correlated with overall health—the better the diet, the better the health. Quality of diet is negatively correlated with incidence of disease—the better the diet, the less frequent or serious the incidence of disease. Generally speaking, correlations, whether positive or negative, are considered high if the absolute values are equal to .70 or above, medium if between .40 and .60, and low if below .30.

Review of Terminology Used in Correlational Reports

Correlation—a relationship that exists between two or more variables, such that individuals' standings on one of the variables tend to be accompanied by similar, or inverse, standings on the other variable(s)

Positive correlation—high standing on one variable tends to accompany high standing on the other variable(s), average standing tends to accompany average, low tends to accompany low

Negative correlation—high standing on one variable tends to accompany low standing on the other variable(s), average tends to accompany average, and low tends to accompany high

Coefficient of correlation (r)—a decimal number that indicates the degree of relationship in a correlation; virtually always a decimal value, the coefficient can vary from 0, which indicates an absence of relationship, to 1.0 or -1.0, both of which indicate perfect relationships but are almost never encountered in research; the positive and negative values have nothing to do with strength or closeness of relationship—they only show the direction of the correlation; generally speaking, correlations, whether positive or negative, are considered high if they are ±.70 or more extreme, medium if between .40 and .60 (or between -.40 and -.60), and low if between .30 and 0 (or between - .30 and 0).

**Exercise 5.2**

Indicate, for each of the following findings, whether the research expresses (S) status, (C) comparison, or (R) relationship.

1. The valedictorians at Elmwood Lincoln High School have been female 37 times.
2. Girls score better in reading than do boys at Cutter Elementary School.
3. Student grades can be predicted, though not perfectly, from family income.
4. The inquiry approach produced higher achievement than did memorization.

---

**The Concept of Significance**

The terms *significant* and *significance* appear in a majority of research reports. Two separate meanings are associated with these terms. The first has to do with whether or not the topic being investigated is worth the time and effort involved. We read that the topic was significant, or that the research made a significant contribution.

The second meaning is associated with statistical treatment of data. Quite different from the first, this meaning has nothing to do with importance. To say a finding is "statistically significant" is to say it is very likely the finding exists in the population as well as in
the sample—that it has not appeared just because of errors made in selecting a representative sample.

It is this second meaning—statistical significance—to which we give attention here. For a finding to be deemed significant, it must meet stringent levels of probability, usually either the .05 level or the .01 level. The .05 level of significance means there is only a 5 percent chance, on average, that a finding of a particular magnitude occurred because of errors made in selecting the sample. The .01 level means there is only a 1 percent chance that the finding is due to sampling error.

Thus, significance tests help us determine whether a finding made in a sample also exists in the population. Suppose a difference or a correlation is found in the sample. We are now left with the following questions: Does the finding for the sample also exist in the population? Or has the finding appeared because a sample was erroneously selected that does not mirror the population? Or has something occurred to bring about differences between samples that were once equal, something that would affect the population as well?

If the significance test strongly suggests that the finding cannot be attributed to error made in selecting the sample, the finding is called significant. However, if there is even a slight chance (usually just over 5 percent) that the finding is due to sampling error, the finding is deemed not significant.

Let us suppose that in a study we have found an apparent relationship between bibliotherapy and student attitude toward school. Our null hypothesis was that “No difference exists in attitude toward school between students who have participated in bibliotherapy and students who have not participated.” We know that the sample might, or might not, accurately reflect the population. To determine whether it does, we test the null hypothesis statistically. Gall, Borg, and Gall (1996) point out that, before administering the test of significance, we should have decided on the level of probability we will accept. This is called the alpha (α) level when stated in advance. We might have selected the .05 level (odds of 19 to 1 against the results being due to sampling error, shown as \( p < .05 \)) or the .01 level (odds of 99 to 1 against the result being due to sampling error, shown as \( p < .01 \)). If we selected an alpha level of .05 and our test showed a probability level less than that, we reject the null hypothesis and call the finding significant. In so doing, we move ahead to assume that the finding in our sample is probably real—that it reflects the population and is not due to errors made in selecting the sample (Bartz, 1976). Following that conclusion, logic allows us to conclude that improved student attitude was brought about by bibliotherapy.

The .05 and .01 probability levels of significance are traditional in research, but there is nothing magic about them. They are very low levels of probability of sample error, and until recently were the two levels commonly shown on probability charts used by researchers. Nowadays, statistics packages for computers can show all probability levels. In any case, researchers want to be very sure that they do not retain the null hypothesis when in fact it is not true. Therefore, they set the probability level so it is very unlikely they will make such an error. But even the very strict odds of .05 and .01 (or .001 often seen reported) leave room for some doubt about any finding. That is why researchers never report any finding with absolute confidence.

Shaver (1992) presents a concise summary of what significance testing does and does not do. He says that significance testing does provide a statement of probability of occurrence in the long run. He cautions however, that significance testing does not indicate the following:

**Tips**

FOR ONLINE COURSES

Provide students with a sample published article (perhaps available through Research Navigator™ so that it is accessible online). Require students to read and then summarize the article in a format similar to that shown in Figure 5.1. Engage in a threaded discussion about their written summaries.

- Why were certain aspects or components of the article chosen for inclusion in students’ summaries, but others not included? Discuss why students may have chosen to highlight different aspects in their summaries.
- Classify the article as either a status, comparison, or correlational report.
- Comment on the types of statistical analyses (i.e., descriptive, inferential, or none at all) used in the study. Were they appropriate for the research questions or hypotheses being addressed in the study? Why or why not?
The probability of a given finding's having occurred by chance. (The concept of chance applies to numerous future repetitions of a study, not to one particular instance of the study.)

That the null hypothesis is true or false. (It only suggests levels of probability.)

Whether a treatment being studied had an effect. (It only helps rule out the possibility of procedural error.)

The magnitude of a result. (The result can be very small and still significant or very large and not significant, depending on the number of participants involved—the more participants, the greater the likelihood of significance.)

The importance of a result. (As noted previously, statistical significance has nothing to do with importance.)

In light of these facts, Shaver recommends that researchers minimize their reliance on tests of significance and place emphasis instead on the size of differences or correlations they find.

Practicum: Exercise 5.3

Explain what the italicized phrases mean.

1. Jones found that girls scored higher than boys on 18 of 30 tasks but that the difference between the groups was not significant.

2. Adolescent boys from the Midwest were an average of one-quarter inch taller than boys from other parts of the country. The difference was highly significant.

3. A correlation of -.83 was found between the two variables.

4. The correlation was significant at the .05 level of significance.

5. The null hypothesis was rejected.

Developmental Activity: Reviewing Published Research

In Chapter 4, you began searching for and reading published research related to your chosen research topic. As you read published research, it is imperative that you analyze what you are reading for several reasons—for example, you need to become a critical consumer of research, it is important to see what research has and has not been conducted on your topic, and because published research will help guide decisions that you will make for your particular study. This activity will help you get started in this process. For the article that you review, respond to questions 1 through 10. You should be able to respond to each item, although this may sometimes require you to make inferences. After reviewing several research studies, respond to question 11.

1. What was the purpose of the study?

2. What was (were) the research question(s)?

3. The following is a brief summary of the literature review: 
4. What was (were) the hypothesis(es)?

5. Based on the types of research previously discussed, this study would be classified as:

6. What type of sampling was used?

7. How were the data collected?

8. How were the validity and reliability of the data assessed?

9. What descriptive and/or inferential analyses were used?

10. The following is a brief summary of the conclusions and recommendations:

11. Now that you have reviewed several studies related to your topic, draft a preliminary outline for the review of related literature that you will eventually write for your study, highlighting the major sections, or subheadings, of your literature review.

I.
II.
III.
IV.
V.

(A downloadable, interactive version of this developmental activity is available from the Companion Website, www.ablongman.com/mertler6e)

Chapter Summary

When reading research reports, scan them quickly to determine (1) the topic, purpose, and importance of the report, (2) the type of report you are reading—either status, difference between groups, or correlational—and (3) the report's findings and conclusions. When taking notes on material as you read, be sure to include subject, date, author, summary, and complete bibliographical citation.

Your ability to interpret research reports will increase as you familiarize yourself with some of the statistical terms and concepts you are likely to encounter. In status reports of a qualitative nature, you will find verbal descriptions of findings and conclusions, with little if any use of statistics. Few specialized terms will be encountered there. In status reports of a quantitative nature, you are likely to see mention of raw scores, mean, median, mode, standard deviation, grade equivalents, age equivalents, percentiles, and stanines. In reports that focus on differences among or between groups, you are likely to see mention of chi-square, difference between means, and analysis of variance. In correlational reports you are likely to see mention of correlation, coefficient of correlation, positive correlation, and negative correlation.

In reports of correlations and group differences, you are almost certain to see mention of the term significance, which in statistics refers not to importance but to whether the researcher's hypothesis can be retained or rejected at a given probability level, such as .05 or .01.
Having earlier learned to locate published research, you are now familiar with procedures, concepts, and terms that help you read, interpret, and summarize quantities of research. That gives you the background necessary to begin planning a research project of your own. But before continuing to that phase of your development, please strengthen your ability by responding to the activities presented here for thought and discussion.

### Activities for Thought and Discussion

1. Outline the strategy suggested for rapidly assessing the large number of reports you are likely to encounter when reviewing literature on a given topic.
2. List the information you were advised to enter on note cards and where the information should be placed.
3. Why is the concept of significance so often employed in research that explores correlations and differences between groups? If you encountered the following statement: “The coefficient of correlation was significant at the .01 level,” what would it mean to you?
4. Explain why knowledge of statistical terms and concepts enables you to interpret research reports more accurately.
5. Recall the information organizer presented at the beginning of the chapter. See if by referring to it you can summarize the chapter contents.
6. Here again are the questions presented at the beginning of the chapter. See how well you can answer them.
   a. Why should research reports be skimmed rapidly before they are read carefully?
   b. What specifically does one look for when skimming reports?
   c. What entries should one make on note cards when annotating published reports?
   d. What are the three general types of research reports encountered in published literature?
   e. What characterizes the three different types of research reports?
   f. What common analysis terminology—verbal and/or statistical—is usually associated with each of the three types of reports?
   g. Why do the different types of research reports utilize different verbal and statistical procedures to analyze their data?
   h. What two different meanings does the term significant have in the context of research?
   i. Why is the term significant often misleading to uninformed readers of research?
**Answers to Chapter Exercises**

5.1. Discuss responses in class in light of the style manual used at your institution.


5.3. 1. The odds are unacceptably high that the finding does not exist in the population but has appeared because of errors made in selecting the sample.

2. The finding is probably true for the population.

3. A high inverse relationship exists between the two variables.

4. If the study were repeated thousands of times, similar findings would occur at least 95 percent of the time.

5. The finding is probably “real” and not the result of sampling error.

**References and Recommended Readings**


