THE GOOD, THE BAD, AND THE UNSCIENTIFIC: EVALUATING EDUCATION RESEARCH

by

Brian L. Carpenter
CEO
National Charter Schools Institute

Introduction

A few years ago I attended a workshop at a charter school conference where a presenter was promoting a reading intervention program he had developed. He flashed a couple of PowerPoint® slides with some graphs claiming that the intervention's effectiveness was confirmed by research he had conducted.

In the minds of many of the teachers and principals also attending, his "research findings" must have sounded credible: His books were flying off the sales table during the intermission.

Was the presenter's reading intervention effective? I couldn't tell from the information presented but I was skeptical. Based on his graphs, I suspected that in actuality he had little or no truly scientific evidence to assess its effectiveness.

My suspicion was confirmed when I later exchanged emails with him. The few details he provided in response to my questions showed that he had violated several key assumptions pertaining to the research design and methods he used.

How important are such assumptions in scientific research? One of the best answers comes from a noted British professor, Andy Field, who says, "when assumptions are broken we stop being able to draw accurate conclusions about reality" (2005, p. 171).

Thus, in this instance, by doing two simple things—being skeptical and checking the researcher's methods—I was able to conclude that this particular reading intervention program lacked exposure to rigorous scientific evaluation, i.e., the presenter's research was unscientific.

Does this mean that I wouldn't recommend his reading intervention to a school? Not necessarily. It does, however, mean that if I did, I would be doing so with the knowledge that it was not validated by scientifically based research—contrary to the claims of its developer.

The key to being able to properly evaluate research begins with understanding what it is exactly that qualifies research as *scientific*. To that end, several excellent resources are extensively cited in this paper and listed in the references section. Among them:

- *Scientific Research in Education* (Shavelson & Towne, 2002)
- *The Voice of Evidence in Reading Research* (McCardle & Chhabra, 2004)
- "The Science of Reading Research" (Lyon, 2004)

The first two items are books, while the latter is an article published by the Association for Supervision and Curriculum Development. All three resources are invaluable for educators and policymakers who evaluate educational research.

While most educators agree that educational public policy, instructional methods and curricula should be based on sound research, people like the reading intervention presenter still get away with selling their misrepresented wares because the people buying them don't know how to evaluate their research claims. Yet our students deserve better. My purpose for this paper is to provide a brief primer on evaluating research by teaching potential buyers the right questions to ask.

By exercising healthy skepticism, you place the burden of proof on the researcher, where it belongs. In effect, your skepticism says, "I'm willing to accept your findings as true if you can convince me that your work upholds the principles of credible evidence."

Ask a Skeptic's First Two Questions

When it comes to evaluating the quality of research, it helps to start with a healthy amount of skepticism. As people, sometimes we're reluctant to do that. After all, we're taught as children to believe the best about people and their work. Being skeptical seems to contradict that.

But there's a difference between being skeptical and being cynical. The skeptic says, "Show me." The cynic says, "Don't show me." The former is open to proof while the latter isn't open to anything. The skeptic is willing to be persuaded while usually nothing persuades the cynic.

By exercising healthy skepticism, you place the burden of proof on the researcher, where it belongs. In effect, your skepticism says, "I'm willing to accept your findings as true if you can convince me that your work upholds the principles of credible evidence." Your skepticism is a filter through which you determine whether the research rises to a level that is scientific and objective in other respects.

Anytime I review research, I begin skeptically by asking myself two preliminary questions:

1. Who published this research?
2. Who funded this research?

These questions may not have any bearing on the scientific quality of the research but answering them first has the effect of either heightening or lessening my skepticism.

The first question goes to objectivity. For example, does the organization that published the research exist to advance an ideological position? If so, it's pretty unlikely it will ever publish findings that contradict its ideology. The findings may nevertheless be true, but the evaluator should recognize that the ideological commitment may have influenced the objectivity of those findings. (Although this tends to be the case with basic research more than applied, the latter is not exempt by any means.)

The second question is similar to the first, but it makes a critical distinction between the organization under whose name the research is published and whose money was actually spent to fund the research.

In many cases, research is funded by grants or fees. There's nothing wrong with that. But it's often helpful to place research in perspective by considering *why* someone may have wanted to fund it.

Consider for example, an "independent" think tank which is funded from teachers' union dues and whose board leadership is composed of union bosses. Its research should heighten the evaluator's skepticism because it was probably funded to advance the ideological position of the organization.

Curriculum publishers also conduct research to demonstrate the effectiveness of their products. Such research may produce valid findings, but again, your skepticism should be heightened because curriculum companies don't fund or publish research just to add to the knowledge base in education.

Once you've answered those first two questions, you're ready to move on to evaluating the research itself.

What Is Scientific Research?

Despite common misperceptions, true scientific research is based on *universal principles* that apply to all disciplines or fields of inquiry (Fletcher & Francis, 2004). This means that what constitutes scientific research in medicine or physics is also true of education research. Yet "education research in the United States is barely 100 years old" (Shavelson & Towne, 2002, p. 13)—an interesting age because it took medicine about that long to accept randomized clinical trials as its research standard (Sweet, 2004).

A second common myth is that scientific research is determined by method, such as quantitative vs. qualitative. Fletcher and Francis note too, that this "dichotomy is false and a poor characterization of the nature of the scientific enterprise" (2004, p. 59).

So what *is* scientifically based research? Interestingly, because of mounting evidence in the area of reading research in the past few years, Congress defined scientific research in the Reading Excellence Act of 1998. Sweet (2004) also provides the definition from the Act itself:

> The term 'scientifically based reading research'—
>
> (A) means the application of rigorous, systematic, and objective procedures to obtain valid knowledge relevant to reading development, reading instruction, and reading difficulties; and
>
> (B) shall include research that—
>
> (i) employs systematic, empirical methods that draw on observation or experiment;
>
> (ii) involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;
>
> (iii) relies on measurements or observational methods that provide valid data across evaluators and observers and across multiple measurements and observations; and
>
> (iv) has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, scientific review. ("Reading and Literacy Grants to State Educational Agencies," Title II (C) Sec. 2252 (5) [20 U.S.C. § 6661a])

Sweet goes on to note that while the Reading Excellence Act has expired, its definition of scientifically based research "laid a solid foundation for Reading First, which is part of the No Child Left Behind Act of 2001," which includes "more than 110 references to the term *scientifically based research* …" (2004, p. 22).

These four principles have been amplified by Shavelson and Towne in *Scientific Research in Education* (2002). In this excellent volume published by the National Research Council, they list the scientific principles of research as follows:

1. Pose significant questions that can be investigated empirically;
2. Link research to relevant theory;
3. Use methods that permit direct investigation of the question;
4. Provide a coherent and explicit chain of reasoning;
5. Replicate and generalize across studies; and
6. Disclose research to encourage professional scrutiny and critique (Shavelson & Towne, 2002, pp. 3-5).

Though not linear in nature, these six principles, as well as how you can evaluate them, are explained in the pages that follow.

There are "more than 110 references to the term *scientifically based research* in the No Child Left Behind Act of 2001 . . . " (Sweet, 2004, p. 22).

> While there are other kinds of inquiry, they are not scientific if they do not involve "observations that can be verified, replicated, and otherwise scrutinized" (Fletcher & Francis, 2004, p. 63).

Without These Six Principles It's Not Scientific

Scientific inquiry is rooted in *empiricism*—that which can be observed through senses (Fletcher & Francis, 2004; Shavelson & Towne, 2002). While there are other kinds of inquiry, they are not scientific if they do not involve "observations that can be verified, replicated, and otherwise scrutinized" (Fletcher & Francis, 2004, p. 63).

Guiding Principle 1

Thus, we begin with the first "guiding principle" (Shavelson & Towne, 2002, p. 52), which requires that scientific research "pose significant questions that can be investigated empirically" (p. 54). This means that education research should deal with important questions and the inquiry process should be based on observation.

A significant research question is one that adds to the body of knowledge in a way "that has an important impact on the conceptual structure or outlook of a field" (Brooks, 1967, as cited by Shavelson and Towne, 2002, p. 57). This might involve investigating a new hypothesis or revising a previous one, but in either case the question should be an important one. Important questions are linked to existing theory and practice in the professional literature.

The empirical requirement means if other researchers cannot make the same observations, the research lacks the important quality of being replicable. If research cannot be replicated, its knowledge claims rest on shaky ground because it is during the process of replicating research that it is most subjected to the rigorous scrutiny of others.

For example, the question, "Do some children deserve vouchers more than others?" cannot be scientifically researched. How deserving one child may be compared to another cannot be empirically tested. The question may be addressed by policy, but that's not the same as claiming a scientific basis for doing so.

You can evaluate principle 1 by asking if other researchers have made similar findings. Further, is the research question significant? Is it answerable by observation?

Guiding Principle 2

One of the characteristics of genuine scientific research is that it should demonstrate a "link research to relevant theory" (Shavelson & Towne, 2002, p. 59).

Theory is the conceptual framework upon which the scientific researcher builds, in an effort to broaden human understanding of a particular phenomenon. This "conceptual framework, model, or theory . . . helps generate the research question" (Fletcher & Francis, 2004, p. 64).

In scientific research the researcher carefully and thoroughly reviews the existing literature and clearly describes its importance in contributing to the question under investigation. A legitimate scientific study will have ample references cited. These references should help "provide a coherent and explicit chain of reasoning" (the fourth guiding principle).

Education research that lacks such references should be regarded as unscientific. Look for, and at, the references.

Guiding Principle 3

The third guiding principle is that the researcher should "use methods that permit direct investigation of the question" (Shavelson & Towne, 2002, p. 62).

In my own reviews of education research, I have found that it is on this particular principle that substantial amounts of research fail to meet the test of science. That is, researchers use methods to make claims not supported by those methods.

For example, correlation is a measure of relationship and cannot be used to make causal claims despite the fact that many people misuse it by inferring a causal link. Often, studies that utilize correlation also violate the statistical assumptions for doing so.

Fletcher and Francis (2004) make a similar observation by noting:

> Investigators often fail to reach valid statistical conclusions because they do not fully understand statistical inference, do not understand the limitations of a statistical procedure (e.g., sensitivity to violations of assumptions), or fail to take into account the operating characteristics (power and robustness) of their procedures and designs (p. 72).

Somewhat tellingly in fact, Gall, Gall & Borg (2003) cite Tuckman who concluded from a review of "four studies that employed experts in research methodology to judge the quality of educational research published in journals and other sources" that "between 40 and 60 percent of the research studies . . . should have been extensively revised prior to publication or should not have been published at all" (p. 113).

For the practitioner or policy-maker evaluating research, such warnings should serve as a red flag to proceed with caution.

Shavelson and Towne succinctly prescribe that the:

> link between [the research] question and method must be clearly explicated and justified . . . [and] a detailed description of method—measurements, data collection procedures, and data analysis—must be available to permit others to critique or replicate the study (see Principle 5)" (2002, p. 63).

This means that if the research methods, data or procedures are not available for evaluation, the study lacks the criteria to claim the findings are scientific.

Obviously an explanation of how to evaluate methods is beyond the scope of this paper, though Shavelson and Towne present some discussion in *Scientific Research in Education* (2002). If you are evaluating research that uses inferential quantitative methods, it may be necessary to take a couple of graduate courses in statistical methods. If this is not practical, a qualified consultant can help you assess a study's methodological applications.

The bottom line on methods is this: There's a lot of substandard research being conducted. Whether you evaluate a study's methods yourself or hire someone to assist you, the salient questions are:

1. Are the researcher's methods appropriate to the question being investigated?
2. Are the assumptions for using the particular methods clearly explained in the research itself?

Failure by a researcher on either question compromises—if not invalidates—the study.

Shavelson and Towne succinctly prescribe that the "link between [the research] question and method must be clearly explicated and justified … [and] a detailed description of method—measurements, data collection procedures, and data analysis—must be available to permit others to critique or replicate the study (see Principle 5)" (2002, p. 63)

Remember: In science, skepticism is a virtue. It is in fact, one of the three "basic rules that govern the scientific process" (The other two are integrity and replication.) (Fletcher & Francis, 2004, p.69).

Guiding Principle 4

The fourth guiding principle useful for evaluating whether research is true to the standards of science is that the study must "provide [a] coherent, explicit chain of reasoning" (Shavelson & Towne, 2002, p. 66).

This is a process of linking theory, research and design and methods to findings and back to theory (Shavelson & Towne, 2002). Logical explanations and reasoned inferential thinking are the means of evaluating how well the research meets these criteria.

In the example cited in the opening paragraphs of this paper, the researcher's sample was too small and not randomly selected, and he did not have a control group. His research design failed to control for extraneous variables and his conclusions were unsupported by his research methods. Overall, his research lacked the explicit chain of reasoning required of scientific research.

As with evaluating methodology, evaluating the chain of reasoning necessitates some training. Still, "It should be clear to other scientists how links between the actual data and the research question were made" (Fletcher & Francis, 2004, p. 66). This means that while you might lack the specific skills to evaluate the chain of reasoning, another qualified person should be able to do it. With weighty decisions involving instructional interventions, the cost of retaining someone to make this evaluation is nominal compared to the impact of implementing a poorly evaluated program.

One thing that can be easily spotted, however, is when the "research" has not been properly written up or was written in a style that is amateurish or incomplete. In such cases, you have good reason to question its credibility.

Remember: In science, skepticism is a virtue. It is, in fact, one of the three "basic rules that govern the scientific process" (The other two are integrity and replication.) (Fletcher & Francis, 2004, p.69). If a researcher seems resistant to answering pointed questions about his chain of reasoning or cannot produce a written study demonstrating it, you're probably looking at claims that are unsupported by science.

Guiding Principle 5

The fifth guiding principle can be contained in a single word: convergence (McCardle & Chhabra, 2004). Convergence means that research findings are more plausibly valid if they are arrived at from multiple designs, methods and studies. Instead of using the term convergence, Shavelson and Towne state this principle as "replicate and generalize across studies" (2002, p. 70).

Replication refers to the process whereby other researchers attempt to test the hypothesis by conducting a similar study. Doing so allows researchers to both confirm the findings (and the methods) as well as extend the findings to other populations.

As Reid G. Lyon points out, "A single study's findings are not sufficient to generalize results to different populations" (2004, p. 15). In other words, in the case of the misinformed researcher introduced at the beginning of this paper, his study—even if it met the other criteria for being scientific—would be considered insufficient until other researchers had achieved similar results with other populations.

For example, a reading intervention program that is effective with children of well-educated parents may prove ineffective with children of less-educated parents. Replication allows researchers to make such generalizations.

How do you know if a study has been replicated? One way to assess this important point is to survey the literature on the topic. There are many excellent web-based sources for doing so, such as ProQuest, ERIC, and Sage. Subscribing to professional journals also gives you access to accumulating research.

Guiding Principle 6

This final principle requires that researchers "disclose research to encourage professional scrutiny and critique" (Shavelson & Towne, 2002, p. 72).

In theory, this means that scientific work should be peer reviewed or subjected to some comparable critique process. The goal of such reviews is to give other researchers the opportunity to objectively evaluate all aspects of the research in order to ensure that it meets the standards of scientific research.

In practice, however, as I noted on page 5, many studies in education are published with serious errors. Shavelson & Towne (2002) cite Lagemann (2000) who noted:

> for many reasons, the education research community has not been nearly as critical of itself as is the case in other fields of scientific study (p. 73).

Thus, you may rightly conclude that the fact a study has been published or even peer reviewed is no proof of its quality. You must still cast a critical eye on its hypotheses, literature review, design, methods, and conclusions.

Develop Your Skills

But you might be asking yourself, "If many peer reviewed studies in education are flawed, how am I as a practitioner supposed to be able to evaluate them?"

The answer is not necessarily simple but it is clear: You must develop the skills necessary for understanding the universal principles of scientific inquiry and you must practice applying them by evaluating research. In the complex information age in which we live, *every* educator must, to some degree, become a practitioner/researcher.

This doesn't mean that every teacher or principal needs to earn a PhD in research or become a psychometrician. But surely we can acquire sufficient skills to ensure that the curricula and instructional methods we use have been sufficiently vetted through the scientific process.

The good news is that, as the references in this paper have noted, there is an increasing understanding beginning to emerge in education research. Research is becoming more sophisticated and in some areas, such as teaching reading, there is no question about what works. As educators, we can advance this understanding by developing our own skills and by insisting on scientifically based resources for our schools.

You must develop the skills necessary for understanding the universal principles of scientific inquiry and you must practice applying them by evaluating research. In the complex information age in which we live, *every* educator must, to some degree, become a practitioner/researcher.

In Conclusion

Nobel Prize Laureate Richard Feynman closed one of his books, "*Surely You're Joking, Mr. Feynman!": Adventures of a Curious Character,* with a poignant remark directed toward those who would conduct scientific research (Feynman, Leighton & Hutchings, 1985). He said that "The first principle [of scientific integrity] is that you must not fool yourself—and you are the easiest person to fool" (p. 343).

Although the context for his admonition is other scientists, I think it applies well to all of us who are called upon to review and evaluate education research and to make decisions affecting the lives of children on the basis of that research. The truth is, we sometimes fool ourselves because a particular theory or intervention sounds good or because it fits nicely with our existing philosophies of education and learning.

Indeed, during the past 100 years, philosophy has quite frequently trumped science when it comes to pedagogical practice (Ravitch, 2000). Decades of abysmal test scores are proof that in large part, we haven't been doing what works. And though it may be hard to accept, E. D. Hirsch, Jr. is correct in saying that "The reason for this state of affairs—tragic for millions of students as well as the nation—is that an army of educators and reading experts are fundamentally wrong in their ideas about education and especially about reading comprehension" (2006, p. 3).

Living in an increasingly competitive global economy which places value on knowledge work (Friedman, 2006), American educators are at a crossroads. We can either continue using unproven pedagogies and curricula or we can develop the skills necessary to evaluate them against scientific research. By asking the right questions, we can assess how well such research adheres to the universal principles of scientific inquiry and we can implement in our schools and classrooms only those things that work.

References & Recommended Resources

Feynman, R. P., Leighton, R., & Hutchings, E. (1985). *"Surely you're joking, Mr. Feynman!": adventures of a curious character*. New York: W.W. Norton.

Field, A. P. (2005). *Discovering statistics using SPSS: (and sex, drugs and rock 'n' roll)* (Second ed.). London ; Thousand Oaks: Sage Publications.

Fletcher, J. M., & Francis, D. J. (2004). Scientifically based educational research. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research*. Baltimore, MD: Paul H. Brookes Publishing Company.

Friedman, T. L. (2006). *The world is flat: a brief history of the twenty-first century* (1st rev. and expanded ed.). New York: Farrar, Straus and Giroux.

Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction*. Boston, MA: Allyn and Bacon.

Hirsch, E. D. (2006). *The knowledge deficit: Creating a reading revolution for a new generation of American achievers*. Boston: Houghton Mifflin.

Lyon, G. R., & Chhabra, V. (2004). The science of reading research *Association for Supervision and Curriculum Development*, 13-17.

McCardle, P., & Chhabra, V. (Eds.). (2004). *The voice of evidence in reading research*. Baltimore, MD: Paul H. Brookes Publishing Co.

Ravitch, D. (2000). *Left back: A century of failed school reforms*. New York: Simon & Schuster.

Shavelson, R. J., & Towne, L. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Sweet, R. W. (2004). The big picture: Where we are nationally on the reading front and how we got here. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research*. Baltimore, MD: Paul H. Brookes Publishing Company.