Self-referent Research

MICHAEL SCRIVEN
Evaluation Institute, University of San Francisco

I

I will discuss a number of items that fall under the heading of self-referent research, broadly conceived. This will include not only some comments on the state of our research on educational research, but also some comments on research dealing with vehicles of educational research, for example, the printed word, the machinery for its processing and production, and the journals which select and package it.

Studies of research practice of the proportion of women presenters at AERA compared to the proportion of women members, or studies of the relative popularity of different research methods, for example, studies of the spread of ANOVA methodology in published educational research articles, constitute one kind of research on research that might be called descriptive metaresearch. This should be distinguished from R & D work on or evaluations of the machinery of research dissemination (e.g., ERIC or microforms, or videodiscs or communicating word processors). In addition to descriptive, R & D, and evaluative research on research, there is policy research on, for example, the policies of research journals. Educational research covers a broad area. Although I cannot guarantee the absence of a text or anthology of research on the many aspects of educational research, I am confident that there is no substantial number of such volumes, or large body of journal articles. However, there certainly have been some excellent individual studies and an occasional series. This paper does not report on such research, but I will discuss its importance and how it might be done or improved.

Why is it important? A case can be made that a categorical imperative, an obligation of professionalism, should require us to treat this as a priority duty. We are all aware of the protracted failure of the great universities to inspire, support or reward research on their own teaching procedures, their own raison d’etre, a failure which is only being rectified because of external pressure. We should also bear in mind the failure of large professional associations like American Medical Association (AMA) to look at themselves until prodded from the outside. The shocking incongruity between the number of well-qualified woman researchers and the number of women who were principal investigators on federal research grants and contracts was sitting there, a tangible, accusatory fact, for many decades without a strong protest from AERA or the American Psychological Association. Are there other skeletons in the closet? We will not know till we look much closer. I suppose my favorite entry in the Skeleton Stakes would be the continued failure to look seriously at the peer review process both in proposal and in personnel evaluation. This is not specifically an educational research problem—it affects all researchers as well as teachers—but it is the key process whereby educational research is funded and rewarded, and it is still done as a group version of personal opinion instead of the highly sophisticated management process that it should and can quite easily be. To convert it from one to the other, one needs calibration training of raters, a registry of rater validity (determined from follow-ups), and rate baselines, standardized sequences (should the group discussions come before written preliminary ratings or after, etc.), anchor definitions of points in the rating scales, standardization of rating scales (to a reasonable degree), and half a dozen other fairly obvious features. On the cost side, can we continue to respond only to external pressure to upgrade these key professional activities in these cost-conscious times? Why are there no cost-effectiveness evaluations going on: of the alternative sizes of reviewer panels; of systemic efforts to replace review panels...
with linear regression equations (following the lessons learned in
the clinical vs. statistical prediction
research); using teleconferencing
rather than traveling for the peer panel discussions; using
two-tier approaches as standard
operating procedure (i.e., a
two to five-page preliminary pro-
posal, with minor funding support
for winners to prepare full pro-
posals); testing fast versus delib-
erate rating; developing indexes
of cost-effectiveness in research,
and so on? What about investiga-
tion of applicants who were re-
jected, to get some idea of the
false negative rate? What about
investigating the several rad-
cially different possible procedures
for coping with large differences
in the preliminary and final rat-
ings by different raters on the
same panels (resolution by discus-
sion; use of a referee; use of back-
up raters, etc.)?

Far be it from me to suggest
that the most productive use of
educational researchers’ time is
to look at educational research.
Philosophers who do nothing but
discuss the nature of philosophy
make little contribution to either
society or intellect. But huge ef-
forts at educational research (or
philosophical research) must be
attended by much smaller but
still significant efforts to improve
the process for doing it, and fund-
ing it, and implementing it; and
we do very little of this.

Take another aspect of profes-
sional activity which could stand
much more serious analysis than
it has ever received. The accredi-
tation process is used by many
professional organizations and in
virtually all states to accredit
training programs for profes-
sionals, in education and else-
where. We recently made a seri-
sous effort to get together informa-
tion on all of those systems about
which such information is public.
It was depressing to discover that
not only are none of them very
good, but 95% of them are unnec-
essarily subject to almost every
flaw and fallacy identified in ele-
mentary texts on research or
evaluation methodology. These
include, most noticeably, a disas-
trous failure to look at the right
criterion variables (they always
look at process instead of outcome
variables); essentially, total fail-
ure to look at interpanel reliabil-
ity; no effort to avoid halo effects,
no standardization of language or
procedures, no calibration train-
ing or testing of participants, no
systematic study of the indirect
costs or side-effects of programs,
no follow-up, no critical review of
the process by clients or their rep-
resentatives—or no visibility for
such reviews, and so on. And this
is the process of accreditation by
professional associations (or by
professional staff of state over-
sight agencies) used for judging
the programs that educate new
professionals! So the way we fund
research, and the way we dis-
pense our professional training
are both shoddy examples of pro-
fessional practice by the stan-
dards of our own textbooks. And
the reason is simple; we own that
turf, or we have until recently,
and the last house to be put in or-
der is one’s own. I fear that the
rest of the world has caught on to
this, and that the big universities
and the big applied research con-
tractors and the big school sys-
tems and professional organiza-
tions have just about lost their
credibility, largely and simply for
the excellent reason that it has
been discovered they did not de-
serve it. But even with our backs
very close to the wall, what have
we done? Where are the con-
structive responses to the new threats?
Every now and then a defensive
report comes out—but where are
the proposed tough new guide-
lines that protect the public inter-
est, the critical studies solicited
by professional groups of their
own practice, done by indepen-
dents? I am glad to say that there
have been one or two small efforts
of which the evaluation standards
project, led by Dan Stufflebeam,
with NIE, NSF and Lilly Support,
is an excellent example.

When one is elected president
of a professional association one
supposes that the chance has
come to set such matters right.
Next, one notices that, through
some incomprehensible failure of
logistics, the position does not au-
tomatically guarantee a year’s in-
stant sabbatical. After that about
all one notices is the steadily ris-
ing pile of unanswered mail. The
largest contribution I will be
making on this issue is this ad-
dress. If I can make you feel as I
do that we are letting the ship
sink slowly because we are not
raising our eyes from our regular
jobs, then next time you serve on
a review panel or a site visiting
team or as a paper referee, or
when you start thinking about
your next research project, you
may act differently, look more
critically at what you are doing,
make a small commitment to up-
grade these powerful processes
just slightly. We need to raise our
consciousness about the stan-
dards embodied in our own group
practices. They are not the stan-
dards we espouse rhetorically or
pedagogically, and there is no ex-
cuse for that. It is more difficult,
at first, to do these jobs right, to
develop the training and testing
materials for peer review panels,
than to drift along in the same old
way; but we must shed the coun-
try-club membership committee
approach in favor of a self-aware,
self-critical, and systematic ap-
proach or there will not be a coun-
try club.

I have discussed our traditional
institutions, the academies and
guilds and entrepreneurs; every-
thing I have said applies to the
agencies and the foundations,
perhaps more so. I doubt if anyone
could be more sympathetic with
the difficulties of their situation,
with the extremes of political
pressure and bureaucratic bun-
gling which impede their
progress. But I am not very sym-
pathetic with their solutions.
When McCarthy or Proxmire
screams, they appoint a short-
lived committee or a consultant
to

Educational Researcher
review whatever he is screaming about. What we lack are committee to upgrade process before someone screams, simply because they need upgrading by any reasonable professional standards. Where, indeed, are the committees to continue to upgrade process after the screaming has died down? We had a little spell of casual investigation of the peer review process after Proxmire screamed, but how many long-term studies have been done which were designed with the kind of minimum care and attention that we give to long-term stu-
as in either case there is an external evaluation of the results, and publication of the evaluations and reactions. No sound-minded manufacturer would research and develop a new product with the incompetence we show in researching and developing a credentialing or research-funding process. It is at the research on research, training of trainers or research on funding—the meta-level—level that we fail. How are matters on the ground floor with straight educational R & D? Not bad, but not exactly perfect.

I ask you to look back on the dies of the effect of food additives on mice? When you have a long track record of doing nothing to set your own house in order, it takes more than a minimum effort to overcome scepticism from Congress and the general citizenry. Yet, instead of using the principle of compensatory effort, we use instead the principle of minimal effort; act like you are doing something remedial for the shortest possible time, and then go back to your knitting. A serious compensatory effort would mean, for example, the systematic gathering and publication of criticisms, the systematic search for unmet needs, the systematic application of competitive bidding in order to get improvements running on a trial basis—or, if that can be done in-house, fine; as long as usual record of the new curricula. We have far more than $100,000 invested in these projects, perhaps $500,000 including indirect costs, and they were developed by a great many of our most experienced and distinguished colleagues, working with (in the early days) more than ample resources. A cynical view of it all, popular with some members of the general public, is that the main long-term contribution of the great curriculum effort was to decrease literacy. Unfair though that may be, it contains a grain of truth that should cause us to pause, for the new curricula did pull time out of core curriculum.

Let me try to explain why I think we are more to blame than most of us believe. Blame accrues, we must remember, not just to those who could have changed matters by their own unaided effort, but to all those who, had they all acted differently, would then jointly have made a difference. Blame is never reduced one iota either by the identification of others who were equally guilty, or by the identification of others without whose help no change could have been made and who showed no interest in change. The obligation on us is to make our maximum effort, in the hope that our example and our influence will change enough others to form an adequate force for change. In short, it is an obligation to bang one's head against walls because it is well known that walls give way before the banging of a thousand heads.

It may be the time for educational research to take a lead in matters of professional duty, to bang its head first. Such a lead is one that our subject matter and the best of our results have earned. We have too often seen ourselves as followers, as workers on areas the application of which are results from other areas. It is most significant that the social sciences, as commonly listed, do not include education, nor do the humanities. It tells us something about status. I would like to see us become a model for other areas, a source for results that transform them, instead of an area for the application of their results. What educational research knows about learning is far more significant for the reform of psychology than is the content of present day psychology for education. Yet, we continue to believe the opposite, to look for guidance from over there, to lament instead of capitalize upon the perpetual victory of the null hypothesis in our studies. We must stand science on its head—but carefully! We must learn to identify and search for significant negative results by contrast with mere negative results. We continue the quest for panaceas instead of prophylactics, for concep-

April 1980
tually simple treatments instead of complex, possibly indescribable but learnable treatments. Instead of studying the possibility of emulating great teachers, we study the possibility of fractionating them. Surely style research has shown that this does not work; the secret is not in the components but in the complex combinations of theory, interacting with task and teacher variables, probably in an unspecified way, but a way that is quite possibly learnable. Most visual recognition skills (e.g., learning to recognize pictures of Einstein or pictures by Eisenstadt) are not decomposable into teachable elements that assemble into the macroskill, but they are easily taught/learned holistically. (This does not mean the whole is more than the sum of its parts; just that it is more than the sum of any parts which we have the vocabulary to describe.) Are we not using the wrong paradigm in our research? Are we ever looking hard enough at our paradigms? We could do worse than to study such questions more seriously than the social sciences have studied them. We might begin to lead with leadership in self-referent research: It is an area where the competition is not severe, though that is changing with the increasing interest in atrogenic research.

Let us consider the example of curriculum reform and materials development in a little more detail. What did we achieve there? Was it a success or a failure of educational research? First, we must clarify the question slightly, for it is often confused with the question of whether we made a mark, had an influence, left substantial changes behind (a confusion that invalidates many of the studies of the impact of research on educational practice). The important question is whether what we did was clearly superior to what would have been done without us. The Federal and many local projects got in and did something, perhaps even met the need. But would it have been met much worse by the private sector publishers, possibly with very little help from the educational researchers? This is the key question both for the evaluation of the Federal interventions and for the evaluation of the contribution of educational research. It is a question that needs an answer if we are to clarify our own sense of worth, and also, more importantly, if we are to make the best decision next time. Well, has it been thoroughly studied; do we have the answer? Most of the research I have seen on the school impact of these curricula only addresses the question of whether there was one, sometimes also the question whether the impact was valuable, but not the question of whether it was essential. One cannot feel too pleased about spending a week on an overload effort to clean up the litter in Yosemite Park, an effort for which one can ill afford the time, if one later discovers that the Park Service would have done it a week later. Well, the publishers have one view of our curriculum contribution and we have another: It is not easy to say who is right. It would be interesting to see the matter seriously studied and the results seriously discussed since on the outcome of that meta research rests the appropriateness of future curricular commitments by the Federal government or by the research community.

The other half of the cynic’s challenge on curricular impact relates to the opportunity costs of the big curriculum projects. Did all that effort divert attention from the basics into “fringe areas”? Did it divert resources? Did it divert expertise? It is hard to say no to all of these questions, and that creates the real possibility that even if what we did in physics and biology and social studies was both valuable and indispensable, it still was too costly, and that we do, in fact, share substantially in the blame for the extent of functional illiteracy today.

Is it absurd to think that every member of AERA should be at least moderately well-informed about issues like this, even if not this issue? Should we not have some knowledge and sense of responsibility about the adequacy of our contribution to the most important educational needs? Educational research is viewed with a great lack of respect in a great many circles, often for absurd reasons such as it has not yet discovered the basic laws of learning. But I would feel less uneasy when these allegations come up if I had access to a substantial body of sustained research on the pay-offs from educational research. We have recently seen the publication of the National Academy of Education study concerned with what at first appears to be this issue. But the study turns out to be quite peripheral, concerned with such questions as whether the results of basic research in psychology and the social sciences trickle down to the educational research journals (which, you will not be surprised to hear, they do). That does not show that educational research benefitted from that other research and, more importantly, it does not show that educational research benefitted anything else.

It may be thought that this line of suggestions is all too naive in suggesting that we should do a certain kind of research on research, when the simple fact is that we do not have the tools for that kind of research. After all, it is one thing to show that certain curricula are used, but how do you decide what other curricula would have existed if these had not? I think that we have the tools to handle that kind of question and I have thought about it in some detail. But suppose we lack such tools. Does not that fact, carefully studied, lead to important conclusions? Since the answer to these questions is so important, should we not make an effort to change the conditions of our interventions, in the future at

Educational Researcher
least, so we will be able to answer such questions? The scenario would be simple enough. Instead of jumping into a possibly freezing swimming pool with both feet, we do frequent tests with our big toe. That is, we do head-to-head runoffs between small samples of the best efforts of commercial publishers, who are recovering their development costs from users, and federal- or state-funded efforts with their much greater use of educational researchers and much larger development costs provided from general tax resources. Results would have, and should have, very limited generalizability, so we would have to adopt this kind of limited competitive piloting as a standard preliminary R & D practice.

Compare that with the federal decision to fund a complete K-12 curriculum at one research and development center, to be done by unknown writers over a decade or more, or with the recent sole-source spending of millions (not only federal) into Children’s Television Workshop to get a “Science Sesame Street” going. Where is the track record data to justify either the sole-sourcing or the large sum? It all points the other way; the only Sesame Street evaluation shows very slight gains under conditions that totally precluded validity and favored false positive results. Where is the evidence that the Sullivan readers or the Cohen reading materials—to mention an early and a recent private venture—are inferior to the federal efforts with their much larger and more expensive commitment to pooled research expertise? It turns out there is a virtual ban on using Federal dollars to fund such comparisons. What does that tell you? It is hard to believe that we are even looking at the big picture at all, let alone acting on it, let alone standing up and fighting to insist that it be considered. I should stress that this line of argument does not stem from any great faith in the genius of free enterprise, just skepticism about the superiority of ill-considered government enterprise. I am sure we need federal R & D efforts, but equally clear, the present ones have uncertain credentials and block the pipeline.

Of course, if you look at what happens to those who stand up and fight to get the big picture considered, you cannot fail to conclude that the establishment has its defense mechanisms pretty well established, even if not its creative efforts. The mechanisms work like this. The fighter is either allowed to exhaust her or himself by struggling with successive levels of the bureaucracy; or is tolerated until time and other troubles take over; or is bought off with a small grant; or—in the unlikely event of partial success—the fourth strategy is unlimbered, namely, to create a task force from which, of course, the troublemaker is excluded on the grounds of bias and which therefore contains no one with the motivation and often no one with the capability for handling the problem. It is no news to students of bureaucracy that this happens, but saying it is expectable is no excuse for letting it continue. Anti-Semitism is expectable in certain social contexts in this country, as is prejudice against blacks and women. A first step in changing such things is to bell the cat, to point the finger, to make the charges explicit in order to force public response, to raise consciousness. We hardly need to be told that this strategy works; it has been used by special interests often enough. Is it too much to ask that we point the finger in the interest of defensible standards and not just in our own interests? I cannot help feeling that an association which will not do that is something less than professional—just as much as one that will not respond to the legitimate demands of minorities and women for fair treatment. I think we should require a task force, a plan and a report to members once a year on these self-evaluative issues, with funded self-referent research where appropriate. There might be a useful opportunity for a novel cooperative effort in which a foundation or agency joined in with some funding and external criticism, and received our report on ourselves plus some external criticism of their own procedures in return. This is no pleasant, self-congratulatory self-review that I am talking about; it is clear from the preceding discussion that we are disgracefully at fault and the main question is whether we will and how best to improve, not whether we should. If I receive some expressions of interest on this, especially from those members (and nonmembers) who have done some of the existing self-referent research, I will (1) propose the task force to the Association’s Governing Board and (2) get a small anthology published (nominations for inclusions welcomed). It is too bad to leave this field open to the kind of performance that we have just seen from Nader’s Raiders.

There is nothing “holier than thou” about this. I should have done more on these issues during my year as President than I did; and with more opportunity than most members, I am simply more at fault. But whatever a president of this Association can do is worth little without the support of many, many members. So the bad news is that no one person can save us: it will take many of us. The good news is that it does not take a great effort from each of us to do it. The most important effort is to raise our consciousness. Once you start applying research standards to research itself (and to its connected processes) you can never turn back, any more than you can now quell the pang that you feel when you hear someone say “All men are created equal.” In the next section, we will turn to a discussion of research methodology.