



EVALUATING RESEARCH: UNDERSTANDING ELEMENTS OF A STUDY HELPS DETERMINE CREDITABILITY

Special Report

University of Pittsburgh Office of Child Development

Serving Children and Families By Promoting

Interdisciplinary Education and Research • University-Community Programs • Dissemination

The influence of research is far reaching. It informs and guides children's policy and practice, provides evidence of the effectiveness of approaches and programs, identifies the characteristics that strengthen and weaken them, and can shape the opinions of policymakers and the public on important issues.

Not all research, however, is equal. Factors ranging from the credibility of those doing the research to the design of the study determine to what degree the findings can be trusted. This report, based on advice published by the National Association of Child Advocates, offers an overview of factors that should be considered when trying to determine the credibility of research related to children and families.

Sources

The source of the study is a factor worth considering. In the absence of detailed information about a study, the reputation of researchers and institutions is a helpful guide. Those who are known in their fields for quality are more likely to produce credible work. The experience of a researcher or institution in the topic being studied is another consideration, although some young researchers produce very high quality work.

It is also important to know a little about who financed the study. Groups with strong political or commercial agendas, for example, are likely to have an interest in research that supports their viewpoints. Although the political position of an organization does not invalidate the studies it supports, it is a factor that should be examined when evaluating the credibility of those studies. Most behavioral studies on children are financed by the government and conducted at universities, which do not have obvious commercial interests.

News Media As A Source

The news media occasionally reports the findings of studies related to children and families, particularly when studies touch on high-profile issues, such as welfare reform or juvenile crime.

However, news media accounts are likely to be incomplete.

Newspapers, magazines, and television and radio news each have constraints that influence their coverage: limits on space in newspapers and magazines, for example, and time limitations in television and radio broadcasts. News organizations focus their reporting on what they perceive to be of interest to the general public. News media accounts also tend to highlight the most attention-grabbing elements of a study.

As a result, news reports of studies tend to be much more concise and far less technical than accounts of the same study that appear in research journals. They necessarily must omit most of the details. In research, however, the "devil is in the details." Credibility is associated with several scholars agreeing with the conclusions.

What Publication Suggests

Studies published in reputable research journals undergo peer review and meet the standards of the publishing organization. These studies, as a result, earn a high-level of credibility.

If a study is unpublished or appears in a publication that does not require peer review – a chapter in an edited book, for example – the fact that it has not undergone the scrutiny of independent experts should be considered when evaluating the weight to give its findings.

That is not to say that all unpublished work or work published without the scrutiny of peer-review is of poor quality. All studies start out unpublished. The reputation of the institution that conducted the study and the reputation of the organization that published it are guides to evaluate the reliability of work that does not appear in peer-reviewed journals. Many professional organizations, such as the American Psychological Association, urge their members not to publicize their research until it has been peer reviewed, but this cannot always be done.

Research Characteristics

Understanding certain characteristics of a study,

such as research design and how the issue in question is defined, helps determine the validity of its findings and more accurately interpret the results and assess their implications for specific purposes.

Research Design Varies

Research design is an important indicator of the quality of a study. Research design is the way a study is structured to answer a question.

Studies using an experimental design provide the best cause-effect information. They compare groups that have been experimentally given different experiences by the researchers. These studies offer the strongest evidence that an outcome was the result of a specific program rather than the result of other conditions, events, or the predisposition of the participants. Among the different types of experimental design, there are two general categories that provide different levels of evidence.

- **Randomized Experimental Design.** True experimental designs compare a *treatment group* – people who have received an intervention – to a similar group of people who did not receive the intervention, known as a *control group* or *comparison group*. The key characteristic of these studies is that participants are randomly assigned to the treatment and control groups. This methodology allows researchers to state with more confidence that the intervention studied was responsible for the outcomes. It is the principal method used to determine a cause-and-effect relation. However, the methodology has its limitations. A single randomized study may not produce the same results in another setting, under different conditions, or with different types of participants. Also, the results are limited to people who are randomly assigned to a treatment, whether they like it or not. Some treatments work better if people choose, like, or have faith in it. Additional trials in other settings that replicate the effects increase confidence in the program's effectiveness.

- **Quasi-Experimental Studies.** Quasi-experimental design compares groups involved in a program or model with groups that are not. These studies *do not* use random assignment to create the groups. Instead, they find groups with similar characteristics to study and give the treatment to one or more and not to the others. Careful statistical controls are required to match comparison groups with treatment groups. To conclude that the treatment caused differences between the groups, one needs to believe the groups or participants were essentially identical.

- **Pre-Post-Test Only Intervention.** They are particularly useful for studying complex systems as they exist in the community. The lowest level of experimental design is the pre-post test design without an adequate control group. Without control groups it is difficult to know if program effects are due to normal growth and development, other programs and services, or other factors that changed between the pre- and post-test, such as changing economic conditions or enthusiasm for any special treatment.

- **Observational Designs.** In some cases, experimental designs are not possible. It is sometimes impractical or unethical to assign children to different treatment groups to attain the control needed for the experiment. For example, researchers, for ethical reasons, cannot assign children to low quality education. Often, researchers simply observe and measure the difference between two or more groups of people with contrasting naturally-occurring experiences. For example, children in early care who have more highly educated teachers score better on school readiness tests than children in care with less educated teachers. Does mean that the better educated teachers produced (i.e., caused) that readiness advantage? Maybe, but maybe not. Perhaps children from more highly educated families go to care settings that have higher educated teachers, but it is their home experience that produces the readiness advantage.

Why Sampling Is Important

Researchers gather information on a sample of people to determine the effects of a program for the full population. Knowing the size of the sample and how it was collected helps determine the reliability of a study and whether its results can reasonably be applied to one group or another or to larger numbers of people.

The minimum size of a sample depends on how large the effects being studied are. A general guideline for a minimum sample size might be 30-50 people. The larger the sample, the smaller the difference needed between groups to attain statistical significance (i.e., the evidence is sufficient to say the differences were not due to just chance).

Even more important than the size of the sample is how it was collected. If researchers are to assume that the findings for a sample of people can be generalized to a larger group, they must be careful to select a sample that fairly represents that group. If sampling is biased toward some types of people (i.e., upper educated, Caucasian), the study's findings might not apply to those people not fairly represented in the sample.

One appropriate sampling procedure is random selection, but it can rarely be implemented. Sometimes participants are randomly selected from a specific group, such as from a single preschool or an agency's client list. But conclusions only apply to that group and similar groups. Some surveys, such as random digit dialing, come close to having a random sample of people in a geographic area.

An important aspect of sampling is the response rate, which is the proportion of people selected to be in the study who actually participated. A low response rate means that a portion of the sample was not studied and suggests that those who did not respond are different in some systematic way from people who did respond. If possible, researchers usually try to determine if those who participate are different in some way than those who refuse, but it is difficult to be certain or comprehensive.

Statistical Significance

Statistics are used to test whether the results researchers find are likely due to the intervention studied and not certain other factors. When studies report a statistically significant outcome, it means that it is unlikely the outcome is simply due to chance.

Here is an example: In a hypothetical case, a study finds that 75% of children given health care had acceptable school attendance but only 50% of children who did not receive health care had acceptable attendance. When researchers report the difference as statistically significant, it means the outcomes of the two groups were not simply due to the fact that any two groups of children would not have identical attendance by chance, even if health care made no difference at all.

In some cases, a finding that may not be statistically significant because the sample size was too small may still be meaningful because it suggests an important change in an outcome. Other times, a result may be statistically significant because of a large sample size but the difference really is not very large or important from a practical or policy standpoint.

Findings Relate To Groups

Studies usually base their findings on comparisons

between groups of people. Such findings tend to be particularly relevant for policy decisions, but less relevant for decisions on individual cases.

More Studies, Better Understanding

No single study tells the whole story. Science, in general, is about the aggregation of specific studies, each building upon the other and each representing different aspects or circumstances of an issue to increase the body of evidence on a particular topic. A deeper, more complete understanding of any issue important to children and families comes when many specific studies are examined together, such as in a literature review (which are often reported in *Developments* as Special Reports).

Quality Matters

Studies of varying quality on the same topic should not be given the same weight. Those of higher-quality – randomized experimental studies, for example – should be given more weight when compared to less rigorous research, especially when deciding cause and effect. But observational studies and quasi-experimental studies may be better indications of what actually happens in society.

Sometime it is difficult to find rigorous studies done on new topics because the body of research evidence is still thin. Available studies may be useful in providing information that suggests what is going on, but they should not be considered definitive until numerous studies are done on the topic and a more complete picture is available.

References

Schaefer, S.A. (2001). *Understanding Research: Top Ten Tips for Advocates and Policymakers*. Washington, DC: National Association of Child Advocates.

This Special Report, written by Jeffery Fraser, is largely based on the above-referenced report. It is not intended to be an original work but a summary for the convenience of our readers.

**University of Pittsburgh Office of Child Development, a program of the University
Center for Social and Urban Research, 400 N. Lexington Avenue,
Pittsburgh, PA 15208 (412) 244-5421. Internet: www.pitt.edu/~ocdweb/.**