The first step in preparing to conduct research of your own is to see what others have already done on the topic. There is not much point in doing research on a problem that has been largely resolved. Also, seeing how other researchers have addressed your problem will be of great help to you in deciding the best strategy to use in tackling it yourself. Equally important, your research review can help you avoid unsuccessful approaches.

Although the literature review is the first step, we don't encounter it in this book until the 10th chapter. That is because this preliminary step requires a good deal of prior knowledge. Until one has substantial general background information about research methods and terminology, such as that provided in chapters 1 - 9, it is difficult to read literature with understanding.

The process of reviewing, critiquing, and synthesizing is largely the same whether you plan to stop when you have learned what the state of the art is in research on the problem or whether you intend to do further research on the issue yourself. The general idea is to find, summarize, organize, and apply existing research knowledge to your question. Your problem might be that you have a practical decision to make and you want first to inform yourself about the research relevant to your problem. Or your problem could be more oriented toward further research, as when you plan to write a dissertation. In this chapter I focus more on how to do a literature review when your goal is to get ready to conduct your own research study. The procedures can be somewhat different when you are reviewing research in preparation to make a practical decision (see Haller & Kleine 2001).

The most fundamental point, regardless of the purpose of your literature review, is that a literature review should be taken as seriously as any other kind of research. A literature review should be replicable—in two ways. First, your search techniques should be replicable. To be replicable, they have to be stated. You need to tell your readers what you searched for, where you searched, and how you searched. Another researcher, trying to follow in your footsteps should be able to arrive at the same or a very similar list of research reports. Second, and this is a tougher standard, another researcher using your procedures should come up with the roughly same general conclusions. As with any other research, if two investigators using the same data (research reports in this case) and the same methods of analyzing the data do not draw the same conclusions, the discrepancy raises important questions about the quality of the research.
There are four common ways that writers of literature reviews put together their findings about the individual research reports they have surveyed:

1. The researcher can generate a boring list. This method of exposition essentially involves the author saying "in the first study I read it said . . . , in the second study I read it said...." That is just typing up your notes, and it is never acceptable. No one would seriously advocate a boring list, but that is in fact what many novice reviewers of the literature produce. A list of summaries of research reports needs to be analyzed and synthesized before it qualifies as a review. Without such analysis and synthesis, it is no more than a pile of summaries.

2. The researcher could write an impressionistic summary--perhaps after being inspired by a visit from the muse of literature reviews. This can occasionally be a good approach, but only if you are certain that your work will be brilliant and creative. Absent that level of self-assurance, however, it is better to have a method.

3. The researcher can conduct and construct a systematic review, proceeding with all the care that one would use in organizing and summarizing any other type of data. That fact that one's "data" in literature reviews are research reports, does not give one license to be sloppy. (See the Review of Educational Research for many excellent examples of systematic reviews.)

4. Finally, one can conduct a meta-analysis, also known as a research synthesis. A meta-analysis is essentially a subset of number 3, the systematic review. It is different in that meta-analysis uses some special quantitative techniques for summing up the findings of the research reports.

For obvious reasons, I will only address the final two ways writers of literature reviews present their findings: the systematic review and the meta-analysis. Despite the persistence of the boring list and the impressionistic summary, neither practice leads to reliable literature reviews.

I. How to find research to review

Whether you do a systematic review or a meta-analysis, two preliminary steps are required. First you need to find the research reports to review. Second, you need principles for deciding what to include in and what to exclude from your review. Ideally, you might want to find and review everything relevant. This is sometimes possible in a field that has not been extensively researched, but you will usually have to settle for a selection from a vast field of research reports. If there is a potentially huge body of research literature on your topic, you will need to
define your topic very precisely to limit the scope of your search. A complete review of all relevant literature may still be beyond reach. Even after careful delimitation of your topic, it will often still be necessary to settle for a sample of the best and most recent research reports addressing your research problem.

Your search strategy will vary depending on the phase of your research. When conducted to support further research, which is the focus of this chapter, literature reviews are often done three times. First, you review in an exploratory way to get the lay of the land on your topic and make your decisions about how you will do your study. Such early literature reviews illustrate what is known as the "paradox of inquiry." To know what to look for, you have to have a good conception of your variables and how they are related, but one of the main purposes of looking is to get a better understanding of your variables and how they are related. If you don't start your literature review until you've decided upon your design in all its details, you will be stuck in the same position as the famous fictional fool, Simplissimus, who refused to go into the water until he had first learned how to swim. In research, like everything else, you've got to start by getting your feet wet.

The second literature review is the big one. Once you've decided on what you will study and how you will study it--your question, your design, and your variables--then you are ready to return to the literature for a much more targeted review. You will want to look very closely at the work of researchers who have investigated the same variables and who operationalized their variables the way you have decided to do. In the pages that follow, I will concentrate mostly on the first and second phases of the literature review. These are done before your research is conducted, but there often is a third phase. It is not at all unusual to do a final review of the same and additional sources after you have conducted your study. In that case, the goal is to help you interpret your results. Revisiting the literature can often provide clues to help you interpret a puzzling result. Because this third review is actually quite common, it is useful to keep a photocopy of every article that has played a major role in your literature review. The literature review process tends to be iterative. It is rarely a one-shot task; you usually need to reexamine the research literature on your topic from time to time.

In the first phase of your literature review you should probably begin with secondary sources, such as well-known text books. Although your goal is to do your own review and synthesis of the original research on your topic, it is very useful to see how others have reviewed and summarized it. Start with a few recent intermediate or advanced texts with chapters on your topic. If you find these texts too difficult, and need to work with an elementary text, you may not be ready to do research on the topic. You should try to investigate a topic about which
you are already quite knowledgeable so that you have a pretty good idea about how to frame your research questions. It is usually more effective to build on your strengths rather than to start from scratch.

Go talk to a reference librarian. Usually librarians are pleased to be asked interesting questions and eager to use their considerable skills to help you. Surprisingly perhaps--despite or because of the wonderful resources for electronic searches--librarians have become even more essential in recent years. That is because electronic search tools are in a period of rapid development. Unless you are a complete novice, don't bother to read a book about how to use the electronic resources for searching the research literature. Whatever you read will probably be out of date by the time you read it. It has been my experience for many years that every time I go to a research library, the tools for finding research have been transformed. What I used last semester is not now available, but something better is.

Despite all the change, there are some consistently important sources, although methods of access to them continue to evolve. In educational research one of the most important data bases is ERIC, (Educational Resources Information Center), which is actually a cluster of data bases in which you can search. As good as ERIC sources are, and they are enormously helpful, they are not sufficient for most topics, nor is any other single source. Many topics are researched in disciplinary isolation. Psychologists, sociologists, and political scientists working on highly similar subjects may know surprisingly little about one another's efforts. Generally one has to supplement ERIC searches with searches in three other data sources: PsychINFO, Sociological Abstracts (Sociofile), and Dissertation Abstracts. For most topics in education, if you have searched in these four, you can be comfortable that you've covered the basics. Each of the four is available in various formats: hard copy, CDs, and on-line. Talk to your librarian to learn what is most accessible for you. See chapter 3 in McMillan (2004) for a good introductory review. Finally, be sure also to search for other literature reviews on your topic.

You can sometimes do most of your search for research reports on your subject without going to a library, and you can get a fairly large proportion (not all) of the actual reports on line. This is particularly the case if you have internet access to the library's resources. Sources unavailable to you as an individual may be available through your library, which pays the subscription fee so that patrons can gain access. Indeed, it may someday be possible to be a good education researcher and write a top-notch dissertation without ever entering a library, but in my view, we are not there yet. Even when you can do much of your searching from home, you can usually do some aspects of it better on the spot. More important, libraries are also a good place to meander in your field, to skim books and to thumb through recent numbers of relevant research journals. You will probably discover that two or three journals
publish many of the articles on your topic. If so, it is often very helpful to check the contents of those journals, especially the recent issues that may not yet have been indexed in data bases. Much of this wandering can be done virtually, electronically. But not all of it. In my experience, even after a thorough electronic search, one can gain new insights and understanding of a field by "hiking" around in it. Exploring electronically is like looking at a field from a spy satellite. Even at the highest magnifications, it just doesn't "feel" the same as being on the ground and getting dust on your shoes.

Ultimately your preferred methods (more than one is usually needed) of searching are a matter of taste. But, however you meander, physically or electronically, leave a trail of bread crumbs so you can find your way back to the interesting things. Keep a record of what you see. Enter the field like a naturalist. Carry a lab notebook, a diary, a journal--or whatever you want to call your record of what you've done. You will probably find that from time to time it will be very helpful to be able to reconstruct your travels through the literature; this happens more frequently than you might imagine. And, as the old Chinese proverb has it, the strongest memory is weaker than the palest ink. And don't forget to hunt for researchers, not only for the reports they write. Send e-mails to authors of studies you find interesting. A surprisingly high percentage will respond and be glad to help. Ultimately, there is no other way to get up-to-date on your topic but to talk to the people who are actively engaged in doing research on it.

Despite the potentially great advantages of searching for sources on the shelves of a well-stocked research library, electronic searches are absolutely crucial. Hardcopy versions of most good data sources exist, but for most searches they are dramatically inferior to the electronic versions. Many kinds of topics can only be effectively searched electronically--because of the word "and." The connector "and" allows you to specify what you are looking for with increasingly precise descriptors. For example, you might be looking for information about college students, (and) who are female, (and) who attended four-year colleges, (and) who received financial aid, (and) who majored in engineering, and so on. Each additional descriptor narrows your search and focuses it more precisely on what you are really looking for; it enables you to use multivariate criteria for defining your search. Another handy search tool for focusing on your search is "not," as in: education, not elementary. The word "or" is the opposite of "and" in searching. While "and" gives you the narrowest search, "or" gives you the broadest, as in: schools or colleges.

Once in a while you will have the opposite problem: your focus will be too narrow. After running your topic through the various ERIC, Dissertation Abstracts, PsychINFO, and Sociofile resources, you won't be able to
find research on your topic. This is especially the case if your object of research is new—or if it has a new name. We are especially good in education at dreaming up new names for old concepts. We have cellars full of old wine, in old bottles, but with new labels. The most effective way to handle a search in which you find that there is "no research" on a topic is to think in terms of variables, not their labels. The new label usually has an older name, or the concept can be described more generally, in which case you'll probably find ample research on it. It is a rare topic indeed that has not been investigated. There is a real art to discovering the right generic descriptor (rather than a specific brand name) for a variable. Fortunately, most data bases include a thesaurus to help you find the terms you need to search effectively. Again, my advice is to talk to a reference librarian if you need to find and to learn how to use the thesaurus for your data base. Even modest research libraries will often have a reference librarian who specializes in your field and who can help you identify the key search terms for your topic.

I have talked in these pages only about searching in the scholarly sources. But general internet search engines can also be very helpful. Of course, the internet is awash in trash—everything from pornography vendors to political extremists, to say nothing of pictures from my family reunion. Like every other new information technology since the invention of printing, the internet can be used for good or ill, or just for fun. Truly crucial sources for educational research are available in the web sites of government agencies (see chapter 4 of McMillan, 2004 for an overview). And, increasingly, this information is available only online. My favorite of these government sources is the site for the National Center for Educational Statistics (nces.ed.gov); it makes available huge amounts of very high quality data. Most states also have web sites where they provide large amounts of the most up-to-date information. Also, an increasing number of research journals are available only online. Finally, it is often useful and fun to use commercial search engines—such as Google or Yahoo—to investigate your topic. It is particularly instructive to search using the same key words in several search engines. One often finds remarkably little overlap on the "same" subject.

II. What to look for when reviewing a research report

To read, critically analyze, and then synthesize the research literature you have gathered, it is often useful to use a checklist of questions to ask of every research report you examine. There are too many things to cover to trust your memory to think of checking them all. An important bonus of this approach is that after you've gone
through your checklist with each research report, and you reach the point where you want to synthesize what you have found, a checklist gives you a way to organize your notes according the answers to a common set of questions.

Here follows the checklist I use. I use it quite specifically as a checklist, as a guide to make sure I don't forget anything important. Every question on the list is not equally pertinent for every article or for every topic. If you use the checklist, you may need to supplement some questions or add new ones for your particular purposes, but this list is a good place to start. To repeat, answering a checklist of questions almost automatically produces a preliminary organization for your notes. This organization will make it easy for you to pull together all of your information about mediating variables (question 9), or sample size (question 17), or statistical significance (question 20), and so on. Being able to organize the findings of your literature review is the key to building your research upon it. I also find it useful to use the same checklist to question myself about my own work.

Table 10.1 presents the checklist. Following the table I briefly discuss some points raised by the questions on it. Since a good checklist will naturally try to ask about all important things, discussing the checklist in the table will involve raising, if only briefly, most major issues in research methods.

**Insert Table 10.1 (from end of chapter) about here**

### A. General questions

1. *What is the hypothesis or the research question, if any, guiding the research?*

   Answering this question first can save you a lot of time. It tells you whether this article should be on your list of those that you need to review. Sometimes it is not clear what the authors of the article were investigating. If you cannot figure it out, it could be because you lack the background, but it also could be because the authors themselves do not have a clear conception. Not all published work is valuable.

   Note that "research questions" or statements of research objectives have largely replaced formal hypotheses (and especially null hypotheses) in introductions to research reports. The difference is mostly a matter of style. Conventions vary by discipline, author, and time period. What does not vary is the reader's need for a clear understanding of the purpose of the research.
2. *Why do the authors believe it is important to investigate this hypothesis/question?*

In addition to situating the research in the context of the problems it addresses, answering this question can help you determine whether an article truly is relevant to the research you plan. Articles ostensibly about the same subject might not really be about your topic or issue. If you are interested in the effects of academic achievement on self-esteem, your search could easily turn up research on the effects of self-esteem on achievement—clearly a different topic, in which you might or might not be interested.

3. *What methods did the authors use to collect evidence? What was their design?*

This question needs to be asked at a couple of levels. Most generally, what research design, broadly speaking, was used? Was it an experimental study, an ethnographic study, a sample survey, a secondary analysis of data, a document analysis, etc.? Within each of those categories you will want to specify further. For example, what kind of experiment was conducted, a true or a quasi-experiment, a within-groups or between-groups experiment, etc.?

4. *Were the methods appropriate to address this problem/question?*

Even if the method was well-executed, that still leaves the question of whether it was the right sort of approach for the particular problem studied. For example, if a survey was used, might interviews or focus groups have been more likely to provide evidence pertinent to the answering the authors’ questions?

5. *What are the main findings or conclusions of the article?*

What do the authors say they learned in their study? Are the findings clearly and concisely stated? If so, this will be of great help to you. The art of reviewing literature is in part the art of summarizing. You will typically review dozens of studies, but you will not want your summaries of conclusions to run to dozens of pages. So you need to be able to give the nub of the matter in less than a page. Sometimes the authors will help you by providing a concise account of their conclusions (usually in the abstract). Often you will have to construct that summary yourself—sometimes because your interests might be different than those of the report's authors. You may want to make different use of their work than they have.
6. Are the conclusions convincing?

You will want to answer this in more than one way. First you will want to record your overall "feeling" for the persuasiveness of the article. Most of the rest of the questions on this checklist help you to explain why you might or might not find the conclusions convincing. But even though you will explain your assessment of the article via your answers to the following questions, you are also likely to have a more visceral and less analytical first reaction to it. It is good to record this reaction. One of the most common sources of a good dissertation topic is a student's annoyance with a well-known article in an important journal. Irritation can motivate the reader to look closely very closely at the research. When a research report gives you pleasure, analyze those reactions too; they help you discover what you value in research. First, trust your instincts. Then, scrutinize them. There is nothing wrong with starting with gut feelings--as long as you use them to give you the energy to engage your mind.

B. Questions about the variables

7. What is the dependent or outcome variable (OV)?

The answer to this question is often a very straightforward and clear one. If it is not, that could be because the article is descriptive and does not have an outcome (or dependent) variable properly speaking. As with question number 1 (what is the hypothesis/question?), answering this one helps you focus your attention sharply on whether a given research report is really one you want to analyze extensively. If you want to study, for example, the causes of dropping out of high school, you'll want to review articles in which dropping out is the outcome variable. However, you probably don't want to review articles on the consequences of dropping out, in which case dropping out would be the predictor (or independent) variable.

8. What are the independent or predictor variables (PVs)?

In addition to recording the PVs studied in the research report, you have to decide about the ones that you are really interested in. It is helpful to make an inventory of all the PVs (or independent variables) that have been suggested as related to your outcome variable (OV). Then you need to decide how to handle this list. If your OV is dropping out of school or college, do you want to look for any and all possible factors that lead to dropping out? Or are you more focussed and only want to look at the effects of one particular variable on dropping out? If so, all the other predictors on your list could still be important, but as control variables. You need to look at variables in which
you aren't interested so as to be able to control for them, to "eliminate them as suspects," as the police say; they are alternative explanations for your findings.

9. Are any mediating or intervening variables identified?

If a mediating variable is identified, this is often the main point of the article. Mediating variables are the links between predictors (PVs) and outcomes (OVs). In our earlier examples, we saw that arthritis was linked to reduced heart attacks by aspirin, which was the mediator. Small schools were linked to higher student achievement by improved teacher attitudes; the attitudes mediated the effect of school size. The mediator is the message in many research reports.

10. Should these or any other mediating/intervening variables have been studied?

Sometimes your critical analysis of a research report will lead you to believe you have discovered an important potential mediator that has been omitted. Such discoveries are one of the ways in which your literature review becomes active and creative. An omitted mediator may become the variable in which you take a special interest and want to study. Often you will discover mediators of interest by reflecting on the possible relationships among the variables in your inventory of all predictors (see question number 8).

11. Are any control variables considered?

First you need to note what the control variables are in each study. Then, as with the predictor (independent) variables, it is helpful to make a list of all the control variables used in the studies you review. Then you can compare the variables controlled in the studies. This can often enable you understand seemingly discrepant findings. If one study shows a significant relation between X and Y and a second study does not, this could be because the second study has controlled for a variable omitted in the first. Also, because another researcher has used a variable as a control, it does not follow that you must. Indeed, you could use that variable as a predictor, in which case information (e.g., a regression coefficient) about another author's control variable is directly relevant to your predictor variable.
12. Should other control variables have been examined?

It is almost always possible to think of other control variables that could have or should have been examined. When an article suggests that one variable influences another, it is usually possible to say, "Yes, but isn't that just because . . . " What follows "because" in the sentence is a control variable. Yes, the students in the new program learned more, but isn't that just because their teachers were enthusiastic volunteers? Often the greatest weakness of a non-experimental study is that the authors have not considered what seems a crucial control variable. The same can be true of experimental studies. We hope that random assignment to experimental and control groups will automatically equalize all potential control variables, including those we don't know about. And, if the sample is fairly large, we can usually count on that form of control to do the job. But if the sample is not large and/or if it is an unrepresentative convenience sample, control variables can be important in experiments too. It is in this circumstance that pre-tests can be so important in experimental research (see Chapter 6).

13. Does the article discuss possible moderating variables and interaction effects? Should it?

Gender, age, and ethnicity are the most common moderators in educational research. Just as when there are obvious control variables that ought to have been considered, the authors should explain what they have done about obvious but omitted moderating variables. As we have seen (in Chapter 9), the statistical examination of a moderating variable and of a control variable generally involves the same technical analysis procedure, such as including it in a regression equation. For example, if we study the effects of participation in an educational program on student learning, do we get the same results before and after controlling for student gender? If not, student gender may moderate the relationship between program participation and student learning.

14. How are the variables defined and measured, i.e., how are they operationalized?

Very often studies presumably about the same thing define and measure that thing so differently that really there is no reason for thinking that the same thing is being studied. A difference in the findings of two studies may stem from the fact that they have measured their variables differently. Say one study measures the effects of parental SES on student achievement by looking at parents' education level and students' grades. Another defines SES as income and achievement as SAT scores. Are they really measuring the same variables and the same relationship? Even if parents' education level and income tend to be fairly well correlated, they are not the same
thing, nor are students' grades and SAT scores. Each may be an indicator of the same underlying variable. Perhaps the two measures of parental SES and the two measures of student achievement should be combined. Or if one is your target predictor variable, the other should be controlled. For example, what is the effect of parents' education level after controlling for their household income (or vice versa)?

15. Are the definitions and measurements of the variables appropriate for this study?

Depending on the purposes of a study, different operationalizations of variables might be appropriate. You will at least want to be able to categorize the studies you use in your literature review according to how they made these key decisions. To understand the works in your literature review you need to know not only how the variables were labeled, e.g., student achievement, but also how variables like student achievement were defined. It surely matters whether student achievement was measured by class rank or GPA or ACT scores or "met standards" on a state test. Each of these is an indicator of student achievement, but each measures it differently. If a conclusion holds even when several different indicators are used, this does much to substantiate the conclusion. But if it differs when different indicators are used, this provides continuing challenges for researchers.

C. Questions about the sample/subjects

16. Who is studied and are the subjects appropriate given the goals of the study?

The most accessible subjects are not always the most relevant. For example, if the subject is attitudes toward the risks of unemployment and the subjects are college students, the mismatch may be so great that the study is of little use.

17. How many are studied and is this enough for the purposes of the study?

Is the sample big enough for generalization, for statistical power, for statistical significance? Often these are technical questions, but sometimes less so. Do we really want to decide about how all high schools function on the basis of a study of 12 high schools?
18. *Is the sample representative of a population? To whom might the conclusions apply?*

How broadly can the conclusions be generalized? This is to point to which questions 16 and 17 are leading you. Were enough of the right sorts of people or schools or cities studied that you can be confident about the applicability of the conclusions? Even if appropriate subjects were studied, and even if many of them were studied, if they aren't a representative sample, one should be wary of using information about them to generalize beyond the sample to a population.

**D. Questions about the conclusions**

19. *Are the findings statistically significant?*

Unless the authors have made an arithmetic error, this is just a matter of reading off the p values. But there is still some judgement involved. Some authors will use $p < .10$ others $p < .05$, and others $p < .01$ as the cut-off at which “significance” is declared. Although .05 is the conventional level, it is just a convention and you may wish to judge whether it is appropriate (either too strict or not strict enough) in each case.

20. *Are the findings scientifically significant?*

Do the conclusions make a genuine contribution to the knowledge that other researchers might use to further their work? Note that this is a separate question from the issue of whether the conclusions can be directly applied to a practical problem. Most research articles are cluttered with the word "significant," especially in their concluding paragraphs. Readers need to use great caution when interpreting this adjective.

21. *How big are the effects discovered?*

You want to compare the size of outcomes as well as their statistical and scientific significance. Statistically significant effects can sometimes be quite small. Measures of "size" include measures of association such as correlation coefficients and the widely used effect size statistic based on standardized mean differences. But "big" and "small" are relative. Reducing the incidence of some problem in a state by a mere 1% may mean saving thousands of students from its effects.
22. Are the findings practically significant?

Practical significance, like size and scientific significance, is also a matter of judgement. Practical significance tends to be determined by the size of effects, but whether a particular size of effect is big "enough" to make an "important" difference is not a statistical question but is one of judgement about the context in which the conclusion is applied. An effect size (ES) of 0.1 for the effects of an inexpensive cancer treatment can be very important practically, since treating cancer is a matter of life or death. On the other hand, an ES of 0.1 for an expensive SAT coaching course would be considered trivial by knowledgeable people.

23. Are the conclusions really supported by the evidence cited in the article?

Readers who have carefully examined the evidence as presented by the authors are sometimes shocked to find what the authors have had the audacity to conclude. Even if the authors get a little exuberant in the concluding paragraphs, good authors of good articles in good journals give the reader the data and the means to check their work, to judge the evidence to see whether it supports the conclusions. The whole reason we go through the above questions about operationalizing variables, representative samples, effect size, and the rest of it is to be able to check. This is also why you need considerable background before you can do your literature review. Otherwise, as a reviewer, you might be limited to merely reading the discussion section where authors tend to speculate freely and use overblown adjectives. A literature review that skips the evidence is not sufficient as a basis for further research.

E. Finally and implied in the answers to the above questions

24. How could the research have been improved?

If your literature review is providing the context for your research, this question is key. If the study is already perfect there isn't much point in you continuing. But no study is perfect—at least I have not encountered one. No study can examine all aspects of a question. The authors of research reports will often, and should usually, discuss the limitations of their work. When the authors are candid, descriptions of research limitations provide you with important guidelines about what needs to be done in your field.
25. *What questions or problems does the article leave unanswered?*

Again, this provides you with direction for your research. The imperfections could merely be that the study needs to be extended. The article could be excellent, but you might wish to see if the conclusions hold in different settings or with different subjects.

26. *How would you go about doing a better job?*

The relevance of this question is obvious. Answer it, and you have the nub of your research design. As with limitations, many authors give you their assessment of what research needs to be done in the future. In fact, it is quite rare to read a research report that does not end with a call for further research. Some of these statements are ritualistic, but others offer serious advice about research needs on a topic.

**Conclusion to checklist review**

After having answered these questions about the research reports you have read, you will be in a good position to come to strong conclusions, even if some of the articles disagree, as they almost surely will. In the first place you should be clear what they agree and disagree about. In the second, you will be ready to make a systematic judgement about which side of the debate you want to come down on. Third, you will also have extensive of evidence about why studies might come to different conclusions. Perhaps they used different methods or operationalized variables differently or used different controls. Perhaps they studied different settings or kinds of subjects. Say you’ve reviewed 25 studies and 16 of them have claimed that X is associated with Y, but 9 say that it isn’t. Have they operationalized variables differently--e.g., is academic achievement defined as grades or scores? Or have they included different control variables--e.g., is the effect different in those studies that controlled for SES? Or did they study different subjects (college students, males, etc.)? Or did they use different designs (experimental versus observational)? Or were the studies conducted in different settings (public versus private schools)? You will be in a position to observe these things, even before you have undertaken any of your own research. You will have an overview of research on the topic. You will be quite likely to know something *new* before you even begin to gather data. The authors of individual studies will often not have your broader perspective. In brief, you should finish your literature review as the possessor of a rare kind of expertise.
III. What to include (and exclude) and how to write it up

Reading all of the research reports you unearth is often a formidable task. Writing a review of that research can be equally daunting. For one thing, it is not always easy to distinguish between research reports that provide interesting background for your project and those that are really essential to your research. You will almost certainly have read and taken careful notes on many works that do not make it into your final review. Cutting these irrelevancies, which may represent days of work, is painful. Remember that the goal of the review is to establish the rationale for your research, not to show how much homework you have done. Good reviewing means summarizing and eliminating--data reduction, if you will. But save your notes! You may find that after you have done your research you need to revisit some of this earlier work.

Writing up your literature review can often be the most conceptually demanding part of your research project. Doing it well requires that you have a good grasp of the entire field and the methods used to study it. You need to have mastered your sources if you are to be able to do the kind of radical summarizing that is necessary. Most of my students in research classes have difficulty when I limit them to 1,000 words (about 4 typed pages) to summarize one article. A good literature review will often discuss 50 articles and other sources. At 4 pages each, that would get you to 200 pages. You don't want to write, and no one wants to read, a 200-page literature review. To twist a famous quotation from Pascal, a noted 17th century philosopher and contributor to statistical theory, "Excuse me for writing such a long lit review; I did not have the time to write a short one." If you have devoted sufficient time to master your subject, you can often write a comparatively short research review. The more drafts you write the shorter they will get.

One way to approach the writing-up problem is to think of your research review as a history of your subject. Professional researchers often see their work this way; they think of their investigations very explicitly in terms of what has gone on before. Your literature review should help you see your work that way too. You will often encounter this historical approach in your reading and read phrases such as the following: "the problem was originally understood as… then, with advances in the field, we added… what is new about the current research is that we have …." One reason for the popularity of this historical approach is that it can leave the reader eager to read the study. After having been given this "running start" in the past, the reader is ready to jump into the future. Substantively and rhetorically the whole point of the literature review is to prepare first yourself and then your readers for what is to come.
Perhaps the best way to learn to write research literature reviews is to reflect on those you read. As you read research reports you will inevitably be reading literature reviews too—generally very short ones. Try to learn from these examples. The *Review of Educational Research* is a journal devoted exclusively to research reviews. Among the techniques that can be appreciated by reading some examples in that journal is the excellent use that can be made of tables to summarize findings. Creswell (2003, pp. 39-41) recommends constructing a physical "map" of the literature that will enable you to picture the relations among the research reports on your topic. Constructing such a map is more than an exercise to help you with your thinking. You may well want to use it when writing up your final review and to help your readers with their thinking about your topic.

As should be obvious from all that has been said above, writing your review using only the discussion sections of the research reports is insufficient. Especially when reviewing quantitative research, use numbers, not adjectives, to summarize the results. Numbers are more precise and take up less space. Knowing that two studies found a significant correlation between SAT scores and GPAs is one thing. Knowing that one found a correlation of .30 and the other of .50 tells you much more. Give the effect sizes, the correlation coefficients, and the mean differences. Don't just say "big," say how big. Don't merely say that a result was statistically significant, tell the reader at which level it was significant and, if available, give the actual $p$-values.

If you write your review well, you will be leading the readers to see the research problem as you do. When they move to the methodology section of your paper or thesis, the readers will not be surprised at how you have decided to investigate your topic. They will, in fact be expecting your approach, because your review should have led them to expect it, to see the research question in your terms. This is a difficult standard, but if you can attain it, it indicates true mastery of your subject.

Sometimes your research review is so important that it becomes its own end, not merely a lead-in to another research project. This is especially the case for research reviews that employ the techniques of meta-analysis.

**IV. Meta-Analysis or Research Synthesis**

Everything I have said above about systematic literature reviews applies—perhaps at a somewhat heightened level of intensity—to meta-analysis. Meta-analysis might better be called “research synthesis,” and it is thus called by many researchers, especially in biomedical research. Whatever one calls it, the basic idea is to
develop quantitative summaries of the findings of quantitative research reports. In addition to calculating an effect size for each individual research finding, one can calculate an overall average effect size for all research findings. Doing this adds a new dimension to reviewing and summarizing research, but it does not change its nature fundamentally. Apart from the technical issues concerning how to do the calculations, the important questions remain the same as in any review of the research literature. And the technical issues involved with meta-analysis are not highly complicated. Meta-analysis does not many use new techniques; rather, well-known measures are applied to new kinds of investigations—in which research reports are the unit of analysis. Readers wishing to conduct a meta-analysis will need more detail than is provided here. My students' two favorite sources for the needed details are by Light & Pillemer (1984) and Lipsey & Wilson (2001).

Meta-analysis was independently invented, at about the same time (1970s), by Robert Rosenthal and Gene Glass. Its techniques have developed rapidly since that time. Among the most important consequences of the emergence of meta-analysis has been its effect on all research reviews, whether or not they use the techniques of meta-analysis. The presence of meta-analysis has ratcheted up standards for all literature reviews. In the past, researchers too often treated the research studies they were reviewing in ways they would have been ashamed to treat any other kind of evidence. Today, most researchers would agree that research reviews should be conducted with as much care as, and they "should be just as replicable as any other piece of scientific work" (Cook et al., 1992, p. viii).

One useful way to think of meta-analysis is to see it as "a form of survey research in which research reports, rather than people, are surveyed" (Lipsey & Wilson, 2001, p. 1). Extending this analogy, traditional reviews, those that do not employ quantitative means of summarizing findings, are more like interviews. Just as it would be silly to say surveys are the "right" way and interviews are the "wrong" way to obtain evidence, so too with narrative and quantitative summaries of research. Each has advantages and disadvantages. Surveys, like meta-analyses, gather the same information from a large number of respondents with the goal of being able to summarize their responses quantitatively. They tend to be strong on external validity or generalizability. Interviews, by contrast, tend to probe in more depth with fewer informants and in ways that can vary from informant to informant. They tend to be better at exploring individuals' meanings and being certain that the respondents' meanings are clear. This can be thought of as a kind of internal validity. But because interviews inevitably use smaller, and usually unrepresentative, samples, they tend to be weak on generalizability. Deciding on the best approach for your
research review is as complicated as any other choice of method for a research project and always involves tradeoffs. Are you going to survey your sources or will you interview them? I usually advocate the kind of mixed-method approach advocated by Light & Pillemer (1984).

Should you do a meta-analysis for your research review? If you are conducting a literature review and it is not a meta-analysis, you probably need to explain why. It is the state of the art in literature reviews. I would even go so far as to say that the burden of proof is on you to show that a meta-analysis would not be better, not only than a traditional literature review, but than the small-scale study you are planning. In short, if you are not doing a meta-analysis, you should have a good reason.

What are some reasons you would not do a meta-analysis? One is that another meta-analysis has recently been done on this topic. The readily analyzable topics tend to get snapped up, in part because meta-analysis is a powerful technique and using it requires no human subjects review. Probably the main reason not to do a meta-analysis is that the number of quantitative studies on the topic is insufficient. Or, even if an apparently adequate number of studies exists, they might not be appropriate for one reason or another. For example, it is not uncommon for research reviews to conclude that studies of a subject are so flawed that is it not possible to come to any clear conclusions (see Mattingly et al., 2002 for an example). A final reason not to do a meta-analysis, of course, is that if everyone only did meta-analyses, we’d eventually run out of works to synthesize. We could delay the inevitable by doing meta-meta-analyses, but ultimately, someone will have to undertake the original research studies that are the foundation on which we build.

My estimate is that for most of the topics that doctoral students in education consider, well over three-quarters will not be suitable for a full meta-analysis. Usually, there will not be enough studies investigating the same variables, measured the same way, and researched with the same methods. However, within a broader review of the research literature, more limited meta-analytic work--what we might call mini-meta-analyses--on a subset of research reports can often be conducted. Within a general review of several dozen reports, one is likely to find one or more clusters of a few studies for which it is possible to pool data in ways that meet the requirements of meta-analysis. When that kind of opportunity presents itself, be pragmatic and use the tools. Looking at things the other way around, I would also recommend a similar sort of pragmatism when conducting full meta-analytic studies. In the course of a meta-analysis, one is very likely to find studies not appropriate for statistical summary. But these should not merely be discarded. If they contain relevant evidence, the fact that they will not fit into the statistical
summary should not lead to their disqualification. For example, studies using multivariate regression are extremely
difficult to pool. But skipping them would seem ill-advised. Also studies generating more qualitative evidence can
often be included in a review, even when they cannot directly contribute to the quantitative summary.

Narratives versus numbers in research reviews

Not surprisingly, the ideological posturings of the quant/qual battle occur too in the realm of research
reviews. As elsewhere, here too the quant/qual distinction obscures more than it clarifies. Some reviews are better
done with narratives, some with numbers, and many can benefit from various combinations of the two. Nonetheless,
proponents and opponents of meta-analysis have sometimes turned their disputes into another episode in clash
between the Quants and the Quals. However, as Light & Pillemer (1984) put it, "the pursuit of good science should
transcend personal preferences for numbers or narrative" (p. 143).

The truly heated polemics between the antagonists and protagonists of meta-analysis have cooled since the
1980s. Extreme accusations on both sides have died down. One of the most common accusations was that meta-
analysts think that they can take bad studies and, merely by adding them up, produce good ones. This is the famous
GIGO criticism: garbage-in-garbage-out. No one would claim that meta-analysis can turn garbage into gold. In my
reading, authors of the typical meta-analysis are very careful to cull out the garbage. Often huge searches reviewing
hundreds of research reports yield only a few dozen that are appropriate for synthesis. GIGO is certainly true, but it
applies as much to narrative reviews as to numerical ones. Generally, meta-analysts make their selection criteria
explicit; this enables readers to decide whether or not what has been synthesized is garbage. In general, bad
practice in meta-analysis is bad practice in any kind of research review.1

The meta-analytic retort to such criticism has often been to claim that traditional narrative reviews turn
gold into garbage. Good studies are thrown together in ways that mostly reveal the biases of the reviewers.
Research reviews in which reports are selected and summarized according to no clear plan or are chosen and
discussed as dictated by the biases of reviewers is the origin of much garbage. This also is a good criticism. Of
course, not many people would seriously argue for slanted reviews. Guarding against bias is as important in meta-
analyses as it is in traditional narrative reviews.
Comparing apples, oranges, and clones

What should you include in your review of the research? What can reasonably be summarized under the same rubric? These are problems for all research reviews, but are made more explicit, and thus more controversial, in meta-analysis. Some answers to these questions may to stem from one's disposition. The research world seems divided between "lumpers" and "splitters." Splitters are fond of saying that you can't compare apples and oranges. To do so ignores important distinctions. Lumpers might say of course you can: apples and oranges are both fruits that grow on trees, unlike strawberries; both are rounder than plums and firmer than peaches. If apples and oranges were exactly alike there would be no point in comparing them; there is no point in comparing clones.

The lump/split distinction has parallels with the quant/qual divide. Advocates of qualitative measurement, given their stress on individual distinctions, tend to be splitters. Advocates of quantitative measurement, with their emphasis on variables that transcend individual examples, are more likely to be lumpers. As usual, these distinctions are overdrawn. Narrative reviewers can sometimes bring together (lump) a broader range of studies--precisely because narrative reviews can more easily accommodate different operationalizations. And numerical reviewers often spend much time on the narrative descriptions of design; only in that way can they be sure they are summarizing comparable studies.

Most broadly, the question is: how much detail can you to sacrifice in order to make a generalization? If you aren't willing to sacrifice any detail, to ignore any differences, you can't generalize at all. Inevitably, generalizations are only generally true--which is why we call them generalizations. On the other hand, if you overlook crucial differences, your generalizations will be foolish. Finding a reasonable middle ground requires much wisdom. The most reasonable answer to the question, "how broadly can we generalize?" is usually a pragmatic one. Does the generalizing help you learn what you are trying to learn? Is your research question best addressed by lumping or splitting? Do you want to study a general variable such as academic achievement? If so, it might best be measured by combining indicators such as grades, class rank, and scores on standardized tests. Or, are you interested in the distinctions between aspects of achievement? In that case you would rightly resist lumping them.

1 See Chapter 13 of Shadish, Cook, & Campbell (2002) for a systematic review of the advantages of meta-analysis
Populations and samples

Once you have decided on your general definitions of variables, on which predictors and outcomes you want to study and how you want to measure them, how many studies do you need to include to do a meta-analysis? If your population is all the research done on a topic, what should your sample be?

The most typical answer by meta-analysts is: use a "total sample;" include everything at least initially. Begin with the broadest sample possible. Be relentless in your search. Leave no stone unturned. Meta-analysts point out that many of the studies you find will eventually have to be excluded because they do not meet technical criteria, usually because they do not provide enough information to enable quantitative summary, so you had better begin by a very thorough and inclusive search.

One pragmatic way to answer to the question, "how many is enough?" has been suggested by Rosenthal (1979). After you have done an initial meta-analysis on the sources you first found, it is possible to calculate the number of research reports with contradictory conclusions that you would have to find before you would have to alter your summary conclusion. If you would have to find a very large number of such studies, and if you have done a thorough search, then your sample size is probably sufficient and your conclusions are probably sound.

The emphasis on thoroughness in searching has led most meta-analysts to advocate hunting extensively for unpublished as well as published research reports. The rationale for this is that published sources may be a biased sample of all the research on the subject. This concern has been supported by some studies that have indicated that research reporting statistically significant findings are more likely to get published. If that is true, then otherwise good research where no significant difference among variables was found, where no significant difference between groups could be discerned (in other terms, where the authors were unable to reject the null hypothesis), would be underrepresented in the population of studies.

There are actually two or three categories of unpublished research. The first category is made up of work that, while unpublished, is publically available. We might call these "quasi-published." This would include, for example, dissertations indexed in Dissertation Abstracts and conference papers indexed in ERIC. I think meta-analysts are right to be systematic in their search for these. The arguments for pursuing other categories of unpublished research, truly unpublished research, strike me as less persuasive. The authors of unpublished research may not have believed their work to be interesting enough to submit for publication. I see no reason to challenge over traditional reviews. For a history of the early debates, see Hunt (1997).
their self-assessment. Or, unpublished works could have been submitted for publication by their authors, but rejected by publishers. Such rejected work could have been rejected because of some sort of publication bias, but it also could have been rejected because it was of low quality. My work as an editor and reviewer for journals makes me think that the latter explanation is much more likely.

While a large number of good-quality unpublished studies on a topic may exist and may differ systematically from the published ones, the possibility strikes me as quite remote. Most meta-analysts disagree with me. Their argument is that a valid survey of a population cannot leave out part of the population. You can sample from the whole population (though meta-analysts seem reluctant to do this too), but even the most technically correct sampling from an unrepresentative part of the population will not yield a representative sample.

True enough, but what is the "population" in a meta-analysis? It is all the research reports on a subject. Is excluding the unpublished part of the population of research reports really parallel to excluding parts of other populations? There seem important differences. Surely if we were surveying all teachers and omitted those who taught in private schools, we would be making a mistake. That is because private school teachers are a knowable part of the population of all teachers. But the population of unpublished studies is in principle an unknown quantity. How many unpublished studies are there? How could anyone possibly know? Assertions about the unknowable are risky. One might rightly assume (not know) that more such privately conducted research, which is available only in the authors' file drawers, would be more common in nations with repressive political regimes. Or extensive unpublished research could exist on highly controversial topics for which limited publications outlets exist. While this seems logical, direct evidence is meager.

There is another difference, and a more important one, between the population of research reports and other populations, such as the population of all teachers, or all cities above a certain size, or all hospital patients. Even if one obtains all research reports--public, private, published, quasi-published, etc.--about such easily known populations, one does not necessarily have an unbiased population. One cannot necessarily generalize from this population to "the truth" about a subject. This population can be biased in the sense that the researchers who wrote the research reports can be biased. The population is what happens to have been studied about a topic. It does not represent the topic; it represents researchers' work on the topic. Indeed, the literature review is often designed as much to find parts of a topic that have not been studied as well as to sum up what is already known. For example, in the not-too-distant past, public opinion research in the U.S. routinely excluded blacks. Medical research on some
conditions did not include women in the subject pool. A total population of this research would be a biased representation of the potential domain of knowledge in the field. These issues should usually be addressed in the limitations section. If you take a representative sample of the population of studies, but, if most or all of the samples of subjects are non-representative convenience samples, your sample is representative of the population of studies, but not of any population of subjects.

To take another example of the difficulty of identifying or conducting a census of "the whole" population, do you consider studies reported in languages other than English? If your research review excludes these, the bias is potentially very large (much larger, perhaps than if it excludes unpublished work). Scholars once learned languages to get access to sources, but this seems pretty rare these days.

Given the near impossibility assembling or even knowing the "true" population of research reports on a subject, what is to be done? In my view it is best simply to say that the "population" for the meta-analysis is the publically available research on a topic (this would include unpublished works such as dissertations and papers presented at conferences, which are "quasi-published"). In the limitations section one would then have to admit that this might omit something important. In short, I would define the population of research for the purposes of meta-analysis as the publically available research reports on a topic written in languages I know how to read. Unless one has some strong independent reason to believe that a treasure trove of unpublished material exists, that should suffice. It will have to do. Now that meta-analysis is a fully accepted research technique, researchers no longer have to pursue fugitive literature well beyond the point of diminishing returns.

Once we have settled on the population, realizing that it will imperfectly coincide with the true population, and that it will imperfectly represent the potential domain of knowledge in the field, should we do a census of the whole population, however defined? Or should one sample from it? That depends on many things, but one point can be stated with certainty here: there is no point in using inferential statistics unless one has sampled from a known population.

Quantitative techniques

So once you've decided all the tricky questions about populations and samples, what are the methods for quantitatively summarizing the results of research? As always in this book, we will concentrate on general
principles rather than the technical details. While the reader actually wanting to conduct a meta-analysis will have to consult a guide, such as Lipsey & Wilson (2001), we can cover many of the basics here.

The key statistics for meta-analysis are various effect size (ES) measures. Indeed, the early history of meta-analysis, in the 1970s and 80s, overlaps that of the great effect size debate in psychological and educational research. Many of the protagonists were the same (Rosenthal, Cohen, Glass) who argued that not only was it important to identify whether a result was statistically significant, but the size of the effect mattered as much if not more. Confidence intervals were also important in this debate, but at its vortex were effect size statistics. As we have seen, effect sizes are standardized measures that can be used to summarize and compare quantitative results across studies. Many textbooks talk about effect size (ES) as though there were only one effect size statistic. In fact, several such statistics are available for the use of someone undertaking a meta-analysis. An even broader range is available for general purposes (Kirk 1996). The ES measure usually referred to as the effect size is the "standardized mean difference." As we have seen, to compute this you subtract the mean of the control group from the mean of the experimental group and divide a standard deviation--either of the scores of the control group or the scores of the control and experimental group combined (Lipsey & Wilson 2001.) The other two effect size statistics used most extensively in meta-analysis are the Pearson $r$ correlation, discussed in chapter 2, and the odds ratio, discussed in chapters 12 and 16. Table 10.2 lists the three most common ES measures used in meta-analysis and the types of variables for which they can be used.

**Table 10.2. Meta-analysis techniques and measurement levels of variables**

<table>
<thead>
<tr>
<th>Analysis Technique</th>
<th>Predictor (independent) variables</th>
<th>Outcome (dependent) variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds ratio</td>
<td>Categorical</td>
<td>Categorical</td>
</tr>
<tr>
<td>Standardized mean difference</td>
<td>Categorical</td>
<td>Continuous</td>
</tr>
<tr>
<td>Pearson $r$ correlation</td>
<td>Continuous</td>
<td>Continuous</td>
</tr>
</tbody>
</table>

In theory, one could also use as ES measures, standardized regression coefficients when the outcome variables were continuous (see Chapter 9). Some have recommended using the adjusted R-squared statistic for talking about overall regression effects. Logistic regression coefficients (see Chapter 16) could be used when the

---

2 Confidence intervals are important here because they can be constructed without using a null hypothesis; $p$-values cannot. Confidence intervals focus on population values, not on the probably of sampling error in the sample statistic and are thus more important for meta-analysis, which tends to use populations, not samples. See Thompson (2002).
outcomes were categorical. But this would work only when the same predictors were used, when the same number of predictors were used, and when they were all measured the same way—a very rare circumstance except in some replication studies (see Chapter 15). Generally, meta-analysis is hard to use with regression studies that include multiple predictor variables. In this case, the typical procedure is to study the bivariate correlations among the variables in the study rather than the regression coefficients.

As with any standardized measure, so too with effect size measures: the gain in comparability is somewhat offset by losing the ability to report in natural units. Effect sizes, like z-scores, are in standard deviation units. But most people do not naturally think in standard deviation units. Mean differences are easier to understand than standardized mean differences. To say that the treatment group scored 5 points higher on the scale is pretty clear. To say that the ES is .33 seems less so. Regression coefficients can also be in natural units. We might say that for every year of experience workers earn an extra $800 in salary; this is straightforward. Saying the correlation between salary and years experience is $r = .45$ can seem too abstract. Abstraction is the price we pay for comparability. While ES statistics are useful for comparing studies, they are better known in meta-analysis for summing studies to come up with an overall mean ES.

One of the key advantages in a meta-analysis is that pooling studies results in a larger sample size and therefore more statistical power. Small effects could be more easily discernable with a meta-analysis. Meta-analysis can also identify effects that were substantial but missed because the sample sizes were too small. This advantage of meta-analysis has been especially important in experimental research which tends to have small sizes.

Meta-analysis can also result in a combined p-value for the studies summarized. In most meta-analyses, this p-value will not be very revealing. The combined sample sizes tend to be very large, which means that fairly trivial effects can yield a significant p-value. As Glass put it in an interview "statistical significance is the least interesting thing about the results. You should describe the results in terms of measures of magnitude—not just, does a treatment affect people, but how much does it affect them? That's was we need to know" (Hunt 1997, pp. 29-30, emphasis in original).

Before meta-analysis, if quantitative summaries of research were attempted, they usually used the method of "vote counting." If you had 35 studies of a subject and 8 said yes, significant effect and 27 said no, no significant effect, the outcome was clear. But is it? What if the 27 were small convenience samples with too little power to
detect significant effects and the 8 were large probability samples? In such cases, even when the vote is very lopsided, merely counting studies is a very poor way to summarize.

When you combine ES into an overall average you have to weight the studies in order to give the large studies more emphasis in your summary. Otherwise a study with 60 subjects could count as much as a study with 1,000. Weighting studies can also be done for the quality of the design, but this is controversial. This controversial practice is more common than most people realize. When studies that do not meet design criteria are eliminated, they are given a weight of zero. So in meta-analysis, as in most research, first much qualitative judgment is involved. Then you use statistics.

Texts on meta-analysis (e.g., Lipsey & Wilson 2001) often suggest hiring coders to help with the meta-analysis. When using such coders, it is very important, to train them well so that they all use the same criteria. That is good advice, but better advice, I think, is to do the coding yourself if at all possible. You could also trust co-authors because they too would have a vested interest in the accuracy of the work. As mentioned above, when discussing data entry, the use of hirelings is probably not good practice if you are serious about your work. Legend has it that Ronald Fisher, probably the greatest 20th-century statistician, when doing agricultural experiments, cleaned out animal cages himself rather have to worry that the hired help would do the work inconsistently or in a way that could bias the results. This caution and the practice of doing your own data gathering and data entry seems to me especially appropriate in a meta-analysis. The "respondents" in a meta-analysis are not anonymous subjects picked by random digit dialing. They are important documents written by professional colleagues. They should be treated with care and respect.

Meta-analysis is largely a descriptive technique. You describe the population of studies. Merely reporting a mean effect size is as risky as merely reporting any other mean without first examining the distribution of individual scores. Measures of dispersion and the shapes of distributions are very important. You don't want to use only the mean to describe your data if your data are bimodal or curvilinear. The full range of descriptive techniques (see Chapter 4) should be considered, including graphics. Scatter diagrams are especially useful in this regard (Light & Pillemer 1984). For example, say you have 50 studies and have calculated the ES for each of them. You can plot the ESs against any other continuous variable, such as their dates of publication or the ages of their subjects. The idea of this work is, as always, is to help you to discover important patterns.
Finally, if you want your study to be meta-analyzed some day, you need to report some of your data quite fully. Means, standard deviations, exact p-values, Pearson r's and Ns are among the most important statistics needed by the meta-analyst. When authors do not make some of this available, it is often possible for meta-analysts to calculate the missing information. But authors should make it available in their publications, or if space constraints keep them from doing so, they can offer in a footnote to make full range of data available. Providing access to what meta-analysts need is increasingly considered a matter of professional ethics.

**Conclusions**

Reviewing the research of others is fundamentally important. It can be an end in its own right or the indispensable first step in preparing to conduct your own research study. While this chapter discusses traditional narrative literature reviews and meta-analytic reviews separately, they are two means of approaching the same goals. Whether the review is narrative or quantitative in orientation, Light & Pillemer (1984) conclude that "by capitalizing on study-level variation" research reviews "show their strongest advantage over even the most carefully executed single study" (p. 45). Others have emphasized finding not variation but similarities among the studies reviewed. Meta-analysis, says Hunt (1997) enables one to "discover patterns in the seemingly hopeless jumble of dissimilar findings" (p. 13).

In either case, not only do research reviews enable you to discover new things, they prepare you to discover still more in your research. Setting about their work innocent of a thorough grounding in what others have done has sometimes facilitated the work of geniuses, but for the rest of us it mostly results in mediocre rehashes of projects that have already been carried out more effectively by someone else.

Once you have assembled your studies for review you'll want to compare and contrast them according to a broad range of criteria. Reviewing some of the main points in the checklist in Table 10.1, we can see that among the things you'll want to examine are (see Light & Pillemer 1984 for more discussion):

- **Settings.** Are results different or similar across settings, such as public and private schools?
- **Participants.** Do the ages or genders of participants in the research moderate some of the findings?
- **Research Design.** Do laboratory experiments yield results different from or similar to those of field experiments?
- **Publication Date.** Do more recent studies show stronger or weaker effects?
• Analysis techniques. How do studies using different analysis techniques (e.g., contingency tables with chi-squares versus log-liner analyses) compare?

• Control Variables. If one study controls for ethnicity and another does not, do their outcomes vary?

There is much more to a good quality research review than most beginning researchers (and many veterans) realize. It is almost never time ill-spent. For many professional practitioners, the results of a research review on a topic can be much more useful than the results of any individual study. A meta-analysis or other research review will probably on average take about the same amount of time and effort as conducting your own small-scale study. A research review may be a better choice for you. But be forewarned. People who do research reviews often get the urge to do their own study, to investigate the topic the right way. That is particularly the case when they suspect that the setting in which they work is different from those that have been researched by others.

This urge to conduct your own study of your own setting often leads professional practitioners in education and other fields to undertake an evaluation study. Meta-analysis is among the most useful techniques for choosing among the array of programs that you might want to institute in your local setting. To see if your selection has had the consequences you were hoping for a program evaluation is needed. Program evaluation is the subject of our next chapter.
Table 10.1. Checklist of questions to ask when critiquing the typical quantitative* research report

A. First, and in general
1. What is the hypothesis or the research question, if any, guiding the research?
2. Why do the authors believe it is important to investigate this hypothesis/question?
3. What method(s) did the author(s) use to collect evidence?
4. Were the methods appropriate to address this problem/question?
5. What are the main findings or conclusions of the article?
6. Are the conclusions convincing?

B. Questions about the variables
7. What is the dependent variable?
8. What are the independent variables?
9. Are any mediating or intervening variables identified?
10. Should these or any other mediating/intervening variables have been studied?
11. Are any control variables considered?
12. Should these control variables have been examined?
13. Does the article discuss the possible moderating variables and interaction effects? Should it?
14. How are the variables defined and measured, i.e., how are they operationalized?
15. Are the definitions and measurements appropriate?

C. Questions about the sample/subjects
16. Who is studied and are the subjects appropriate given the goals of the study?
17. How many are studied and is this enough for the purposes of the study?
18. Is the sample representative of a population? How broadly can the conclusions be generalized?

D. Questions about the conclusions
19. Are the findings statistically significant?
20. Are the findings scientifically significant?
21. How big are the effects discovered?
22. Are the findings practically significant?
23. Are the conclusions really supported by the evidence cited in the article?

E. Finally and implied in the answers to the above questions
24. How could the research have been improved?
25. What questions or problems does the article leave unanswered?
26. How could you go about doing a better job?

*One would want to ask roughly the same sorts of questions about qualitative articles, but much of the terminology would tend to be different.

**Note, for other discussions of these and related points see Gall, Gall & Borg (1999), pages 516-524; Stern & Kaloff (1996), pages 144-156; Haller & Kleine (2001), chapter 3; and Meltzoff 1997, pages 164-166; Light & Pillemer (1984), chapter 6.