INTRODUCTION

In the recent move toward standards-based reform in public education, many educational reform efforts require schools to demonstrate that they are achieving educational outcomes with students performing at a required level of achievement. Federal and state legislation, in particular, has codified this standards-based movement and tied funding and other incentives to student achievement.

At first, demonstrating student learning may seem like a simple task, but reflection reveals that it is a complex challenge requiring educators to use specific knowledge and skills. Standards-based reform has many curricular and instructional prerequisites. The curriculum must represent the most important knowledge, skills, and attributes that schools want their students to acquire because these learning outcomes will serve as the basis of assessment instruments. Likewise, instructional methods should be appropriate for the designed curriculum. Teaching methods should lead to students learning the outcomes that are the focus of the assessment standards.

Standards- and assessment-based educational reforms seek to obligate schools and teachers to supply evidence that their instructional methods are effective. But testing is only one of three ways to gather evidence about the effectiveness of instructional methods. Evidence of instructional effectiveness can come from any of the following sources:

- Demonstrated student achievement in **formal testing** situations implemented by the teacher, school district, or state;
- **Published findings of research-based evidence** that the instructional methods being used by teachers lead to student achievement; or
- **Proof of reason-based practice** that converges with a research-based consensus in the scientific literature. This type of justification of educational practice becomes important when direct evidence may be lacking (a direct test of the instructional efficacy of a particular method is absent), but there is a theoretical link to research-based evidence that can be traced.

Each of these methods has its pluses and minuses. While testing seems the most straightforward, it is not necessarily the clear indicator of good educational practice that the public seems to think it is. The meaning of test results is often not immediately clear. For example, comparing averages or other indicators of overall performance from tests across classrooms, schools, or school districts takes no account of the resources and support provided to a school, school district, or individual professional. Poor outcomes do not necessarily indict the efforts of physicians in Third World countries who work with substandard equipment and supplies. Likewise, objective evidence of below-grade or below-standard mean performance of a group of students should not necessarily indict their teachers if essential resources and supports (e.g., curriculum materials, institutional aid, parental cooperation) to support teaching efforts were lacking. However, the extent to which children could learn effectively even in under-equipped schools is not known because evidence-based practices are, by and large, not implemented. That is, there is evidence that children experiencing academic difficulties can achieve more educationally if they are taught with effective methods; sadly, scientific research about what works does not usually find its way into most classrooms.

Testing provides a useful professional calibrator, but it requires great contextual sensitivity in interpretation. It is not the entire solution for assessing the quality of instructional efforts. This is why research-based and reason-based educational practice are also crucial for determining the quality and impact of programs. Teachers thus have the responsibility to be effective users and interpreters of research. Providing a survey and synthesis of the most effective practices for a variety of key curriculum goals (such as literacy and numeracy) would seem to be a helpful idea, but no document could provide all of that information. (Many excellent research syntheses exist, such as the National Reading Panel, 2000; Snow, Burns, & Griffin, 1998; Swanson, 1999, but the knowledge base about effective educational practices is constantly being updated, and many issues remain to be settled.)

As professionals, teachers can become more effective and powerful by developing the skills to recognize scientifically based practice and, when the evidence is not available, use some basic research concepts to draw conclusions on their own. This paper offers a primer for those skills that will allow teachers to become independent evaluators of educational research.

The Formal Scientific Method and Scientific Thinking in Educational Practice

When you go to your family physician with a medical complaint, you expect that the recommended treatment has proven to be effective with many other patients who have had the same symptoms. You may even ask why a particular medication is being recommended for you. The doctor may summarize the background knowledge that led to that recommendation and very likely will cite summary evidence from the drug's many clinical trials and perhaps even give you an overview of the theory behind the drug's success in treating symptoms like yours.

All of this discussion will probably occur in rather simple terms, but that does not obscure the fact that the doctor has provided you with *data* to support a *theory* about your complaint and its treatment. The doctor has shared knowledge of medical *science* with you. And while everyone would agree that the practice of medicine has its "artful" components (for example, the creation of a healing relationship between doctor and patient), we have come to expect and depend upon the scientific foundation that underpins even the artful aspects of medical treatment. Even when we do not ask our doctors specifically for the data, we assume it is there, supporting our course of treatment.

Actually, Vaughn and Dammann (2001) have argued that the correct analogy is to say that teaching is in part a craft, rather than an art. They point out that craft knowledge is superior to alternative forms of knowledge such as superstition and folklore because, among other things, craft knowledge is compatible with scientific knowledge and can be more easily integrated with it. One could argue that in this age of education reform and accountability, educators are being asked to demonstrate that their craft has been integrated with science—that their instructional models, methods, and materials can be likened to the evidence a physician should be able to produce showing that a specific treatment will be effective. As with medicine, constructing teaching practice on a firm scientific foundation does not mean denying the craft aspects of teaching.

Architecture is another professional practice that, like medicine and education, grew from being purely a craft to a craft based firmly on a scientific foundation. Architects wish to design

beautiful buildings and environments, but they must also apply many foundational principles of engineering and adhere to structural principles. If they do not, their buildings, however beautiful they may be, will not stand. Similarly, a teacher seeks to design lessons that stimulate students and entice them to learn—lessons that are sometimes a beauty to behold. But if the lessons are not based in the science of pedagogy, they, like poorly constructed buildings, will fail.

Education is informed by formal scientific research through the use of archival research-based knowledge such as that found in peer-reviewed educational journals. Preservice teachers are first exposed to the formal scientific research in their university teacher preparation courses (it is hoped), through the instruction received from their professors, and in their course readings (e.g., textbooks, journal articles). Practicing teachers continue their exposure to the results of formal scientific research by subscribing to and reading professional journals, by enrolling in graduate programs, and by becoming lifelong learners.

Scientific thinking in practice is what characterizes reflective teachers—those who inquire into their own practice and who examine their own classrooms to find out what works best for them and their students. What follows in this document is, first, a "short course" on how to become an effective consumer of the archival literature that results from the conduct of formal scientific research in education and, second, a section describing how teachers can think scientifically in their ongoing reflection about their classroom practice.

Being able to access mechanisms that evaluate claims about teaching methods and to recognize scientific research and its findings is especially important for teachers because they are often confronted with the view that "anything goes" in the field of education—that there is no such thing as best practice in education, that there are no ways to verify what works best, that teachers should base their practice on intuition, or that the latest fad must be the best way to teach, please a principal, or address local school reform. The "anything goes" mentality actually represents a threat to teachers' professional autonomy. It provides a fertile environment for gurus to sell untested educational "remedies" that are not supported by an established research base.

TEACHERS AS INDEPENDENT EVALUATORS OF RESEARCH EVIDENCE

One factor that has impeded teachers from being active and effective consumers of educational science has been a lack of orientation and training in how to understand the scientific process and how that process results in the cumulative growth of knowledge that leads to validated educational practice. Educators have only recently attempted to resolve educational disputes scientifically, and teachers have not yet been armed with the skills to evaluate disputes on their own.

Educational practice has suffered greatly because its dominant model for resolving or adjudicating disputes has been more political (with its corresponding factions and interest groups) than scientific. The field's failure to ground practice in the attitudes and values of science has made educators susceptible to the "authority syndrome" as well as fads and gimmicks that ignore evidence-based practice.

When our ancestors needed information about how to act, they would ask their elders and other wise people. Contemporary society and culture are much more complex. Mass communication allows virtually anyone (on the Internet, through self-help books) to proffer advice, to appear to be a "wise elder." The current problem is how to sift through the avalanche of misguided and uninformed advice to find genuine knowledge. Our problem is not information; we have tons of information. What we need are quality control mechanisms.

Peer-reviewed research journals in various disciplines provide those mechanisms. However, even with mechanisms like these in behavioral science and education, it is all too easy to do an "end run" around the quality control they provide. Powerful information dissemination outlets such as publishing houses and mass media frequently do not discriminate between good and bad information. This provides a fertile environment for gurus to sell untested educational "remedies" that are not supported by an established research base and, often, to discredit science, scientific evidence, and the notion of research-based best practice in education. As Gersten (2001) notes, both seasoned and novice teachers are "deluged with misinformation" (p. 45).

We need tools for evaluating the credibility of these many and varied sources of information; the ability to recognize research-based conclusions is especially important. Acquiring those tools means understanding scientific values and learning methods for making inferences from the research evidence that arises through the scientific process. These values and methods were recently summarized by a panel of the National Academy of Sciences convened on scientific inquiry in education (Shavelson & Towne, 2002), and our discussion here will be completely consistent with the conclusions of that NAS panel.

The scientific criteria for evaluating knowledge claims are not complicated and could easily be included in initial teacher preparation programs, but they usually are not (which deprives teachers from an opportunity to become more efficient and autonomous in their work right at the beginning of their careers). These criteria include:

- the **publication** of findings in refereed journals (scientific publications that employ a process of peer review),
- the **duplication** of the results by other investigators, and
- a **consensus** within a particular research community on whether there is a critical mass of studies that point toward a particular conclusion.

In their discussion of the evolution of the American Educational Research Association (AERA) conference and the importance of separating research evidence from opinion when making decisions about instructional practice, Levin and O'Donnell (2000) highlight the importance of enabling teachers to become independent evaluators of research evidence. Being aware of the importance of research published in peer-reviewed scientific journals is only the first step because this represents only the most minimal of criteria. Following is a review of some of the principles of research-based evaluation that teachers will find useful in their work.

PUBLICLY VERIFIABLE RESEARCH CONCLUSIONS: REPLICATION AND PEER REVIEW

Source credibility: the consumer protection of peer reviewed journals. The front line of defense for teachers against incorrect information in education is the existence of peer-reviewed journals in education, psychology, and other related social sciences. These journals publish empirical research on topics relevant to classroom practice and human cognition and learning. They are the first place that teachers should look for evidence of validated instructional practices.

As a general quality control mechanism, peer review journals provide a "first pass" filter that teachers can use to evaluate the plausibility of educational claims. To put it more concretely, one ironclad criterion that will always work for teachers when presented with claims of uncertain validity is the question: *Have findings supporting this method been published in recognized scientific journals that use some type of peer review procedure?* The answer to this question will almost always separate pseudoscientific claims from the real thing.

In a peer review, authors submit a paper to a journal for publication, where it is critiqued by several scientists. The critiques are reviewed by an editor (usually a scientist with an extensive history of work in the specialty area covered by the journal). The editor then decides whether the weight of opinion warrants immediate publication, publication after further experimentation and statistical analysis, or rejection because the research is flawed or does not add to the knowledge base. Most journals carry a statement of editorial policy outlining their exact procedures for publication, so it is easy to check whether a journal is in fact, peer-reviewed.

Peer review is a minimal criterion, not a stringent one. Not all information in peer-reviewed scientific journals is necessarily correct, but it has at the very least undergone a cycle of peer criticism and scrutiny. However, it is because the presence of peer-reviewed research is such a minimal criterion that its absence becomes so diagnostic. The failure of an idea, a theory, an educational practice, behavioral therapy, or a remediation technique to have adequate documentation in the peer-reviewed literature of a scientific discipline is a very strong indication to be wary of the practice.

The mechanisms of peer review vary somewhat from discipline to discipline, but the underlying rationale is the same. Peer review is one way (replication of a research finding is another) that science institutionalizes the attitudes of objectivity and public criticism. Ideas and experimentation undergo a honing process in which they are submitted to other critical minds for evaluation. Ideas that survive this critical process have begun to meet the criterion of public verifiability. The peer review process is far from perfect, but it really is the only external consumer protection that teachers have.

The history of reading instruction illustrates the high cost that is paid when the peer-reviewed literature is ignored, when the normal processes of scientific adjudication are replaced with political debates and rhetorical posturing. A vast literature has been generated on best practices that foster children's reading acquisition (Adams, 1990; Anderson, Hiebert, Scott, & Wilkinson, 1985; Chard & Osborn, 1999; Cunningham & Allington, 1994; Ehri, Nunes, Stahl, & Willows, 2001; Moats, 1999; National Reading Panel, 2000; Pearson, 1993; Pressley, 1998; Pressley, Rankin, & Yokol, 1996; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2002; Reading Coherence Initiative, 1999; Snow, Burns, & Griffin, 1998; Spear-Swerling & Sternberg, 2001). Yet much of this literature remains unknown to many teachers, contributing to the frustrating lack of clarity about accepted, scientifically validated findings and conclusions on reading acquisition.

Teachers should also be forewarned about the difference between professional education journals that are magazines of opinion in contrast to journals where primary reports of research, or reviews of research, are peer reviewed. For example, the magazines *Phi Delta Kappan* and *Educational Leadership* both contain stimulating discussions of educational issues, but neither is a peer-reviewed journal of original research. In contrast, the *American Educational Research Journal* (a flagship journal of the AERA) and the *Journal of Educational Psychology* (a flagship journal of the American Psychological Association) are both peer-reviewed journals of original research. Both are main sources for evidence on validated techniques of reading instruction and for research on aspects of the reading process that are relevant to a teacher's instructional decisions.

This is true, too, of presentations at conferences of educational organizations. Some are data-based presentations of original research. Others are speeches reflecting personal opinion about

educational problems. While these talks can be stimulating and informative, they are not a substitute for empirical research on educational effectiveness

Replication and the importance of public verifiability. Research-based conclusions about educational practice are public in an important sense: they do not exist solely in the mind of a particular individual but have been submitted to the scientific community for criticism and empirical testing by others. Knowledge considered "special"—the province of the thought of an individual and immune from scrutiny and criticism by others—can never have the status of scientific knowledge. Research-based conclusions, when published in a peer reviewed journal, become part of the public realm, available to all, in a way that claims of "special expertise" are not.

Replication is the second way that science uses to make research-based conclusions concrete and "public." In order to be considered scientific, a research finding must be presented to other researchers in the scientific community in a way that enables them to attempt the same experiment and obtain the same results. When the same results occur, the finding has been *replicated*. This process ensures that a finding is not the result of the errors or biases of a particular investigator. Replicable findings become part of the converging evidence that forms the basis of a research-based conclusion about educational practice.

John Donne told us that "no man is an island." Similarly, in science, no researcher is an island. Each investigator is connected to the research community and its knowledge base. This interconnection enables science to grow cumulatively and for research-based educational practice to be built on a convergence of knowledge from a variety of sources. Researchers constantly build on previous knowledge in order to go beyond what is currently known. This process is possible only if research findings are presented in such a way that any investigator can use them to build on.

Philosopher Daniel Dennett (1995) has said that science is "making mistakes in public. Making mistakes for all to see, in the hopes of getting the others to help with the corrections" (p. 380). We might ask those proposing an educational innovation for the evidence that they have in fact "made some mistakes in public." Legitimate scientific disciplines can easily provide such evidence. For example, scientists studying the psychology of reading once

thought that reading difficulties were caused by faulty eye movements. This hypothesis has been shown to be in error, as has another that followed it, that so-called visual reversal errors were a major cause of reading difficulty. Both hypotheses were found not to square with the empirical evidence (Rayner, 1998; Share & Stanovich, 1995). The hypothesis that reading difficulties can be related to language difficulties at the phonological level has received much more support (Liberman, 1999; National Reading Panel, 2000; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2002; Shankweiler, 1999; Stanovich, 2000).

After making a few such "errors" in public, reading scientists have begun, in the last 20 years, to get it right. But the only reason teachers can have confidence that researchers are now "getting it right" is that researchers made it open, public knowledge when they got things wrong. Proponents of untested and pseudoscientific educational practices will never point to cases where they "got it wrong" because they are not committed to public knowledge in the way that actual science is. These proponents do not need, as Dennett says, "to get others to help in making the corrections" because they have no intention of correcting their beliefs and prescriptions based on empirical evidence.

Education is so susceptible to fads and unproven practices because of its tacit endorsement of a personalistic view of knowledge acquisition—one that is antithetical to the scientific value of the public verifiability of knowledge claims. Many educators believe that knowledge resides within particular individuals—with particularly elite insights—who then must be called upon to dispense this knowledge to others. Indeed, some educators reject public, depersonalized knowledge in social science because they believe it dehumanizes people. Science, however, with its conception of publicly verifiable knowledge, actually *democratizes* knowledge. It frees practitioners and researchers from slavish dependence on authority.

Subjective, personalized views of knowledge degrade the human intellect by creating conditions that subjugate it to an elite whose "personal" knowledge is not accessible to all (Bronowski, 1956, 1977; Dawkins, 1998; Gross, Levitt, & Lewis, 1997; Medawar, 1982, 1984, 1990; Popper, 1972; Wilson, 1998). Empirical science, by generating knowledge and moving it into the public domain, is a liberating force. Teachers can consult the research and decide for themselves whether the state of the literature is as the expert portrays it. All teachers can benefit from some rudimentary grounding in the most fundamental principles of scientific inference. With knowledge of a few uncomplicated research principles, such as

control, manipulation, and randomization, anyone can enter the open, public discourse about empirical findings. In fact, with the exception of a few select areas such as the eye movement research mentioned previously, much of the work described in noted summaries of reading research (e.g., Adams, 1990; Snow, Burns, & Griffin, 1998) could easily be replicated by teachers themselves.

There are many ways that the criteria of replication and peer review can be utilized in education to base practitioner training on research-based best practice. Take continuing teacher education in the form of inservice sessions, for example. Teachers and principals who select speakers for professional development activities should ask speakers for the sources of their conclusions in the form of research evidence in peer-reviewed journals. They should ask speakers for bibliographies of the research evidence published on the practices recommended in their presentations.

THE SCIENCE BEHIND RESEARCH-BASED PRACTICE RELIES ON SYSTEMATIC EMPIRICISM

Empiricism is the practice of relying on observation. Scientists find out about the world by examining it. The refusal by some scientists to look into Galileo's telescope is an example of how empiricism has been ignored at certain points in history. It was long believed that knowledge was best obtained through pure thought or by appealing to authority. Galileo claimed to have seen moons around the planet Jupiter. Another scholar, Francesco Sizi, attempted to refute Galileo, not with observations, but with the following argument:

> There are seven windows in the head, two nostrils, two ears, two eyes and a mouth; so in the heavens there are two favorable stars, two unpropitious, two luminaries, and Mercury alone undecided and indifferent. From which and many other similar phenomena of nature such as the seven metals, etc., which it were tedious to enumerate, we gather that the number of planets is necessarily seven...ancient nations, as well as modern Europeans, have adopted the division of the week into seven days, and have named them from the seven planets; now if we increase the number of planets, this whole system falls to the ground...moreover, the satellites are invisible to the naked eye and

> therefore can have no influence on the earth and therefore would be useless and therefore do not exist. (Holton & Roller, 1958, p. 160)

Three centuries of the demonstrated power of the empirical approach give us an edge on poor Sizi. Take away those years of empiricism, and many of us might have been there nodding our heads and urging him on. In fact, the empirical approach is not necessarily obvious, which is why we often have to teach it, even in a society that is dominated by science.

Empiricism pure and simple is not enough, however. Observation itself is fine and necessary, but pure, unstructured observation of the natural world will not lead to scientific knowledge. Write down every observation you make from the time you get up in the morning to the time you go to bed on a given day. When you finish, you will have a great number of facts, but you will not have a greater understanding of the world. Scientific observation is termed *systematic* because it is structured so that the results of the observation reveal something about the underlying causal structure of events in the world. Observations are structured so that, depending upon the outcome of the observation, some theories of the causes of the outcome are supported and others rejected.

Teachers can benefit by understanding two things about research and causal inferences. The first is the simple (but sometimes obscured) fact that statements about best instructional practices are statements that contain a causal claim. These statements claim that one type of method or practice causes superior educational outcomes. Second, teachers must understand how the logic of the experimental method provides the critical support for making causal inferences.

Science addresses testable questions

Science advances by positing theories to account for particular phenomena in the world, by deriving predictions from these theories, by testing the predictions empirically, and by modifying the theories based on the tests (the sequence is typically theory –> prediction –> test –> theory modification). What makes a theory testable? A theory must have specific implications for observable events in the natural world.

Science deals only with a certain class of problem: the kind that is empirically solvable. That does not mean that different classes of problems are inherently solvable or unsolvable and that this division is fixed forever. Quite the contrary: some problems that are currently unsolvable may become solvable as theory and empirical techniques become more sophisticated. For example, decades ago historians would not have believed that the controversial issue of whether Thomas Jefferson had a child with his slave Sally Hemings was an empirically solvable question. Yet, by 1998, this problem had become solvable through advances in genetic technology, and a paper was published in the journal *Nature* (Foster, Jobling, Taylor, Donnelly, Deknijeff, Renemieremet, Zerjal, & Tyler-Smith, 1998) on the question.

The criterion of whether a problem is "testable" is called the *falsifiability* criterion: a scientific theory must always be stated in such a way that the predictions derived from it can potentially be shown to be false. The falsifiability criterion states that, for a theory to be useful, the predictions drawn from it must be specific. The theory must go out on a limb, so to speak, because in telling us what *should* happen, the theory must also imply that certain things will *not* happen. If these latter things *do* happen, it is a clear signal that something is wrong with the theory. It may need to be modified, or we may need to look for an entirely new theory. Either way, we will end up with a theory that is closer to the truth.

In contrast, if a theory does not rule out any possible observations, then the theory can never be changed, and we are frozen into our current way of thinking with no possibility of progress. A successful theory cannot posit or account for every possible happening. Such a theory robs itself of any predictive power.

What we are talking about here is a certain type of intellectual honesty. In science, the proponent of a theory is always asked to address this question *before* the data are collected: "What data pattern would cause you to give up, or at least to alter, this theory?" In the same way, the falsifiability criterion is a useful consumer protection for the teacher when evaluating claims of educational effectiveness. Proponents of an educational practice should be asked for evidence; they should also be willing to admit that contrary data will lead them to abandon the practice. True scientific knowledge is held tentatively and is subject to change based on contrary evidence. Educational remedies not based on scientific evidence will often fail to put themselves at risk by specifying what data patterns would prove them false.

Objectivity and Intellectual Honesty

Objectivity, another form of intellectual honesty in research, means that we let nature "speak for itself" without imposing our wishes on it—that we report the results of experimentation as accurately as we can and that we interpret them as fairly as possible. (The fact that this goal is unattainable for any single human being should not dissuade us from holding objectivity as a value.)

In the language of the general public, open-mindedness means being open to possible theories and explanations for a particular phenomenon. But in science it means that and something more. Philosopher Jonathan Adler (1998) teaches us that science values another aspect of open-mindedness even more highly: "What truly marks an open-minded person is the willingness to follow where evidence leads. The open-minded person is willing to defer to impartial investigations rather than to his own predilections...Scientific method is attunement to the world, not to ourselves" (p. 44).

Objectivity is critical to the process of science, but it does not mean that such attitudes must characterize each and every scientist for science as a whole to work. Jacob Bronowski (1973, 1977) often argued that the unique power of science to reveal knowledge about the world does not arise because scientists are uniquely virtuous (that they are completely objective or that they are never biased in interpreting findings, for example). It arises because fallible scientists are immersed in a process of *checks and balances*—a process in which scientists are always there to criticize and to root out errors. Philosopher Daniel Dennett (1999/2000) points out that "scientists take themselves to be just as weak and fallible as anybody else, but recognizing those very sources of error in themselves...they have devised elaborate systems to tie their own hands, forcibly preventing their frailties and prejudices from infecting their results" (p. 42). More humorously, psychologist Ray Nickerson (1998) makes the related point that the vanities of scientists are actually put to use by the scientific process, by noting that it is "not so much the critical attitude that individual scientists have taken with respect to their own ideas that has given science its success...but more the fact that individual scientists have been highly motivated to demonstrate that hypotheses that are held by some other scientists are false" (p. 32). These authors suggest that the strength of scientific knowledge comes not because scientists are virtuous, but from the social process where scientists constantly cross-check each others' knowledge and conclusions.

The public criteria of peer review and replication of findings exist in part to keep checks on the objectivity of individual scientists. Individuals cannot hide bias and nonobjectivity by personalizing their claims and keeping them from public scrutiny. Science does not accept findings that have failed the tests of replication and peer review precisely because it wants to ensure that all findings in science are in the public domain, as defined above. Purveyors of pseudoscientific educational practices fail the test of objectivity and are often identifiable by their attempts to do an "end run" around the public mechanisms of science by avoiding established peer review mechanisms and the information-sharing mechanisms that make replication possible. Instead, they attempt to promulgate their findings directly to consumers, such as teachers.

THE PRINCIPLE OF CONVERGING EVIDENCE

The principle of converging evidence has been well illustrated in the controversies surrounding the teaching of reading. The methods of systematic empiricism employed in the study of reading acquisition are many and varied. They include case studies, correlational studies, experimental studies, narratives, quasi-experimental studies, surveys, epidemiological studies and many others. The results of many of these studies have been synthesized in several important research syntheses (Adams, 1990; Ehri et al., 2001; National Reading Panel, 2000; Pressley, 1998; Rayner et al., 2002; Reading Coherence Initiative, 1999; Share & Stanovich, 1995; Snow, Burns, & Griffin, 1998; Snowling, 2000; Spear-Swerling & Sternberg, 2001; Stanovich, 2000). These studies were used in a process of establishing converging evidence, a principle that governs the drawing of the conclusion that a particular educational practice is research-based.

The principle of converging evidence is applied in situations requiring a judgment about where the "preponderance of evidence" points. Most areas of science contain competing theories. The extent to which a particular study can be seen as uniquely supporting one particular theory depends on whether other competing explanations have been ruled out. A particular experimental result is never equally relevant to all competing theories. An experiment may be a very strong test of one or two alternative theories but a weak test of others. Thus, research is considered highly convergent when a series of experiments consistently supports a given

theory while collectively eliminating the most important competing explanations. Although no single experiment can rule out all alternative explanations, taken collectively, a series of partially diagnostic experiments can lead to a strong conclusion if the data converge.

Contrast this idea of converging evidence with the mistaken view that a problem in science can be solved with a single, crucial experiment, or that a single critical insight can advance theory and overturn all previous knowledge. This view of scientific progress fits nicely with the operation of the news media, in which history is tracked by presenting separate, disconnected "events" in bite-sized units. This is a gross misunderstanding of scientific progress and, if taken too seriously, leads to misconceptions about how conclusions are reached about research-based practices.

One experiment rarely decides an issue, supporting one theory and ruling out all others. Issues are most often decided when the community of scientists gradually begins to agree that the *preponderance* of evidence supports one alternative theory rather than another. Scientists do not evaluate data from a single experiment that has finally been designed in the perfect way. They most often evaluate data from dozens of experiments, each containing some flaws but providing part of the answer.

Although there are many ways in which an experiment can go wrong (or become *confounded*), a scientist with experience working on a particular problem usually has a good idea of what most of the critical factors are, and there are usually only a few. The idea of converging evidence tells us to examine the pattern of flaws running through the research literature because the nature of this pattern can either support or undermine the conclusions that we might draw.

For example, suppose that the findings from a number of different experiments were largely consistent in supporting a particular conclusion. Given the imperfect nature of experiments, we would evaluate the extent and nature of the flaws in these studies. If all the experiments were flawed in a similar way, this circumstance would undermine confidence in the conclusions drawn from them because the consistency of the outcome may simply have resulted from a particular, consistent flaw. On the other hand, if all the experiments were flawed in *different* ways, our confidence in the conclusions increases because it is less likely that the consistency in the results was due to a contaminating factor that confounded all the

experiments. As Anderson and Anderson (1996) note, "When a conceptual hypothesis survives many potential falsifications based on different sets of assumptions, we have a robust effect." (p. 742).

Suppose that five different theoretical summaries (call them A, B, C, D, and E) of a given set of phenomena exist at one time and are investigated in a series of experiments. Suppose that one set of experiments represents a strong test of theories A, B, and C, and that the data largely refute theories A and B and support C. Imagine also that another set of experiments is a particularly strong test of theories C, D, and E, and that the data largely refute theories D and E and support C. In such a situation, we would have strong converging evidence for theory C. Not only do we have data supportive of theory C, but we have data that contradict its major competitors. Note that no one experiment tests all the theories, but taken together, the entire set of experiments allows a strong inference.

In contrast, if the two sets of experiments each represent strong tests of B, C, and E, and the data strongly support C and refute B and E, the overall support for theory C would be less strong than in our previous example. The reason is that, although data supporting theory C have been generated, there is no strong evidence ruling out two viable alternative theories (A and D). Thus research is highly convergent when a series of experiments consistently supports a given theory while collectively eliminating the most important competing explanations. Although no single experiment can rule out all alternative explanations, taken collectively, a series of partially diagnostic experiments can lead to a strong conclusion if the data converge in the manner of our first example.

Increasingly, the combining of evidence from disparate studies to form a conclusion is being done more formally by the use of the statistical technique termed meta-analysis (Cooper & Hedges, 1994; Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Rosenthal, 1995; Schmidt, 1992; Swanson, 1999) which has been used extensively to establish whether various medical practices are research based. In a medical context, meta-analysis:

> involves adding together the data from many clinical trials to create a single pool of data big enough to eliminate much of the statistical uncertainty that plagues individual trials...The great virtue of meta-analysis is that clear findings can emerge from a group of studies whose findings are scattered all over the map. (Plotkin,1996, p. 70)

The use of meta-analysis for determining the research validation of educational practices is just the same as in medicine. The effects obtained when one practice is compared against another are expressed in a common statistical metric that allows comparison of effects across studies. The findings are then statistically amalgamated in some standard ways (Cooper & Hedges, 1994; Hedges & Olkin, 1985; Swanson, 1999) and a conclusion about differential efficacy is reached if the amalgamation process passes certain statistical criteria. In some cases, of course, no conclusion can be drawn with confidence, and the result of the meta-analysis is inconclusive.

More and more commentators on the educational research literature are calling for a greater emphasis on meta-analysis as a way of dampening the contentious disputes about conflicting studies that plague education and other behavioral sciences (Kavale & Forness, 1995; Rosnow & Rosenthal, 1989; Schmidt, 1996; Stanovich, 2001; Swanson, 1999). The method is useful for ending disputes that seem to be nothing more than a "he-said, she-said" debate. An emphasis on meta-analysis has often revealed that we actually have more stable and useful findings than is apparent from a perusal of the conflicts in our journals.

The National Reading Panel (2000) found just this in their meta-analysis of the evidence surrounding several issues in reading education. For example, they concluded that the results of a meta-analysis of the results of 66 comparisons from 38 different studies indicated "solid support for the conclusion that systematic phonics instruction makes a bigger contribution to children's growth in reading than alternative programs providing unsystematic or no phonics instruction" (p. 2-84). In another section of their report, the National Reading Panel reported that a meta-analysis of 52 studies of phonemic awareness training indicated that "teaching children to manipulate the sounds in language helps them learn to read. Across the various conditions of teaching, testing, and participant characteristics, the effect sizes were all significantly greater than chance and ranged from large to small, with the majority in the moderate range. Effects of phonemic awareness training on reading lasted well beyond the end of training" (p. 2-5).

A statement by a task force of the American Psychological Association (Wilkinson, 1999) on statistical methods in psychology journals provides an apt summary for this section. The task force stated that investigators should not "interpret a single study's results as having importance independent of the effects reported elsewhere in the relevant literature" (p. 602).

Science progresses by convergence upon conclusions. The outcomes of one study can only be interpreted in the context of the present state of the convergence on the particular issue in question.

THE LOGIC OF THE EXPERIMENTAL METHOD

Scientific thinking is based on the ideas of *comparison, control,* and *manipulation.* In a true experimental study, these characteristics of scientific investigation must be arranged to work in concert.

Comparison alone is not enough to justify a causal inference. In methodology texts, correlational investigations (which involve comparison only) are distinguished from true experimental investigations that warrant much stronger causal inferences because they involve comparison, control, and manipulation. The mere existence of a relationship between two variables does not guarantee that changes in one are causing changes in the other. *Correlation does not imply causation.*

There are two potential problems with drawing causal inferences from correlational evidence. The first is called the third-variable problem. It occurs when the correlation between the two variables does not indicate a direct causal path between them but arises because both variables are related to a third variable that has not even been measured.

The second reason is called the directionality problem. It creates potential interpretive difficulties because even if two variables have a direct causal relationship, the direction of that relationship is not indicated by the mere presence of the correlation. In short, a correlation between variables A and B could arise because changes in A are causing changes in B or because changes in B are causing changes in A. The mere presence of the correlation does not allow us to decide between these two possibilities.

The heart of the experimental method lies in manipulation and control. In contrast to a correlational study, where the investigator simply observes whether the natural fluctuation in two variables displays a relationship, the investigator in a true experiment manipulates the variable thought to be the cause (the *independent variable*) and looks for an effect on the

variable thought to be the effect (the *dependent variable*) while holding all other variables constant by control and randomization. This method removes the third-variable problem because, in the natural world, many different things are related. The experimental method may be viewed as a way of prying apart these naturally occurring relationships. It does so because it isolates one particular variable (the hypothesized cause) by manipulating it and holding everything else constant (control).

When manipulation is combined with a procedure known as *random assignment* (in which the subjects themselves do not determine which experimental condition they will be in but, instead, are randomly assigned to one of the experimental groups), scientists can rule out alternative explanations of data patterns. By using manipulation, experimental control, and random assignment, investigators construct stronger comparisons so that the outcome eliminates alternative theories and explanations.

THE NEED FOR BOTH CORRELATIONAL METHODS AND TRUE EXPERIMENTS

As strong as they are methodologically, studies employing true experimental logic are not the only type that can be used to draw conclusions. Correlational studies have value. The results from many different types of investigation, including correlational studies, can be amalgamated to derive a general conclusion. The basis for conclusion rests on the convergence observed from the variety of methods used. This is most certainly true in classroom and curriculum research. It is necessary to amalgamate the results from not only experimental investigations, but correlational studies, nonequivalent control group studies, time series designs, and various other quasi-experimental designs and multivariate correlational designs, All have their strengths and weaknesses. For example, it is often (but not always) the case that experimental investigations are high in internal validity, but limited in external validity, whereas correlational studies are often high in external validity, but low in internal validity.

Internal validity concerns whether we can infer a causal effect for a particular variable. The more a study employs the logic of a true experiment (i.e., includes manipulation, control, and randomization), the more we can make a strong causal inference. External validity concerns

the generalizability of the conclusion to the population and setting of interest. Internal and external validity are often traded off across different methodologies. Experimental laboratory investigations are high in internal validity but may not fully address concerns about external validity. Field classroom investigations, on the other hand, are often quite high in external validity but because of the logistical difficulties involved in carrying them out, they are often quite low in internal validity. That is why we need to look for a convergence of results, not just consistency from one method. Convergence increases our confidence in the external and internal validity of our conclusions.

Again, this underscores why correlational studies can contribute to knowledge. First, some variables simply cannot be manipulated for ethical reasons (for instance, human malnutrition or physical disabilities). Other variables, such as birth order, sex, and age, are inherently correlational because they cannot be manipulated, and therefore the scientific knowledge concerning them must be based on correlational evidence. Finally, logistical difficulties in classroom and curriculum research often make it impossible to achieve the logic of the true experiment. However, this circumstance is not unique to educational or psychological research. Astronomers obviously cannot manipulate all the variables affecting the objects they study, yet they are able to arrive at conclusions.

Complex correlational techniques are essential in the absence of experimental research because complex correlational statistics such as multiple regression, path analysis, and structural equation modeling that allow for the partial control of third variables when those variables can be measured. These statistics allow us to recalculate the correlation between two variables after the influence of other variables is removed. If a potential third variable can be measured, complex correlational statistics can help us determine whether that third variable is determining the relationship. These correlational statistics and designs help to rule out certain causal hypotheses, even if they cannot demonstrate the true causal relation definitively.

Stages of Scientific Investigation: The Role of Case Studies and Qualitative Investigations

The educational literature includes many qualitative investigations that focus less on issues of causal explanation and variable control and more on *thick description,* in the manner of the anthropologist (Geertz, 1973, 1979). The context of a person's behavior is described as much as possible from the standpoint of the participant. Many different fields (e.g., anthropology, psychology, education) contain case studies where the focus is detailed description and contextualization of the situation of a single participant (or very few participants).

The usefulness of case studies and qualitative investigations is strongly determined by how far scientific investigation has advanced in a particular area. The insights gained from case studies or qualitative investigations may be quite useful in the early stages of an investigation of a certain problem. They can help us determine which variables deserve more intense study by drawing attention to heretofore unrecognized aspects of a person's behavior and by suggesting how understanding of behavior might be sharpened by incorporating the participant's perspective.

However, when we move from the early stages of scientific investigation, where case studies may be very useful, to the more mature stages of theory testing—where adjudicating between causal explanations is the main task—the situation changes drastically. Case studies and qualitative description are not useful at the later stages of scientific investigation because they cannot be used to confirm or disconfirm a particular causal theory. They lack the comparative information necessary to rule out alternative explanations.

Where qualitative investigations are useful relates strongly to a distinction in philosophy of science between the context of *discovery* and the context of *justification*. Qualitative research, case studies, and clinical observations support a context of discovery where, as Levin and O'Donnell (2000) note in an educational context, such research must be regarded as "preliminary/exploratory, observational, hypothesis generating" (p. 26). They rightly point to the essential importance of qualitative investigations because "in the early stages of inquiry into a research topic, one has to look before one can leap into designing interventions, making predictions, or testing hypotheses" (p. 26). The orientation provided by qualitative

investigations is critical in such cases. Even more important, the results of quantitative investigations—which must sometimes abstract away some of the contextual features of a situation—are often contextualized by the thick situational description provided by qualitative work.

However, in the context of justification, variables must be measured precisely, large groups must be tested to make sure the conclusion generalizes and, most importantly, many variables must be controlled because alternative causal explanations must be ruled out. Gersten (2001) summarizes the value of qualitative research accurately when he says that "despite the rich insights they often provide, descriptive studies cannot be used as evidence for an intervention's efficacy...descriptive research can only suggest innovative strategies to teach students and lay the groundwork for development of such strategies" (p. 47). Qualitative research does, however, help to identify fruitful directions for future experimental studies.

Nevertheless, here is why the sole reliance on qualitative techniques to determine the effectiveness of curricula and instructional strategies has become problematic. As a researcher, you desire to do one of two things.

Objective A

The researcher wishes to make some type of statement about a relationship, however minimal. That is, you at least want to use terms like *greater* than, or *less* than, or *equal* to. You want to say that such and such an educational program or practice is *better* than another. "Better than" and "worse than" are, of course, quantitative statements—and, in the context of issues about what *leads to* or *fosters* greater educational achievement, *they are causal statements as well.* As quantitative causal statements, the support for such claims obviously must be found in the experimental logic that has been outlined above. To justify such statements, you must adhere to the canons of quantitative research logic.

Objective B

The researcher seeks to adhere to an exclusively qualitative path that abjures statements about relationships and never uses comparative terms of magnitude. The investigator desires to simply engage in thick description of a domain that may well prompt hypotheses when later

work moves on to the more quantitative methods that are necessary to justify a causal inference.

Investigators pursuing Objective B are doing essential work. They provide quantitative information with suggestions for richer hypotheses to study. In education, however, investigators sometimes claim to be pursuing Objective B but slide over into Objective A without realizing they have made a crucial switch. They want to make comparative, or quantitative, statements, but have not carried out the proper types of investigation to justify them. They want to say that a certain educational program is better than another (that is, it *causes* better school outcomes). They want to give educational strictures that are assumed to hold for a population of students, not just to the single or few individuals who were the objects of the qualitative study. They want to condemn an educational practice (and, by inference, deem an alternative quantitatively and causally better). But instead of taking the necessary course of pursuing Objective A, they carry out their investigation in the manner of Objective B.

Let's recall why the use of single case or qualitative description as evidence in support of a particular causal explanation is inappropriate. The idea of alternative explanations is critical to an understanding of theory testing. The goal of experimental design is to structure events so that support of one particular explanation simultaneously disconfirms other explanations. Scientific progress can occur only if the data that are collected rule out some explanations. Science sets up conditions for the natural selection of ideas. Some survive empirical testing and others do not.

This is the honing process by which ideas are sifted so that those that contain the most truth are found. But there must be selection in this process: data collected as support for a particular theory must not leave many other alternative explanations as equally viable candidates. For this reason, scientists construct control or comparison groups in their experimentation. These groups are formed so that, when their results are compared with those from an experimental group, some alternative explanations are ruled out.

Case studies and qualitative description lack the comparative information necessary to prove that a particular theory or educational practice is superior, because they fail to test an alternative; they rule nothing out. Take the seminal work of Jean Piaget for example. His case

studies were critical in pointing developmental psychology in new and important directions, but many of his theoretical conclusions and causal explanations did not hold up in controlled experiments (Bjorklund, 1995; Goswami, 1998; Siegler, 1991).

In summary, as educational psychologist Richard Mayer (2000) notes, "the domain of science includes both some quantitative and qualitative methodologies" (p. 39), and the key is to use each where it is most effective (see Kamil, 1995). Likewise, in their recent book on research-based best practices in comprehension instruction, Block and Pressley (2002) argue that future progress in understanding how comprehension works will depend on a healthy interaction between qualitative and quantitative approaches. They point out that getting an initial idea of the comprehension processes involved in hypertext and Web-based environments will involve detailed descriptive studies using think-alouds and assessments of qualitative decision making. Qualitative studies of real reading environments will set the stage for more controlled investigations of causal hypotheses.

THE PROGRESSION TO MORE POWERFUL METHODS

A final useful concept is the progression to more powerful research methods ("more powerful" in this context meaning more diagnostic of a causal explanation). Research on a particular problem often proceeds from weaker methods (ones less likely to yield a causal explanation) to ones that allow stronger causal inferences. For example, interest in a particular hypothesis may originally emerge from a particular case study of unusual interest. This is the proper role for case studies: to suggest hypotheses for further study with more powerful techniques and to motivate scientists to apply more rigorous methods to a research problem. Thus, following the case studies, researchers often undertake correlational investigations to verify whether the link between variables is real rather than the result of the peculiarities of a few case studies. If the correlational studies support the relationship between relevant variables, then researchers will attempt experiments in which variables are manipulated in order to isolate a causal relationship between the variables.

SUMMARY OF PRINCIPLES THAT SUPPORT RESEARCH-BASED INFERENCES ABOUT BEST PRACTICE

Our sketch of the principles that support research-based inferences about best practice in education has revealed that:

- Science progresses by investigating solvable, or testable, empirical problems.
- To be testable, a theory must yield predictions that could possible be shown to be wrong.
- The concepts in the theories in science evolve as evidence accumulates. Scientific knowledge is not infallible knowledge, but knowledge that has at least passed some minimal tests. The theories behind research-based practice can be proven wrong, and therefore they contain a mechanism for growth and advancement.
- Theories are tested by systematic empiricism. The data obtained from empirical research are in the public domain in the sense that they are presented in a manner that allows replication and criticism by other scientists.
- Data and theories in science are considered in the public domain only after publication in peer-reviewed scientific journals.
- Empiricism is systematic because it strives for the logic of control and manipulation that characterizes a true experiment.
- Correlational techniques are helpful when the logic of an experiment cannot be approximated, but because these techniques only help rule out hypotheses, they are considered weaker than true experimental methods.
- Researchers use many different methods to arrive at their conclusions, and the strengths and weaknesses of these methods vary. Most often, conclusions are drawn only after a slow accumulation of data from many studies.

Scientific Thinking in Educational Practice: Reason-based Practice in the Absence of Direct Evidence

Some areas in educational research, to date, lack a research-based consensus, for a number of reasons. Perhaps the problem or issue has not been researched extensively. Perhaps research into the issue is in the early stages of investigation, where descriptive studies are suggesting interesting avenues, but no controlled research justifying a causal inference has been completed. Perhaps many correlational studies and experiments *have* been conducted on the issue, but the research evidence has not yet converged in a consistent direction.

Even if teachers know the principles of scientific evaluation described earlier, the research literature sometimes fails to give them clear direction. They will have to fall back on their own reasoning processes as informed by their own teaching experiences. In those cases, teachers still have many ways of reasoning scientifically.

Tracing the Link from Scientific Research to Scientific Thinking in Practice

Scientific thinking in can be done in several ways. Earlier we discussed different types of professional publications that teachers can read to improve their practice. The most important defining feature of these outlets is whether they are peer reviewed. Another defining feature is whether the publication contains primary research rather than presenting opinion pieces or essays on educational issues. If a journal presents primary research, we can evaluate the research using the formal scientific principles outlined above.

If the journal is presenting opinion pieces about what constitutes best practice, we need to trace the link between those opinions and archival peer-reviewed research. We would look to see whether the authors have based their opinions on peer-reviewed research by reading the reference list. Do the authors provide a significant amount of original research citations (is their opinion based on more than one study)? Do the authors cite work other than their own (have the results been replicated)? Are the cited journals peer-reviewed? For example, in the

case of best practice for reading instruction, if we came across an article in an opinion-oriented journal such as *Intervention in School and Clinic*, we might look to see if the authors have cited work that has appeared in such peer-reviewed journals as *Journal of Educational Psychology, Elementary School Journal, Journal of Literacy Research, Scientific Studies of Reading,* or *the Journal of Learning Disabilities.*

These same evaluative criteria can be applied to presenters at professional development workshops or papers given at conferences. Are they conversant with primary research in the area on which they are presenting? Can they provide evidence for their methods and does that evidence represent a scientific consensus? Do they understand what is required to justify causal statements? Are they open to the possibility that their claims could be proven false? What evidence would cause them to shift their thinking?

An important principle of scientific evaluation—the connectivity principle (Stanovich, 2001)—can be generalized to scientific thinking in the classroom. Suppose a teacher comes upon a new teaching method, curriculum component, or process. The method is advertised as totally new, which provides an explanation for the lack of direct empirical evidence for the method. A lack of direct empirical evidence should be grounds for suspicion, but should not immediately rule it out. The principle of connectivity means that the teacher now has another question to ask: "OK, there is no direct evidence for this method, but how is the theory behind it (the causal model of the effects it has) connected to the research consensus in the literature surrounding this curriculum area?" Even in the absence of direct empirical evidence on a particular method or technique, there could be a theoretical link to the consensus in the existing literature that would support the method.

For further tips on translating research into classroom practice, see Warby, Greene, Higgins, & Lovitt (1999). They present a format for selecting, reading, and evaluating research articles, and then importing the knowledge gained into the classroom.

Let's take an imaginary example from the domain of treatments for children with extreme reading difficulties. Imagine two treatments have been introduced to a teacher. No direct empirical tests of efficacy have been carried out using either treatment. The first, Treatment A, is a training program to facilitate the awareness of the segmental nature of language at the phonological level. The second, Treatment B, involves giving children training in vestibular sensitivity by having them walk on balance beams while blindfolded. Treatment A and B are

equal in one respect—neither has had a direct empirical test of its efficacy, which reflects badly on both. Nevertheless, one of the treatments has the edge when it comes to the principle of connectivity. Treatment A makes contact with a broad consensus in the research literature that children with extraordinary reading difficulties are hampered because of insufficiently developed awareness of the segmental structure of language. Treatment B is not connected to any corresponding research literature consensus. Reason dictates that Treatment A is a better choice, even though neither has been directly tested.

Direct connections with research-based evidence and use of the connectivity principle when direct empirical evidence is absent give us necessary cross-checks on some of the pitfalls that arise when we rely solely on personal experience. Drawing upon personal experience is necessary and desirable in a veteran teacher, but it is not sufficient for making critical judgments about the effectiveness of an instructional strategy or curriculum. The insufficiency of personal experience becomes clear if we consider that the educational judgments—even of veteran teachers—often are in conflict. That is why we have to adjudicate conflicting knowledge claims using the scientific method.

Let us consider two further examples that demonstrate why we need controlled experimentation to verify even the most seemingly definitive personal observations. In the 1990s, considerable media and professional attention were directed at a method for aiding the communicative capacity of autistic individuals. This method is called facilitated communication. Autistic individuals who had previously been nonverbal were reported to have typed highly literate messages on a keyboard when their hands and arms were supported over the typewriter by a so-called facilitator. These startlingly verbal performances by autistic children who had previously shown very limited linguistic behavior raised incredible hopes among many parents of autistic children.

Unfortunately, claims for the efficacy of facilitated communication were disseminated by many media outlets before any controlled studies had been conducted. Since then, many studies have appeared in journals in speech science, linguistics, and psychology and each study has unequivocally demonstrated the same thing: the autistic child's performance is dependent upon tactile cueing from the facilitator. In the experiments, it was shown that when both child and facilitator were looking at the same drawing, the child typed the correct name of the drawing. When the viewing was occluded so that the child and the facilitator were shown

different drawings, the child typed the name of the facilitator's drawing, not the one that the child herself was looking at (Beck & Pirovano, 1996; Burgess, Kirsch, Shane, Niederauer, Graham, & Bacon, 1998; Hudson, Melita, & Arnold, 1993; Jacobson, Mulick, & Schwartz, 1995; Wheeler, Jacobson, Paglieri, & Schwartz, 1993). The experimental studies directly contradicted the extensive case studies of the experiences of the facilitators of the children. These individuals invariably deny that they have inadvertently cued the children. Their personal experience, honest and heartfelt though it is, suggests the wrong model for explaining this outcome. The case study evidence told us something about the social connections between the children and their facilitators. But that is something different than what we got from the controlled experimental studies, which provided direct tests of the claim that the technique unlocks hidden linguistic skills in these children. Even if the claim had turned out to be true, the verification of the proof of its truth would not have come from the case studies or personal experiences, but from the necessary controlled studies.

Another example of the need for controlled experimentation to test the insights gleaned from personal experience is provided by the concept of learning styles—the idea that various modality preferences (or variants of this theme in terms of analytic/holistic processing or "learning styles") will interact with instructional methods, allowing teachers to individualize learning. The idea seems to "feel right" to many of us. It does seem to have some face validity, but it has never been demonstrated to work in practice. Its modern incarnation (see Gersten, 2001, Spear-Swerling & Sternberg, 2001) takes a particularly harmful form, one where students identified as auditory learners are matched with phonics instruction and visual and/or kinesthetic learners matched with holistic instruction. The newest form is particularly troublesome because the major syntheses of reading research demonstrate that many children can benefit from phonics-based instruction, not just "auditory" learners (National Reading Panel, 2000; Rayner et al., 2002; Stanovich, 2000). Excluding students identified as "visual/kinesthetic" learners from effective phonics instruction is a bad instructional practice—bad because it is not only not research based, it is actually contradicted by research.

A thorough review of the literature by Arter and Jenkins (1979) found no consistent evidence for the idea that modality strengths and weaknesses could be identified in a reliable and valid way that warranted differential instructional prescriptions. A review of the research evidence by Tarver and Dawson (1978) found likewise that the idea of modality preferences did not

hold up to empirical scrutiny. They concluded, "This review found no evidence supporting an interaction between modality preference and method of teaching reading" (p. 17). Kampwirth and Bates (1980) confirmed the conclusions of the earlier reviews, although they stated their conclusions a little more baldly: "Given the rather general acceptance of this idea, and its common-sense appeal, one would presume that there exists a body of evidence to support it. Unfortunately...no such firm evidence exists" (p. 598).

More recently, the idea of modality preferences (also referred to as learning styles, holistic versus analytic processing styles, and right versus left hemispheric processing) has again surfaced in the reading community. The focus of the recent implementations refers more to teaching to strengths, as opposed to remediating weaknesses (the latter being more the focus of the earlier efforts in the learning disabilities field). The research of the 1980s was summarized in an article by Steven Stahl (1988). His conclusions are largely negative because his review of the literature indicates that the methods that have been used in actual implementations of the learning styles idea have not been validated. Stahl concludes: "As intuitively appealing as this notion of matching instruction with learning style may be, past research has turned up little evidence supporting the claim that different teaching methods are more or less effective for children with different reading styles" (p. 317).

Obviously, such research reviews cannot prove that there is no possible implementation of the idea of learning styles that could work. However, the burden of proof in science rests on the investigator who is making a new claim about the nature of the world. It is not incumbent upon critics of a particular claim to show that it "couldn't be true." The question teachers might ask is, "Have the advocates for this new technique provided sufficient proof that it works?" Their burden of responsibility is to provide proof that their favored methods work. Teachers should not allow curricular advocates to avoid this responsibility by introducing confusion about where the burden of proof lies. For example, it is totally inappropriate and illogical to ask "Has anyone proved that it can't work?" One does not "prove a negative" in science. Instead, hypotheses are stated, and then must be tested by those asserting the hypotheses.

Reason-based practice in the classroom

Effective teachers engage in scientific thinking in their classrooms in a variety of ways: when they assess and evaluate student performance, develop Individual Education Plans (IEPs) for their students with disabilities, reflect on their practice, or engage in action research. For example, consider the assessment and evaluation activities in which teachers engage. The scientific mechanisms of systematic empiricism—iterative testing of hypotheses that are revised after the collection of data—can be seen when teachers plan for instruction: they evaluate their students' previous knowledge, develop hypotheses about the best methods for attaining lesson objectives, develop a teaching plan based on those hypotheses, observe the results, and base further instruction on the evidence collected.

This assessment cycle looks even more like the scientific method when teachers (as part of a multidisciplinary team) are developing and implementing an IEP for a student with a disability. The team must assess and evaluate the student's learning strengths and difficulties, develop hypotheses about the learning problems, select curriculum goals and objectives, base instruction on the hypotheses and the goals selected, teach, and evaluate the outcomes of that teaching. If the teaching is successful (goals and objectives are attained), the cycle continues with new goals. If the teaching has been unsuccessful (goals and objectives have not been achieved), the cycle begins again with new hypotheses. We can also see the principle of converging evidence here. No one piece of evidence might be decisive, but collectively the evidence might strongly point in one direction.

Scientific thinking in practice occurs when teachers engage in action research. Action research is research into one's own practice that has, as its main aim, the improvement of that practice. Stokes (1997) discusses how many advances in science came about as a result of "use-inspired research" which draws upon observations in applied settings. According to McNiff, Lomax, and Whitehead (1996), action research shares several characteristics with other types of research: "it leads to knowledge, it provides evidence to support this knowledge, it makes explicit the process of enquiry through which knowledge emerges, and it links new knowledge with existing knowledge" (p. 14). Notice the links to several important concepts: systematic empiricism, publicly verifiable knowledge, converging evidence, and the connectivity principle.

Teachers and Research
Commonality in a "What Works" Epistemology

Many educational researchers have drawn attention to the epistemological commonalities between researchers and teachers (Gersten, Vaughn, Deshler, & Schiller, 1997; Stanovich, 1993/1994). A "what works" epistemology is a critical source of underlying unity in the world views of educators and researchers (Gersten & Dimino, 2001; Gersten, Chard, & Baker, 2000). Empiricism, broadly construed (as opposed to the caricature of white coats, numbers, and test tubes that is often used to discredit scientists) is about watching the world, manipulating it when possible, observing outcomes, and trying to associate outcomes with features observed and with manipulations. This is what the best teachers do. And this is true despite the grain of truth in the statement that "teaching is an art." As Berliner (1987) notes: "No one I know denies the artistic component to teaching. I now think, however, that such artistry should be research-based. I view medicine as an art, but I recognize that without its close ties to science it would be without success, status, or power in our society. Teaching, like medicine, is an art that also can be greatly enhanced by developing a close relationship to science (p. 4)."

In his review of the work of the Committee on the Prevention of Reading Difficulties for the National Research Council of the National Academy of Sciences (Snow, Burns, & Griffin, 1998), Pearson (1999) warned educators that resisting evaluation by hiding behind the "art of teaching" defense will eventually threaten teacher autonomy. Teachers need creativity, but they also need to demonstrate that they know what evidence is, and that they recognize that they practice in a profession based in behavioral science. While making it absolutely clear that he opposes legislative mandates, Pearson (1999) cautions:

> We have a professional responsibility to forge best practice out of the raw materials provided by our most current and most valid readings of research...If professional groups wish to retain the privileges of teacher prerogative and choice that we value so dearly, then the price we must pay is constant attention to new knowledge as a vehicle for fine-tuning our individual and collective views of best practice. This is the path that other professions, such as medicine, have taken in order to maintain their professional prerogative, and we must take it, too. My fear is that if the professional groups in education fail to assume this responsibility

> squarely and openly, then we will find ourselves victims of the most onerous of legislative mandates (p. 245).

Those hostile to a research-based approach to educational practice like to imply that the insights of teachers and those of researchers conflict. Nothing could be farther from the truth. Take reading, for example. Teachers often do observe exactly what the research shows—that most of their children who are struggling with reading have trouble decoding words. In an address to the Reading Hall of Fame at the 1996 meeting of the International Reading Association, Isabel Beck (1996) illustrated this point by reviewing her own intellectual history (see Beck, 1998, for an archival version). She relates her surprise upon coming as an experienced teacher to the Learning Research and Development Center at the University of Pittsburgh and finding "that there were some people there (psychologists) who had not taught anyone to read, yet they were able to describe phenomena that I had observed in the course of teaching reading" (Beck, 1996, p. 5). In fact, what Beck was observing was the triangulation of two empirical approaches to the same issue—two perspectives on the same underlying reality. And she also came to appreciate how these two perspectives fit together: "What I knew were a number of whats—what some kids, and indeed adults, do in the early course of learning to read. And what the psychologists knew were some whys—why some novice readers might do what they do" (pp. 5-6).

Beck speculates on why the disputes about early reading instruction have dragged on so long without resolution and posits that it is due to the power of a particular kind of evidence—evidence from personal observation. The determination of whole language advocates is no doubt sustained because "people keep noticing the fact that some children or perhaps many children—in any event a subset of children—especially those who grow up in print-rich environments, don't seem to need much more of a boost in learning to read than to have their questions answered and to point things out to them in the course of dealing with books and various other authentic literacy acts" (Beck, 1996, p. 8). But Beck points out that it is equally true that proponents of the importance of decoding skills are also fueled by personal observation: "People keep noticing the fact that some children or perhaps many children—in any event a subset of children—don't seem to figure out the alphabetic principle, let alone some of the intricacies involved without having the system directly and systematically presented" (p. 8). But clearly we have lost sight of the basic fact that the two observations are

not mutually exclusive—one doesn't negate the other. This is just the type of situation for which the scientific method was invented: a situation requiring a consensual view, triangulated across differing observations by different observers.

Teachers, like scientists, are ruthless pragmatists (Gersten & Dimino, 2001; Gersten, Chard, & Baker, 2000). They believe that some explanations and methods are better than others. They think there is a real world out there—a world in flux, obviously—but still one that is trackable by triangulating observations and observers. They believe that there are valid, if fallible, ways of finding out which educational practices are best. Teachers believe in a world that is predictable and controllable by manipulations that they use in their professional practice, just as scientists do. Researchers and educators are kindred spirits in their approach to knowledge, an important fact that can be used to forge a coalition to bring hard-won research knowledge to light in the classroom.

References

Adams, M. J. (1990). Beginning to read: Thinking and learning about print. Cambridge, MA: MIT Press.

Adler, J. E. (1998, January). Open minds and the argument from ignorance. Skeptical Inquirer, 22(1), 41-44.

Anderson, C. A., & Anderson, K. B. (1996). Violent crime rate studies in philosophical context: A destructive testing approach to heat and Southern culture of violence effects. Journal of Personality and Social Psychology, 70, 740–756.

Anderson, R. C., Hiebert, E. H., Scott, J., & Wilkinson, I. (1985). Becoming a nation of readers. Washington, D. C.: National Institute of Education.

Arter, A. and Jenkins, J. (1979). Differential diagnosis-prescriptive teaching: A critical appraisal, Review of Educational Research, 49, 517-555.

Beck, A. R., & Pirovano, C. M. (1996). Facilitated communications' performance on a task of receptive language with children and youth with autism. Journal of Autism and Developmental Disorders, 26, 497-512.

Beck, I. L. (1996, April). Discovering reading research: Why I didn't go to law school. Paper presented at the Reading Hall of Fame, International Reading Association, New Orleans.

Beck, I. (1998). Understanding beginning reading: A journey through teaching and research. In J. Osborn & F. Lehr (Eds.), Literacy for all: Issues in teaching and learning (pp. 11-31). New York: Guilford Press.

Berliner, D. C. (1987). Knowledge is power: A talk to teachers about a revolution in the teaching profession. In D. C. Berliner & B. V. Rosenshine (Eds.), Talks to teachers (pp. 3-33). New York: Random House.

Bjorklund, D. F. (1995). Children's thinking: Developmental function and individual differences (Second Edition). Pacific Grove, CA: Brooks/Cole.

Block, C. C., & Pressley, M. (Eds.). (2002). Comprehension instruction: Research-based best practices. New York: Guilford Press.

Bronowski, J. (1956). Science and human values. New York: Harper & Row.

Bronowski, J. (1973). The ascent of man. Boston: Little, Brown.

Bronowski, J. (1977). A sense of the future. Cambridge: MIT Press.

Burgess, C. A., Kirsch, I., Shane, H., Niederauer, K., Graham, S., & Bacon, A. (1998). Facilitated communication as an ideomotor response. Psychological Science, 9, 71-74.

Chard, D. J., & Osborn, J. (1999). Phonics and word recognition in early reading programs: Guidelines for accessibility. Learning Disabilities Research & Practice, 14, 107-117.

Cooper, H. & Hedges, L. V. (Eds.), (1994). The handbook of research synthesis. New York: Russell Sage Foundation.

Cunningham, P. M., & Allington, R. L. (1994). Classrooms that work: They all can read and write. New York: HarperCollins.

Dawkins, R. (1998). Unweaving the rainbow. Boston: Houghton Mifflin.

Dennett, D. C. (1995). Darwin's dangerous idea: Evolution and the meanings of life. New York: Simon & Schuster.

Dennett, D. C. (1999/2000, Winter). Why getting it right matters. Free Inquiry, 20(1), 40-43.

Ehri, L. C., Nunes, S., Stahl, S., & Willows, D. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's Meta-Analysis. Review of Educational Research, 71, 393-447.

Foster, E. A., Jobling, M. A., Taylor, P. G., Donnelly, P., Deknijff, P., Renemieremet, J., Zerjal, T., & Tyler-Smith, C. (1998). Jefferson fathered slave's last child. Nature, 396, 27-28.

Fraenkel, J. R., & Wallen, N. R. (1996). How to design and evaluate research in education (Third Edition). New York: McGraw-Hill.

Geertz, C. (1973). The interpretation of cultures. New York: Basic Books.

Geertz, C. (1979). From the native's point of view: On the nature of anthropological understanding. In P. Rabinow & W. Sullivan (Eds.), Interpretive social science (pp. 225-242). Berkeley: University of California Press.

Gersten, R. (2001). Sorting out the roles of research in the improvement of practice. Learning Disabilities: Research & Practice, 16(1), 45-50.

Gersten, R., Chard, D., & Baker, S. (2000). Factors enhancing sustained use of research-based instructional practices. Journal of Learning Disabilities, 33(5), 445-457.

Gersten, R., & Dimino, J. (2001). The realities of translating research into classroom practice. Learning Disabilities: Research & Practice, 16(2), 120-130.

Gersten, R., Vaughn, S., Deshler, D., & Schiller, E. (1997).What we know about using research findings: Implications for improving special education practice. Journal of Learning Disabilities, 30(5), 466-476.

Goswami, U. (1998). Cognition in children. Hove, England: Psychology Press.

Gross, P. R., Levitt, N., & Lewis, M. (1997). The flight from science and reason. New York: New York Academy of Science.

Hedges, L. V., & Olkin, I. (1985). Statistical Methods for Meta-Analysis. New York: Academic Press.

Holton, G., & Roller, D. (1958). Foundations of modern physical science. Reading, MA: Addison-Wesley.

Hudson, A., Melita, B., & Arnold, N. (1993). A case study assessing the validity of facilitated communication. Journal of Autism and Developmental Disorders, 23, 165–173.

Hunter, J. E., & Schmidt, F. L. (1990). Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage.

Jacobson, J. W., Mulick, J. A., & Schwartz, A. A. (1995). A history of facilitated communication: Science, pseudoscience, and antiscience. American Psychologist, 50, 750–765.

Kamil, M. L. (1995). Some alternatives to paradigm wars in literacy research. Journal of Reading Behavior, 27, 243-261.

Kampwirth, R., and Bates, E. (1980). Modality preference and teaching method: A review of the research, Academic Therapy, 15, 597-605.

Kavale, K. A., & Forness, S. R. (1995). The nature of learning disabilities: Critical elements of diagnosis and classification. Mahweh, NJ: Lawrence Erlbaum Associates.

Levin, J. R., & O'Donnell, A. M. (2000). What to do about educational research's credibility gaps? Issues in Education: Contributions from Educational Psychology, 5, 1-87.

Liberman, A. M. (1999). The reading researcher and the reading teacher need the right theory of speech. Scientific Studies of Reading, 3, 95-111.

Magee, B. (1985). Philosophy and the real world: An introduction to Karl Popper. LaSalle, IL: Open Court.

Mayer, R. E. (2000). What is the place of science in educational research? Educational Researcher, 29(6), 38-39.

McNiff, J.,Lomax, P., & Whitehead, J. (1996). You and your action research project. London: Routledge.

Medawar, P. B. (1982). Pluto's republic. Oxford: Oxford University Press.

Medawar, P. B. (1984). The limits of science. New York: Harper & Row.

Medawar, P. B. (1990). The threat and the glory. New York: Harper Collins.

Moats, L. (1999). Teaching reading is rocket science. Washington, DC: American Federation of Teachers.

National Reading Panel: Reports of the Subgroups. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 2, 175-220.

Pearson, P. D. (1993). Teaching and learning to read: A research perspective. Language Arts, 70, 502-511.

Pearson, P. D. (1999). A historically based review of preventing reading difficulties in young children. Reading Research Quarterly, 34, 231-246.

Plotkin, D. (1996, June). Good news and bad news about breast cancer. Atlantic Monthly, 53–82.

Popper, K. R. (1972). Objective knowledge. Oxford: Oxford University Press.

Pressley, M. (1998). Reading instruction that works: The case for balanced teaching. New York: Guilford Press.

Pressley, M., Rankin, J., & Yokol, L. (1996). A survey of the instructional practices of outstanding primary-level literacy teachers. Elementary School Journal, 96, 363-384.

Rayner, K. (1998). Eye movements in reading and information processing: 20 Years of research. Psychological Bulletin, 124, 372-422.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2002, March). How should reading be taught? Scientific American, 286(3), 84-91.

Reading Coherence Initiative. (1999). Understanding reading: What research says about how children learn to read. Austin, TX: Southwest Educational Development Laboratory.

Rosenthal, R. (1995). Writing meta-analytic reviews. Psychological Bulletin, 118, 183-192.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276-1284.

Shankweiler, D. (1999). Words to meaning. Scientific Studies of Reading, 3, 113-127.

Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. Issues in Education: Contributions from Educational Psychology, 1, 1-57.

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

Siegler, R. S. (1991). Children's thinking (Second Edition). Englewood Cliffs, NJ: Prentice Hall.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press.

Snowling, M. (2000). Dyslexia (Second Edition). Oxford: Blackwell.

Spear-Swerling, L., & Sternberg, R. J. (2001). What science offers teachers of reading. Learning Disabilities: Research & Practice, 16(1), 51-57.

Stahl, S. (December, 1988). Is there evidence to support matching reading styles and initial reading methods? Phi Delta Kappan, 317-327.

Stanovich, K. E. (1993/1994). Romance and reality. The Reading Teacher, 47(4), 280-291.

Stanovich, K. E. (2000). Progress in understanding reading: Scientific foundations and new frontiers. New York: Guilford Press.

Stanovich, K. E. (2001). How to think straight about psychology (Sixth Edition). Boston: Allyn & Bacon.

Stokes, D. E. (1997). Pasteur's quadrant: Basic science and technological innovation. Washington, DC: Brookings Institution Press.

Tarver, S. G., & Dawson, E. (1978). Modality preference and the teaching of reading: A review, Journal of Learning Disabilities, 11, 17-29.

Swanson, H. L. (1999). Interventions for students with learning disabilities: A meta-analysis of treatment outcomes. New York: Guilford Press.

Vaughn, S., & Dammann, J. E. (2001). Science and sanity in special education. Behavioral Disorders, 27, 21-29.

Warby, D. B., Greene, M. T., Higgins, K., & Lovitt, T. C. (1999). Suggestions for translating research into classroom practices. Intervention in School and Clinic, 34(4), 205-211.

Wheeler, D. L., Jacobson, J. W., Paglieri, R. A., & Schwartz, A. A. (1993). An experimental assessment of facilitated communication. Mental Retardation, 31, 49–60.

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54, 595-604.

Wilson, E. O. (1998). Consilience: The unity of knowledge. New York: Knopf.

For additional copies of this document:

Download PDF or HTML versions at
www.nifl.gov/partnershipforreading

or

Contact the National Institute for Literacy at ED Pubs
PO Box 1398, Jessup, Maryland 20794-1398

Phone 1-800-228-8813
Fax 301-430-1244

edpubs@inet.ed.gov