The Fairy Godmother—and Her Warts

Making the Dream of Evidence-Based Policy Come True

Carol H. Weiss
Harvard Graduate School of Education
Erin Murphy-Graham
University of California
Anthony Petrosino
WestED
Allison G. Gandhi
American Institutes for Research

Abstract: Evaluators sometimes wish for a Fairy Godmother who would make decision makers pay attention to evaluation findings when choosing programs to implement. The U.S. Department of Education came close to creating such a Fairy Godmother when it required school districts to choose drug abuse prevention programs only if their effectiveness was supported by "scientific" evidence. The experience showed advantages of such a procedure (e.g., reduction in support for D.A.R.E., which evaluation had found wanting) but also shortcomings (limited and in some cases questionable evaluation evidence in support of other programs). Federal procedures for identifying successful programs appeared biased. In addition, the Fairy Godmother discounted the professional judgment of local educators and did little to improve the fit of programs to local conditions. Nevertheless, giving evaluation more clout is a worthwhile way to increase the rationality of decision making. The authors recommend research on procedures used by other agencies to achieve similar aims.

Keywords: *evidence-based policy; Drug Abuse Resistance Education (D.A.R.E.); research utilization; influence of evaluation*

A widespread assumption is that research and evaluation should influence the making of public policy. In recent years, this belief has flourished under the label of *evidence-based policy*. It holds that the findings of research[1] should help policy makers develop wise policy that is based in strong understanding of current conditions. It calls on policy makers to learn what research has to say before they dive into the swamps of the policy process.

Evaluation is a type of research that should be particularly well used because it deals in the currency of policy and programs. Evaluation gives direction about which interventions have had good effects and which kinds of interventions have been less successful. Therefore, an

Carol H. Weiss, Ph.D., Beatrice Whiting Professor of Education, Harvard Graduate School of Education, Gutman 467 Appian Way, Cambridge, MA 02138.

Authors' Note: Robert Wood Johnson Foundation Grant 039080 to the Study on Decisions in Education at the Harvard Graduate School of Education provided funding for this research.

American Journal of Evaluation, Vol. 29 No. 1, March 2008 29-47
DOI: 10.1177/1098214007313742
© 2008 American Evaluation Association

alert policy maker should derive guidance on which road to travel in the future. But despite the logic of attending to evaluation findings when developing or amending policies, evaluations do not always have influence. With a few notable exceptions (e.g., Gueron & Pauly, 1991; Henig, 2007; Schweinhart, Barnes, & Weikart, 1993), many investigations in many countries have shown that evaluation has usually had only modest influence on policy and practice (Baklien, 1983; Becker, 1984; Breslau, 1998; Bulmer, 1986; Furubo, 1994; Lampinen, 1992). Most studies seem to be used in selective bits, reinterpreted to fit existing preferences, or ignored (Lemieux-Charles & Champagne, 2004; Rosenstock & Lee, 2002).

For a long time, evaluators' reaction to this situation was to moan and lament the neglect of evaluation. In some government agencies, the reaction was to coax and cajole attention to evaluation, offer incentives, appeal to rationality, and hoist the evidence-based policy banner. But evidence-based policy was more an aspiration than a reality.

Within the past decade, a number of government agencies have adopted an innovative strategy to impose the use of evaluation evidence on local policy makers and practitioners. This "imposed use" (Weiss et al., 2005) requires applicants for federal program funds to show that the program they wish to run has been scientifically evaluated and found successful. Funding is contingent on evaluation evidence. Regulations of this type have been adopted in such fields as violence prevention, teenage pregnancy prevention, reduction of HIV risk behaviors, and prevention of drug abuse. The federal government mandates that applicants for program grants demonstrate that the programs they plan to run are "scientifically proven."

Thus, federal agencies are requiring that research and evaluation gain a strong voice in policy. Their strategy of "imposed use" has become a Fairy Godmother, apparently making wishes come true for greater evaluation influence on policy. She waves her wand, and lo, policy makers refer to data from evaluation.

A study that we conducted on drug abuse prevention programming has given us an opportunity to put the Fairy Godmother to the test: Does imposed use work? We found that when evaluation evidence of success is required for program funding, a fair number of local school districts pay attention to evaluation. They turn to programs that evaluation has suggested are effective. The federal agency with its magic wand of program funding is obliging school districts to take evidence seriously. The story looks like a victory for rational decision making. However, when we look closer, we see some problems. Imposed use is not a clear winner.

This article is the account of one effort to give greater priority to evaluation evidence and the slip-ups that marred the attempt. The article first reviews the literature on the relationship between evaluation and policy. Then, it discusses the findings of our own study on the influence of evaluation on school-based drug abuse prevention programs. Then, we use our findings to analyze the potency of the Fairy Godmother's wand. In other words, we take advantage of our data to examine whether "imposed use" advances the cause of evidence in policy making. What we find is that the Fairy Godmother has warts on her face. We try to answer the question of whether the effort is worth the problems it stirs up. After all, having warts is not a terminal disease.

Previous Literature on the Relationship Between Research and Policy

Social scientists' belief that they have a contribution to make to public policy and professional practice is not a new phenomenon. More than 300 years ago, Thomas Hobbes expressed the hope that his book *Leviathan* would some day "fall into the hands of a sovereign" who would "convert the truth of speculation into the utility of practice" (cited in Sowell, 1980,

p. 341). Malthus, Marx, and Herbert Spencer were among the many early social scientists who expected their work to influence policy.

Early in the 20th century, but picking up speed and intensity after World War II, social science organizations repeatedly called on the federal government to make increased use of the knowledge that the social sciences have to offer. The National Academy of Sciences and other organizations issued reports to that effect beginning in the 1960s (e.g., National Academy of Sciences, 1968; National Science Foundation, 1968). In many policy areas, such as economics, health, and foreign affairs, social scientists were already working with government agencies (for a comprehensive review of the relationships of social science and the federal government up until that point, see Lyons, 1969).

The War on Poverty was a major spur to the salience of research and evaluation. Just about every grant for programming—in education, job training, social services, crime and delinquency, health, mental health, housing, and other social areas—came with an "Evaluate Me" tag. A new profession was born: evaluator. Soon, there were journals, professional associations, university courses, textbooks, and scores of evaluation firms and thousands of consultants.

However, despite all the exhortations, admonitions, and blue-ribbon reports, there is widespread agreement that social science research has not generally had major influence on policy or practice. Lyons's 1969 book on the relations between government and social science is titled *The Uneasy Partnership*. Anderson writes of a "vexed" relationship between the two realms that is characterized by tension and ambiguity (Anderson, 2003, p. 103). Others have called the relationship haphazard, imperfect, and troubled (Lindblom & Cohen, 1979; Weiss, 1995).

Researchers have analyzed the many reasons why research has had spotty success in influencing both policy and practice in fields as diverse as education, criminal justice, international development, and health care. Many of us have already reviewed the literature on "research utilization" or "research influence" (e.g., Alkin, Daillak, & White, 1979; Beyer & Trice, 1982; Cousins & Leithwood, 1986; Davies, Nutley, & Smith, 2000; Preskill & Torres, 1999; Shulha & Cousins, 1997; Walter, Nutley, & Davies, 2005; Walter, Nutley, Percy-Smith, McNeish, & Frost, 2004; Weiss, 1977). Here, we do a brief overview, with emphasis on the most recent additions to the literature.

The obstacles to productive collaboration between research and policy can be classified into three main categories: shortcomings in research and researchers, shortcomings in policy makers and practitioners, and shortcomings in the links between the two realms.

Shortcomings in Research and Researchers

Observers have acknowledged some of the intrinsic weaknesses in research and evaluation. Among the common complaints are untrustworthy evidence, unresponsiveness to decision makers' needs, fragmented data, and evidence that fails to cumulate or yields contradictory findings (Saunders, 2005). Some writers complain that evaluators are too responsive to their governmental sponsors and fail to look at outcomes from a program participant's perspective (Taylor, 2005). Others are concerned with the fact that like all social knowledge, evaluation findings are historically contingent. They emerge from a particular time and place, inhabited by certain people in particular patterns of interaction. Their findings do not represent eternal truths. As Mulgar (2005) recently noted, people and systems change, and as they do, audiences need to maintain a healthy skepticism about the validity of even the most robust research evidence.

Other writers have faulted the systems in which evaluators work. Owing to the dependence of many evaluators on contracts from government agencies, evaluators often work under limited time constraints with insufficient funds for good comparative design and longitudinal

follow-up. Again because of their dependence on funds from program agencies, evaluators are sometimes reluctant to report findings of shortfalls and program failures; their next evaluation contract may be put in jeopardy. St. Pierre and Puma (2000) urge more straightforward talk to officials in control.

As scholars have long noted, social science is replete with inconsistent and conflicting findings:

> As the twenty-first century unfolds, and scientific research in nearly every field is growing almost explosively, new findings daily "overthrow" old ones . . . findings are often confusing and conflicting about central issues of theory and practice . . . some [studies] show effects in one direction and some in the opposite, and some show effects that are close to zero. (Rosenthal & DiMatteo, 2001, p. 60)

In evaluation, evaluators often promote the application of findings about program effectiveness one study at a time. Instead, a synthesis of all good data on the program's effectiveness would probably be more valid and more convincing. This realization has led to an emphasis on meta-analysis. Meta-analysts combine the findings of multiple studies to reach balanced assessments of program success over a range of settings and times. However, recently, critics such as Pawson (2006) have criticized meta-analysis because it strips away most traces of program context. There is little information about the particulars of the situations in which the program was implemented or the ways in which it was run. A policy maker who wants to know how well a program fits his or her local environment will receive little help.

Of course, meta-analysis can examine the effects of context on program outcomes if the original studies provide data about such features as content, sponsorship, participants, staff, and environment that characterized the programs studied. When sufficient information on context is available in the study reports, meta-analysis is a powerful tool not only for synthesizing the outcomes of programs over multiple settings but also for identifying the features of context that are associated with better outcomes. However, many of the original evaluations do not report on contextual factors. Without data in the original reports, meta-analysis is unable to say anything useful about the effects of program context, and it is unclear why studies of different program sites yield inconsistent information on outcomes.

Dopson and Fitzgerald (2005), in their review of health care research and its effects on practice, emphasize that research has generally focused on individuals and ignored the organizational context of implementation. They stress that context is complex, multifaceted, and interactive, "not a backcloth to action." Research that does not pay attention to the context in which research results need to be implemented loses its appeal to potential actors. In education, a similar point is made by Simmons (2005), Pogrow (1996), and others. Simmons writes that the current emphasis on evidence of "what works" overlooks a second critical question: "What works, given the needs and values of my students and community and the condition and capacity of my school and district?" (p. 9; see also Pawson, 2006). Greenberg et al. (Greenberg, Mandell, & Onstott, 2000; Greenberg & Shroder, 2004) report on a study of the influence of welfare-to-work experiments on state welfare departments. They found support for the contention that knowledge of implementation is vital. They write that "much more important to most states than empirical evidence of treatment effects obtained through random assignment evaluations was information about how these programs actually operated in the field" (Greenberg et al., 2000, p. 379).

Shortcomings in the Context, Beliefs, and Actions of Policy Makers

A recent report from Scotland's Office of Chief Researcher, Knowledge Transfer Team (G. Clark & L. Kelly), Scottish Executive Social Research (2005) summarizes a set of assertions about the limitations of policy makers in dealing with research. In their experience, the policy community

> worked to unrealistic, short term timescales; frequently asked academics to "dumb down" their research findings; sometimes "moved the goal posts" half way through a research project; apparently ignored some research outcomes from projects; did not always take the time to read and analyze full academic outputs; sometimes lost interest in project or events when policy priorities changed. . . . (p. 20)

As an account of a major effort at collaborative work, the report indicates some successes, but the difficulties led to considerable disappointment.

Earlier summaries of the disconnect between research and policy included decision makers' sense that findings are inconclusive, their insistence on prompt reporting before full analysis has been completed, their focus on a big slogan–like finding and inattention to details, qualifications, and caveats.

How and when research has an influence on policy seems to depend largely on the political agenda and ideology of the government of the day rather than "the nature of the evidence, however compelling" (Bowen & Zwi, 2005, p. 602; Davies et al., 2000; Rigby, 2005; Shonkoff, 2000). Where research strives to bring order and clarity to human affairs, politics is sometimes better served by ambiguity. Mulgar (2005) notes that "full revelation is nice in theory but can be deeply destabilizing and destructive of self- and mutual respect. Where different groups have diametrically opposed views or interests, the assertion of rationality and evidence may have little impact" (p. 24).

Shortcomings in the Links Between Researchers, Policy Makers, and Practitioners

Most broadly, observers have stated that researchers and policy makers constitute two different communities. This "two-community" idea is one of the early explanations for the lack of uptake of research knowledge in the policy domain (e.g., Caplan, 1977). Researchers and policy makers differ in language (Bobrow, 2006; Lakoff, 1996), time horizon (Kemm, 2006), responsiveness to constituencies, written versus spoken communication (Weiss, 1989), concern with validity of evidence, attention to a preset agenda, and dozens of other elements in their days. Here is the argument again in 2003:

> There are fundamental differences between the research and policy communities that sometimes act as barriers to effective communication between the two, such as different time frames for results; different languages for communication; different priorities for knowledge; and lack of understanding about the realities of each other's work environments. (Nova Scotia Health Research Foundation, cited in Clark & Kelly [2005], p. 20)

Many authors recognize that communication of research results is essential but say that communication alone is not sufficient to get people to act (Walter et al., 2004). Echoing Knott and Wildavsky's 1980 article titled "If Dissemination Is the Answer, What Is the Problem?" Petersilia (1987), in her review of criminal justice research, reported upbeat conclusions about the influence of research on policy and practice but found that a weak point was assisting decision makers in implementing research findings. She concludes that potential users "need

direct and practical help to adapt policies and programs to their particular situations" (p. 104). Think tanks have become increasingly influential in public discussions because of aggressive marketing of their evidence and expertise (Rich, 2004).

Remedies for policy makers' inattention to research have been proposed. Some of the suggestions are increasing their capacity to understand and apply research findings (e.g., Boudett, City, & Murnane, 2005), developing partnerships between decision-making organizations and researchers, development of appropriate organizational structures to bring research and decision making closer together, and mobilization and allocation of sufficient resources to grease the wheels. Walter et al. (2005) did an outstanding systematic review of strategies for increasing the uptake of research. They conclude that interactive approaches such as partnerships between researchers and practitioners or policy makers are mildly effective, as are organizational and technical support and resources for adoption and implementation. Less successful are efforts at capacity building, leadership development, and structural change for bringing about greater reliance on evidence.

To sum up the recent literature, by and large, current writing supports earlier analyses of the research–policy nexus. One theme that has achieved greater prominence is the significance of the context in which a program is carried out. Despite considerable effort at bridging the gap between evidence and policy, rare is the study that leads to direct change in direction.

An Effort to Overcome the Obstacles to Evaluation Influence

The many obstacles to using social science findings in policy and practice have called forth new solutions. One that has recently appeared in the United States is "imposed use" (Weiss et al., 2005)—that is, a requirement by federal funding agencies that recipients use the grant to run only programs that have reported scientific evidence of success. Government funds will not go to support programs that do not have evidence of effectiveness.

Imposed use thus potentially deals with one main category of obstacles to use: inattention by policy makers to evaluation evidence. Policy makers may still be prone to ignore evaluation for any of the multiform reasons discussed in the literature. But now they have to pay attention or they do not receive federal funds. Although they may not have heeded evaluation on their own initiative, their dependence on federal grants for operating funds makes attention to evaluation more likely. Imposed use, requiring applicants to cite scientific evidence to justify their choice of program, has the potential to give research and evaluation greater influence on policy. That is why we are calling it the Fairy Godmother of research influence.

But a new issue arises: How are local agencies to know which programs are scientifically acceptable? Local agencies are often not equipped to find evidence on their own. If they do manage to locate program evaluations, they may find inconsistent and even contradictory data about the effectiveness of particular programs. They may latch on to a study that is flawed or superseded by later study and fail to examine issues of validity. Or they may be baffled by the complexities. To overcome such problems, at least one funding agency has undertaken to review research evidence itself. The federal office that funds drug abuse prevention programs in schools, with the help of a panel of outside experts, has reviewed evaluation evidence and drawn up a list of "scientifically proven" programs.

The creation of such a list alleviates several problems for local grantees. They do not need to search for scientific reports; they do not need to judge the scientific merit of the studies they find; they do not have to mediate among conflicting study results. In fact, state and local policy makers need not have read evaluation findings or even heard of evaluation studies.

The federal agency can draw up a list of programs that evaluation has found to be effective. The local grantee needs only to read the list of programs and pick a program that has met federal criteria.

The Case of Drug Abuse Prevention Programs

The following section gives a brief account of school-based drug abuse programs to see how imposed use works out in practice. We have written extensively about drug abuse prevention programs in this and other publications, based on an empirical study we did of 4 states, 16 school districts, and the state and federal agencies involved in the programming (Birkeland, Murphy-Graham, & Weiss, 2005; Gandhi, Murphy-Graham, Petrosino, Chrismer, & Weiss, 2007; Petrosino, Birkeland, Hacsi, Murphy-Graham, & Weiss, 2006; Weiss et al., 2005). Here, we recap those elements of the story that allow us to examine the extent to which "imposed use" of research serves to solve the problems of inert policy makers, inadequate dissemination, and poor-quality evidence.

In 1983, when crime and violence stemming from drug use had escalated to crisis levels, the Los Angeles Police Department decided that the best way to deal with the crisis was to prevent the start of drug abuse among young people. The chief of police, Darryl Gates, brought in a researcher employed by the Los Angeles school system to collaborate on developing a drug abuse prevention curriculum that police officers could teach in the public schools. The program was called Drug Abuse Resistance Education (D.A.R.E.). Trainers prepared L.A. police officers to offer the program to fifth- or sixth-grade classes in the schools.

The curriculum is highly structured and is offered usually for 45 to 60 min a week for one semester of 16 or 17 weeks. Each week covers one lesson in the curriculum. For example, in the original curriculum, Lesson 5 is about learning ways of saying no to drugs, and Lesson 6 is on building self-esteem. (The curriculum has been modified several times throughout the years.) Police officers receive about 90 hr of training in the program and are expected to follow the curriculum as written.

From its inception, the program was greeted with acclaim by parents, students, and teachers. Other police departments began clamoring for admission to the L.A. training sessions. Soon the D.A.R.E. program had spread to communities up and down California and then rapidly to other states. The federal government and a number of state governments began providing funds for the program; corporate sponsors, private organizations, and individuals contributed. By the early 1990s, D.A.R.E. was by far the most popular drug abuse prevention program in the country. The parent organization, D.A.R.E. America, which had separated from the Los Angeles Police Department and become a nonprofit organization, stated that by 1995 60% to 80% of all school districts in the country were running D.A.R.E. Beginning in 1990, every U.S. president declared a National Day of D.A.R.E. to support the program, and members of Congress passed legislation mentioning D.A.R.E.—the only program specifically named—as an eligible recipient of federal drug abuse prevention funds.

Evaluations of the D.A.R.E. Program

Early evaluations of the program found that students receiving D.A.R.E. in school did better than students who were not exposed to the program on measures of knowledge, attitudes, and self-reported drug use (e.g., DeJong, 1987). Later evaluations often found mixed results with positive findings for the students receiving D.A.R.E. on attitudes toward drugs and attitudes toward the police but not often on use of drugs. In 1994, researchers at the Research Triangle Institute did a meta-analysis of evaluations of D.A.R.E., limiting their

investigation to eight studies that used a randomized trial or a strong quasi-experimental design (Ennett, Tobler, Ringwalt, & Flewelling, 1994). The results showed that D.A.R.E. had positive impacts on knowledge but much smaller effects on attitudes and self-reported drug use. The researchers compared the results of the D.A.R.E. evaluations with those from a meta-analysis of other drug programs (Tobler, 1986). They concluded that programs that Tobler (1986) had characterized as "interactive" (including role-play and discussion rather than lecture and information only) did better than D.A.R.E. (Ennett et al., 1994).

Other evaluations began to report most of the evaluation activity centered on D.A.R.E. because of its overwhelming presence in the country. Some of the evaluations that came out in the mid-1990s used randomized control designs, with large samples of students, and followed up for long periods of time after the students had completed the D.A.R.E. program (Clayton et al., 1996; Dukes, Stein, & Ullman, 1997; Rosenbaum, Gordon, & Hanson, 1998). The collective result of these studies was to question the effect of D.A.R.E. in keeping youngsters off drugs. By the time students reached their high school years, those exposed to D.A.R.E. looked no different in their drug use behavior from students who had not been so exposed. For a program as popular as D.A.R.E. had become, this was equivalent to news that the emperor had no clothes.

The media had a field day. They gave the evaluation results wide coverage. The message was "D.A.R.E. doesn't work." Stories appeared in *The New York Times*, *Boston Globe*, *U.S. News and World Report*, and scores of other outlets. All the network news programs covered the story, some repeatedly. Television news magazines like *20/20* had segments on the evaluations. On ABC, for example, a reporter concluded the story with a statement that D.A.R.E. is wasting taxpayers' money.

Our Study

In 2000, we applied for and received a grant from the Robert Wood Johnson Foundation to study the influence of evaluation evidence on decision making. We had selected the case of D.A.R.E. for our study because at the time we wrote the proposal, it seemed that despite all the negative evidence, school districts were continuing their support of D.A.R.E. It seemed a strategic instance of ignoring evaluation evidence, and we were interested in understanding why.

We selected four states that had each been the site of a major evaluation of D.A.R.E. and four districts in each state. After extensive pilot tests of our interview protocols, in 2001, we did face-to-face interviews in these 16 districts with officials in the school system and the law enforcement agency involved in the choice of prevention program. We also conducted in-person interviews with relevant officials in the state agencies involved with prevention and in six federal agencies that had a say over federal drug abuse prevention activities. Interviews were taped, transcribed verbatim, coded, and analyzed using computer software. We wrote composite case studies of each district and each state and a chronological account and analysis of federal action. Two and a half years later, we did follow-up telephone interviews with officials in each of the 16 school districts to find out which prevention program they were currently running and why they had made the choice. Although not drawn on here, we also conducted interviews with 10 evaluators who did major studies and did an extensive document analysis of press reports about D.A.R.E. during a 20-year period.

The major financial support for D.A.R.E. and other school-based drug prevention programs was the Safe and Drug-Free Schools (SDFS) program run by the U.S. Department of Education. SDFS did not have the authority to decide which programs should be supported; most of the funds were passed through to the states for distribution. The head of the SDFS

was uneasy about how the money was spent. He wanted greater accountability for the allocation of federal funds. The SDFS office decided that school districts should be required to use evidence to justify their applications for programs. The department drafted a set of principles of effectiveness in 1997 calling on school districts to gauge the effectiveness of their program in reducing illegal drug use. D.A.R.E. America, the umbrella organization of the local programs, objected to the "principles," which seemed designed to give academic expertise more weight than practical experience. Because of its political support in high places, D.A.R.E. was able to delay the effective date of the principles. However, after considerable negotiation, the department released the principles in 1998. They were embedded in the No Child Left Behind (NCLB) legislation of 2002.

It was one thing for the SDFS program to establish principles requiring school districts to run prevention programs that were scientifically proved to reduce drug use. It was another for school districts to know which programs were eligible for funding. School districts had difficulty knowing which programs were scientifically "proven." In an innovative move, SDFS convened a committee of outside prevention experts, state and local education representatives, and educators to develop a list of proven and promising prevention programs. The panel included four prevention researchers who had developed their own programs to reduce substance abuse. Two of the people whom we interviewed in our study assured us that the program developer did not participate in the discussion about his or her own program, in fact left the room when a decision about that program was made. A member of the SDFS staff told us that she felt program developers' presence on the panel was an asset; "we saw that really as strength because they had that [program] perspective." The discussion, she said, was "definitely" unbiased.

The panel drew up a list of programs that satisfied the criterion of effectiveness. The *List of Exemplary and Promising Prevention Programs* was released in 2001. It identified 9 prevention programs as "exemplary" and 33 as "promising." The "promising" label was given to programs that did not have sufficient evaluative data to justify the higher classification. D.A.R.E. was not on the list at all. Glenn Levant, the then-president of D.A.R.E. America, criticized the composition and findings of the expert panel:

> The expert panel had people that are selling their programs to the Department of Education. There were at least five of the people on the expert panel who have programs for sale, and all of their programs were approved. I have never heard of an expert panel that was allowed to be comprised of people that are vendors. . . . [T]here is a very conspicuous conflict of interest with the selection panel.

The Department of Education did not continue to update its list of exemplary and promising programs. It subsequently referred school districts to the registries and lists maintained by the Center for Substance Abuse Prevention (CSAP), Substance Abuse and Mental Health Services Administration (SAMHSA) in the Department of Health and Human Services, and similar agencies.

D.A.R.E. never made it onto the Department of Education's list, the SAMHSA list, or any government-approved list. However, there is provision in NCLB for a waiver of the principles "to allow innovative activities or programs that demonstrate substantial likelihood of success" (NCLB Act 2002). This provision means that a school district might conduct its own scientific evaluation to assess the effectiveness of D.A.R.E., but to the best of our knowledge, no district has done so.

Imposed Use: What Did It Accomplish?

The provision that drug prevention grant funds can go only to programs that have been scientifically found to be effective has had real consequences. When we began our fieldwork in 2001, we found that six school districts had already ended their implementation of D.A.R.E. (Birkeland et al., 2005). The follow-up telephone interviews we conducted in 2003 found that an additional two districts had dropped D.A.R.E. and several of the six districts still running D.A.R.E. were running it in only a limited number of sites and were running other drug prevention programs as well. Most of the reasons given for dropping or scaling back the D.A.R.E. program had to do with government demands for scientifically proven programs, although cuts in police budgets were occasionally involved. In practice, districts were often choosing a program from the list of approved programs the SDFS office had released.

These events look like a victory for evaluation. School districts were paying attention to the results of evaluations that showed that exposure to the D.A.R.E. program did not stop students from using illegal drugs. The strategy of requiring grant recipients to attend to evaluation evidence in their choice of prevention program seems to have solved the problems that have long impeded the use of evidence in policy and practice. Is the Fairy Godmother of imposed use successful? Can we declare victory for rationality?

Possible Warts

The success of "imposed use" in gaining the admission of evidence into the decision arena depends on the assumptions that good evidence was collected and assembled and that the list of proven and promising programs was a fair and accurate representation of the evidence. We have some reservations about each of these assumptions.

Many of the evaluations of the D.A.R.E. program that gained national attention were randomized field trials, and meta-analysts chose only the best designed studies for their syntheses of the evidence. Although critics have found shortcomings in some of the studies (e.g., Gorman, 2002), no evaluation is immune from criticism. The evaluations were generally acknowledged to be of high quality.

It is the way that the evaluation evidence has been interpreted and used by the SDFS panel and other best practice lists that gives us pause. The office of SDFS would not tell us the sources that the panel had used to compile the list of effective programs. A staff member told us that the information on the sources used is "not available anywhere" and is "confidential." Instead, she referred us to the Web sites of each program for supporting data. Consequently, we selected five programs that most consistently appeared not only on the Department of Education list but also on lists of model programs prepared by other agencies, several of which listed the sources that supported their listing. The programs are Life Skills Training, Midwestern Prevention Program, Project ALERT, Project Northland, and CASASTART (for a description of the lists from which we chose those programs, see Petrosino, 2003). We reviewed all the evaluations of each of these five programs that were available at the time that SDFS prepared its enumeration of proven programs and were cited as sources on these lists. Most of the studies used randomized control groups (for full discussion, see Gandhi et al., 2007).

When looking at the totality of evidence that was used to rate programs as "model" or "promising," we found that some of the evidence looks shaky. The main reasons for concern are the fact that a program's developers did almost all the evaluations, limited evidence of positive findings, the choice of "special" sites for study, the use of subgroups rather than the full sample in report of the results, few long-term follow-up studies, and the composition and procedures of the expert panel.

Identity of the evaluator. In the case of all five programs, all or most of the evaluations were conducted by the program developer. Life Skills Training was evaluated more frequently than any of the other programs we reviewed; there were a total of 19 papers reporting results of evaluation. Eighteen of the 19 evaluation reports were written by the program's developer and his colleagues. Of the 7 evaluation reports of Midwestern Prevention Project, all were written by the developer. So were all of the evaluations of Project ALERT and Project Northland, and all except one of CASASTART.

It is perfectly reasonable for a new program to be studied by its developer. The developer wants to see how well the program is attaining its objectives so that he or she can improve it in its formative period. However, it is less reasonable to rely on only the developer's data to classify a program as exemplary. Developers are likely to study well-implemented versions of the program rather than run-of-the-mill implementations. Inasmuch as their interest is exploring the outcomes of a faithful implementation, they are likely to gravitate to well-run sites. The developer is likely to give the program the benefit of the doubt in reporting, expatiating on the good news and underplaying less favorable results. Inevitably, there is the possibility of conflict of interest and bias in reporting. Researchers have found that when a program developer is involved in the evaluation, the results are more likely to be positive (e.g., Petrosino & Soydan, 2005).

Limited evidence of positive findings. The lists generally required only a few evaluations showing positive findings to achieve the "approved" classification. With only one or two studies, data were limited. In some cases, the same data set was drawn on for different evaluation reports, and there were no independent replications.

Furthermore, most of the lists required only one statistically significant finding for a program to make it on to the list. Most evaluations used multiple outcome measures, such as use of marijuana, cocaine, heroin, tobacco, and alcohol, and use of varying frequencies, such as yesterday, last week, in the last month, in the last 6 months, and so on. For example, an evaluation of Project ALERT computed significance tests for 100 comparisons between a program and control condition (six outcome measures, six different substances, three risk levels, and two types of programs). Of these 100 tests, only 2 were statistically significant (Ellickson, Bell, & McGuigan, 1993), 1 in the positive direction and 1 in the negative direction (increase in weekly cigarette use for prior users in the adult-led program). If it took significant improvement in only one category to be "exemplary," the bar was set quite low. With large sample sizes, statistical significance was not hard to reach.

Subgroup comparisons. Some evaluation reports did not concentrate on comparing the participants in the program to the control group. Rather, they focused on different groups of participants, such as those who had never, sometimes, or fairly frequently used drugs before the program. Different versions of the program were compared (e.g., program led by teachers or by peers). Another focus was the degree of fidelity with which the program was implemented. A program developer wants to know whether the program works well when carried out as intended. But these kinds of comparisons might give a misleading impression that the generic program is effective when in fact it is effective only when run in particular ways for certain kinds of participants. Furthermore, these subgroups were not randomly assigned to program and control conditions (Gorman, 2002; Sherman, 2003), and thus, interpretation of the comparisons can be problematic.

Few long-term follow-ups. Almost none of the studies looked at outcomes more than 2 years after the end of the program. At immediate posttest, relatively few reports showed substantial

success, and even fewer studies showed substantial success at longer follow-ups. The few evaluations that followed up after the end of the program generally reported that any positive early results dissipated at longer term follow-up. Although the programs on the lists showed positive effects on some measure at some time, many of the effects seem to disappear as follow-up periods grew longer.

Composition and procedures of the expert panel. As described in the previous section, an expert panel was convened to review the evaluation evidence and draw up a list of effective prevention programs to which school districts could refer. We have noted that some of the members of the panel had developed their own drug abuse prevention programs, and even though they did not participate in decisions about the status of their own programs, they were part of the decision making on other programs.

Remember that the 500-pound gorilla in the room was D.A.R.E., by far the most popular prevention program in the country. D.A.R.E. had a great deal of political support both locally and nationally. It was well known and widely covered by the media. Its bumper stickers and t-shirts were familiar sights. It ran summer camps, kindergarten programs, and well-attended ceremonies. Parents and politicians were among its boosters. For prevention experts, it was a program created not by university experts but by a police department, aided by a curriculum development specialist in a school system. Furthermore, D.A.R.E. had not acknowledged the validity or importance of the evaluations that had been conducted on the program. Whenever evaluations came out that threw doubt on D.A.R.E.'s effectiveness, the reaction of D.A.R.E. America was likely to be hostility and recourse to political pressure.

The long-standing ill will had an effect on the lists of exemplary programs. If the rules about entry to the list had been followed for D.A.R.E., as it was for other programs, D.A.R.E. would have been included on the list. The scores of randomized field trials of D.A.R.E. certainly showed some statistically significant results. The problem was that there was much more data. Hundreds of studies of D.A.R.E. had been done, and long-term follow-ups showed that whatever effects on drug-taking behavior showed up right after the program, such effects soon dissipated. Therefore, the program could not be considered exemplary. Nevertheless, the list included other programs that, with much less data, were showing the same pattern. Not many programs seemed to have strong long-lasting effects on the drug behavior of young people.

Lack of belief in evaluation evidence. In our study, four districts dropped the D.A.R.E. program and implemented one included on the SDFS list. They did so even though they were satisfied with D.A.R.E. and believed it worked in their local context. In fact, we heard much skepticism about evaluation evidence from local respondents both because it did not reflect conditions in their own district and because they believed the studies measured the wrong things. They had never expected D.A.R.E. to keep young people off drugs completely. They were happy that D.A.R.E. had helped kids come to know police officers in a nonenforcement mode and thus improved relationships between cops and families. They were glad that police officers came to know "good" kids through D.A.R.E. and were more understanding. Moreover, the cooperation between the school system and law enforcement agencies fostered by D.A.R.E. had important positive fallout (Birkeland et al., 2005). These school districts were following what they saw as the rules (implementing a program from the list), but they saw losses as well as gains from the process.

Bureaucratic exercise more than "influence of research." The image conveyed by the phrase "influence of research and evaluation" is one of thoughtful review of evidence.

However, in this case, school districts were more likely to view the Principles of Effectiveness and the list in which they were embodied as just another bureaucratic task. To get SDFS money, they had to conform to requirements and pick a program approved by central authority. In fact, there was no requirement that they pick a program from the approved list, just pick a scientifically proven program. But to most districts in our study, the two options were the same. To satisfy requirements and select a proven program, they needed to choose from the list. It was another bureaucratic hoop to jump through.

Discussion

"Imposed use" has limitations. We have seen that there are questions about the strength of the evidence and questions about the procedures that selected some programs into the "exemplary" category. Another issue is that federally imposed use discounts the experience of local educators. It seems to override their understanding of local conditions and the local school population (Birkeland et al., 2005). School districts believe that they have to take a program from a list that is based on evaluation of program effectiveness in another place at another time. Our review of the recent literature indicated that sensitivity to local context is seen as an important factor in encouraging decision makers to take evaluation to heart. The SDFS program does not mandate a single program but provides a menu of programs from which school districts can choose. But given the fact that the list is national in scope and takes its evidence from wherever studies were done, it is not easily tailored to local conditions. There is very little evidence in the evaluations about implementation, staffing, and local circumstances in the sites in which the program was run. Therefore, the list does not give much help to districts on the fit of the evidence to their own local conditions.

Another and more serious concern that imposed use raises is whether this procedure represents the use of evidence at all. It looks rather like a bureaucratic exercise. To get federal funding, the school district has to write an application and follow directions. It has to look at the SDFS list (or now other sources of evaluative information) and choose a program from the list. The habits of mind that are being encouraged are not like the review of research evidence. They are more like toeing the line and doing what the authorities say.

We investigated the possibility that evaluation evidence convinced federal officials about the effectiveness of particular programs. That might be victory enough. Certainly federal officials left D.A.R.E. off the approved list, a decision consistent with most of what we know from evaluations of D.A.R.E. However, evaluation may not have been the primary source of their decision. We heard speculation during our study that prevention professionals in government had been convinced for a long time that D.A.R.E. was not a good bet. They had long years of experience in substance abuse prevention and had considered opinions of their own. Furthermore, D.A.R.E. had been developed and promoted by police departments, not by prevention experts. Prevention science, with its roots in academia and its desire to become a recognized academic specialty, had little in common with the quasi-military, authoritarian organizations of law enforcement, and we heard that prevention officials in government were skeptical of a successful prevention program arising from law enforcement.

In our 2003 interview, we asked the head of the SDFS program about D.A.R.E.'s absence from the approved list, and he declined to comment. He did say, "This is a very controversial issue with D.A.R.E. . . . there is a long history there." In our 2000 interview, he was more forthcoming:

Interviewer: Have you—have you been keeping up with the evaluations of D.A.R.E.?
Respondent: Yes.
Interviewer: And what do you conclude from those?
Respondent: Well, you can conclude a couple of things. One is that it hasn't been terribly effective, and, well, first of all, I think that many of us who've been around for a long period of time would probably say that they overmarketed themselves.
Interviewer: So before all these kind of negative, or at least null, evaluations of D.A.R.E. came through, were you more positive about it [D.A.R.E.]?
Respondent: No.
Interviewer: You were never a supporter of D.A.R.E.?
Respondent: No.

Another strike against D.A.R.E. was its tendency to mobilize its considerable political support whenever evidence suggested that it was not successful. It tried to substitute political pressure for empirical evidence. When officials in federal agencies had tried to rein in D.A.R.E., they made little headway for years because many political leaders were more impressed by public acclaim for D.A.R.E. than they were by evaluation. Evaluation supported the case of prevention officials in the agencies and buttressed their eventual victory, but it was perhaps not the source of their stand against D.A.R.E.

In one school district that we studied, the effect of D.A.R.E. evaluations was clear and dramatic. School staff heard of the evaluation results and dropped D.A.R.E. But if evaluation cannot claim credit for the federal decision about D.A.R.E. and if it had only occasional direct impact on other local school districts, what can we infer about imposed use? Is it a promising policy strategy?

We have collateral evidence from the review of another federal program that requires scientific evidence. Although we did not study the Reading First program, the inspector general of the Department of Education made a careful review (Office of Inspector General, U.S. Department of Education, 2006). Reading First is another program authorized under the NCLB legislation that calls for restricting federal funds to programs that are found successful by scientific research. Rather than draw up a list of approved programs, the Reading First officials convened expert panels to review all applications submitted by the states for funding. The expert panels reviewed each submission, often asked for revisions, and then reviewed the resubmissions that were requested.

Questions arose about the procedures that the Reading First staff were using. Some states complained that they were being pushed to implement particular instructional and assessment packages. As the Center on Education Policy summarized the complaints, while some observers approved of the call for research-based reading material, "others . . . have seen Reading First as too rigid, as promoting a particular philosophy for teaching reading that relies heavily on phonics and decoding, and as funneling funds to particular consultants and textbook companies" (Center on Education Policy [CEP], 2006, pp. 2-3). In its 2005 survey, CEP found that 60% of districts with Reading First grants reported that they changed their reading program to qualify for a Reading First subgrant (from the state) (CEP, 2006, p. 4).

The inspector general of the Department of Education reviewed the procedures and issued a report. The report found significant bias in the actions of the Reading First director and the former assistant secretary of the Office of Elementary and Secondary Education. The officials obviously favored phonics-based instruction, a model for teaching that requires the use of programs like Reading Mastery. Many members of the expert panels had direct connections to that style of instruction or to publishers of that kind of material. The inspector general's report states, "The Department [of Education] intervened to influence a state's selection of

reading programs. . . . The Department intervened to influence reading programs being used by LEAs [local education agencies] after the application process was completed" (Office of Inspector General, 2006, pp. 22, 24).

Somewhat similar complaints were made about Department of Education efforts to distill evidence on effective math education. Schoenfeld (2006), an advisor to the department-funded project that synthesized research evidence on mathematics curriculum effectiveness, wrote a report critical of the procedures and the results. He states that the report that he submitted was rejected and pulled from publication.

The development of a list of exemplary and promising programs, for all its limitations, appears to us to be a more open and honest procedure than the backstairs manipulation described for the Reading First program.

Conclusion

Efforts to increase the influence of research and evaluation in policy making have foundered on shortcomings in the policy system, limitations in the research system, and splintered communications between the two realms. The regulations that imposed use of research on policy makers effectively overcame their inattention to research and their unwillingness to take it seriously in making decisions. That is why we call "imposed use" the Fairy Godmother. She succeeded in putting research and evaluation on the policy table. However, she did not do much about limitations on research and researchers, and only by creation of a list of approved programs did she improve communication between evidence and policy making, at least communication of sorts. But the List did not communicate some of the essential ingredients of an evaluation perspective—such as skepticism, openness to new ideas, and new perspectives on programming.

Throughout the years, the practice of evaluation has improved in rigor and sophistication, and the evidence available is much better than it used to be. However, it is by no means perfect. Critics can point out, and have pointed out, a raft of limitations in the best evaluations of school-based drug abuse prevention programs. It is hard to claim that "science" has yielded an enduring "truth."

Furthermore, as this case demonstrates, basing policy on evidence requires somebody, or some body, to ordain and proclaim the "best evidence." An authorized group has to select that evidence which is worth attending to and give it visibility and standing. If evidence-based policy is to be "imposed," organizational arrangements have to be made to reach an authoritative answer—or set (or range) of authoritative answers. Here again our case suggests some frailty. The procedures used, even by a prestigious group of outside experts, seem to reflect a degree of bias and favoritism. Experts, it seems, may be as subject to human frailties as the rest of us.

So where do we come out on inviting the Fairy Godmother to the policy ball? We have some clues from this one case, but we do not know enough. It would be useful to know how the varied means that government agencies are using actually affect the rationality of local decisions. Given the limited evidence we have, it seems sensible (if a little clichéd) to ask for more research.

In the time since our study, agencies have tried a variety of mechanisms to encourage greater reliance on evidence. As we noted, SDFS changed its procedures and referred applicants to other agencies' lists and registries of exemplary programs. Other program agencies have given greater latitude to states to establish methods to check that applicants are basing their programs on evidence. The CSAP convened an expert panel to assist applicants for its state incentive grants for community programs. The panel released a sophisticated guide on

methods for combining evidence from multiple sources (CSAP, 2007). In the United Kingdom, the Government Social Research Unit did a study of how government agencies are trying to improve their capacity to integrate research into policy formation (Campbell, Benita, Coates, Davies, & Penn, 2007). Further investigation could follow up on the consequences of these kinds of alternative mechanisms.

We confess to partiality on the side of basing decisions on good evidence. In policy circles, a requirement for evidence of program effectiveness is likely to increase the rationality quotient in decision making. It gives greater weight to data about what has proved effective in the past for attaining program objectives. It helps to counter the constant pressure of interests and ideologies and the bent of political actors to rely on their own limited experience. The Fairy Godmother of imposed use may indeed have warts, and much care has to be exercised by everyone involved to improve the quality of the evidence available and the procedures used to select the evidence that is worthy of attention. But some version of the Fairy Godmother, in this or other generations, would add a needed perspective to fortify and brace the influence of knowledge.

On the other hand, evaluative evidence is not the only important component of policy making. The ends to which policy and programming are directed are at least as important as the means for getting there. Evidence-based policy assumes that the ends are given, and the issue is to choose the best way to get there. But there are circumstances where appropriate ends are much in contention and ethical and moral considerations must come into play (Biesta, 2007).

Then, the issue of context has to be faced. The professional judgment of the people on the scene is influenced not only by self-interest and constrained values; it is also grounded in practical wisdom and tacit knowledge. They know the local history, the people involved, the interpretations that participants provide, and all the other experiences that frame a given program. Their perspective would seem to be worth a place in decision making (Burton & Chapman, 2004; Sanderson, 2002).

Rather than accept the current version of imposed use as the best we can hope for, let us study the consequences of alternative requirements for integrating evidence into policy. Maybe the Canadian Health Services Research Foundation (2006) is right that deliberative processes can "take the summary of relevant research as the foundational starting point, inviting in values, interests, and experiences as adjuncts to, and not substitutes for, the research base" (p. 1). Or maybe that is too utopian a vision, and we have to figure out feasible strategies for giving research greater utility and visibility in the policy process. Policy making is often contentious, and research needs help to hold its own against interests, ideologies, and rival information.

Note

1. We often use the word *research* to refer to all types of systematic empirical inquiry, including evaluation. We use *evaluation* to refer specifically to research that examines the processes and outcomes of social interventions.

References

Alkin, M. C., Daillak, R., & White, P. (1979). *Using evaluations: Does evaluation make a difference?* Beverly Hills, CA: Sage.

Anderson, L. (2003). *Pursuing truth, exercising power: Social science and public policy in the 21st century*. New York: Columbia University Press.

Baklien, B. (1983). The use of social science in a Norwegian ministry, as a tool of policy or a mode of thinking? *Acta Sociologica*, *26*, 33-47.

Becker, H. (1984). The case of Germany, experiences from the education council. In T. Husen & M. Kogan (Eds.), *Educational research and policy: How do they relate?* (pp. 103-119). Oxford, UK: Pergamon.

Beyer, J. M., & Trice, H. M. (1982). The utilization process: A conceptual framework and synthesis of empirical findings. *Administrative Science Quarterly*, *27*, 591-622.

Biesta, G. (2007). Why "what works" won't work: Evidence-based practice and the democratic deficit in educational research. *Education Theory*, *57*(1), 1-22.

Birkeland, S., Murphy-Graham, E., & Weiss, C. H. (2005). Good reasons for ignoring good evaluation: Evidence from a study of the Drug Abuse Resistance Education (D.A.R.E.) program. *Evaluation and Program Planning*, *28*(3), 247-256.

Bobrow, D. B. (2006). Social and cultural factors: Constraining and enabling. In M. Moran, M. Rein, & R. E. Goodin (Eds.), *The Oxford handbook of public policy* (pp. 572-586). Oxford, UK: Oxford University Press.

Boudett, K. P., City, E. A., & Murnane, R. J. (Eds.). (2005). *Data wise: A step-by-step guide to using assessment results to improve teaching and learning*. Cambridge, MA: Harvard Education Press.

Bowen, S., & Zwi, A. B. (2005). Pathways to "evidence-informed" policy and practice: A framework for action. *Public Library of Science*, *2*(7), e166, 600-605.

Breslau, D. (1998). *In search of the unequivocal: The political economy of measurement in U.S. labor market policy*. Westport, CT: Praeger.

Bulmer, M. (1986). *Social science and social policy*. London: Allen and Unwin.

Burton, M., & Chapman, M. (2004). Problems of evidence based practice in community health and social services. *Journal of Learning Disabilities*, *8*, 56-70.

Campbell, S., Benita, S., Coates, E., Davies, P., & Penn, G. (2007). *Analysis for policy: Evidence-based policy in practice*. London: Government Social Research Unit, HM Treasury.

Canadian Health Services Research Foundation. (2006). An emerging role for deliberative processes. *Links*, *9*(4), 1-6.

Caplan, N. (1977). A minimal set of conditions necessary for the utilization of social science knowledge in policy formulation at the national level. In C. H. Weiss (Ed.), *Using research in public policy making* (pp. 183-197). Lexington, MA: Lexington-Heath.

Center for Substance Abuse Prevention, Substance Abuse and Mental Health Administration, Department of Health and Human Services. (2007, January). *Identifying and selecting evidence-based interventions: Guidance document for the strategic prevention framework, State Incentive Grant Program*. Washington, DC: Author.

Center on Education Policy. (2006). *Keeping watch on reading first*. Washington, DC: Author.

Clark, G., & Kelly, L. (2005). *New directions for knowledge transfer and knowledge brokerage in Scotland*. Scotland: Office of Chief Researcher, Knowledge Transfer Team, Scottish Executive Social Research [On-line]. Retrieved from http://www.scotland.gov.uk/Resource/Doc/69582/0018002.pdf

Clayton, R., Cattarello, A. M., & Johnstone, B. M. (1996). The effectiveness of Drug Abuse Resistance Education (Project D.A.R.E.): 5-year follow up results. *Preventative Medicine*, *25*, 307-318.

Cousins, B., & Leithwood, C. K. (1986). Current empirical research on evaluation utilization. *Review of Educational Research*, *36*(3), 331-364.

Davies, H., Nutley, S., & Smith, P. (Eds.). (2000). *What works? Evidence-based policy and practice in public services*. Bristol, UK: Policy Press.

DeJong, W. (1987). A short term evaluation of project D.A.R.E. (Drug Abuse Resistance Education): Preliminary indicators of effectiveness. *Journal of Drug Education*, *17*, 279-294.

Dopson, S., & Fitzgerald, L. (Eds.). (2005). *Knowledge to action? Evidence-based health care in context*. Oxford, UK: Oxford University Press.

Dukes, R. L., Stein, J. A., & Ullman, J. B. (1997). Long-term impact of Drug Abuse Resistance Education (D.A.R.E.): Results of a 6-year follow-up. *Evaluation Review*, *21*(4), 483-500.

Ellickson, P. L., Bell, R. M., & McGuigan, K. (1993). Preventing adolescent drug use: Long-term results of a junior high program. *American Journal of Public Health*, *83*, 856-861.

Ennett, S. T., Tobler, N. S., Ringwalt, N. S., & Flewelling, F. L. (1994). Resistance education: A meta-analysis of project D.A.R.E. outcome evaluations. *American Journal of Public Health*, *84*(9), 1394-1401.

Furubo, J. E. (1994). Learning from evaluation: The Swedish experience. In F. L. Leeuw, R. C. Rist, & R. E. Sonnichsen (Eds.), *Can governments learn? Comparative perspectives on evaluation and organisational learning* (pp. 45-65). New Brunswick, NJ: Transaction.

Gandhi, A. G., Murphy-Graham, E., Petrosino, A., Chrismer, S. S., & Weiss, C. H. (2007). The devil is in the details: Examining the evidence for "proven" school-based drug abuse prevention programs. *Evaluation Review*, *31*(1), 43-74.

Gorman, D. M. (2002). The "science" of drug and alcohol prevention: The case of the randomized trial of the Life Skills Training program. *International Journal of Drug Policy*, *13*, 21-26.

Greenberg, D., Mandell, M., & Onstott, M. (2000). The dissemination and utilization of welfare-to-work experiments in state policymaking. *Journal of Policy Analysis and Management*, *19*(3), 367-382.

Greenberg, D., & Shroder, M. (2004). *The digest of social experiments* (3rd ed.). Washington, DC: Urban Institute Press.
Gueron, J. M., & Pauly, E. (1991). *From welfare to work*. New York: Russell Sage.
Henig, J. R. (2007, May). *The evolving relationship between researchers and public policy*. Paper presented at American Enterprise Institute Conference, Washington, DC.
Kemm, J. (2006). The limitations of "evidence-based" public health. *Journal of Evaluation in Clinical Practice*, *12*(3), 319-324.
Knott, J., & Wildavsky, A. (1980). If dissemination is the answer, what is the problem? *Knowledge: Creation, Diffusion, Utilization*, *1*, 537-578.
Lakoff, G. (1996). *Moral politics*. Chicago: University of Chicago Press.
Lampinen, O. (1992). *The utilisation of social science research in public policy*. Helsinki, Finland: VAPK-Publishing.
Lemieux-Charles, L., & Champagne, F. (Eds.). (2004). *Using knowledge and evidence in health care: Multidisciplinary perspectives*. Toronto, Ontario, Canada: University of Toronto Press.
Lindblom, C. E., & Cohen, D. K. (1979). *Usable knowledge: Social science and social problem solving*. New Haven, CT: Yale University Press.
Lyons, G. M. (1969). *The uneasy partnership: Social science and the federal government in the twentieth century*. New York: Russell Sage.
Mulgar, M. (2005). Government, knowledge and the business of policy making: The potential and limits of evidence-based policy. *Evidence and Policy*, *1*(2), 215-226.
National Academy of Sciences. (1968). *The behavioral sciences and the federal government*. Washington, DC: Author.
National Science Foundation. (1968). *Knowledge into action: Improving the nation's use of the social sciences*. Washington, DC: Author.
Office of Chief Researcher, Knowledge Transfer Team (G. Clark & L. Kelly), Scottish Executive Social Research. (2005). *New directions for knowledge transfer and knowledge brokerage in Scotland*. Edinburgh, UK: Office of Chief Researcher.
Office of Inspector General, U.S. Department of Education. (2006, September). *The Reading First program's grant application process; final inspection report*. Available at http://ED-OIG/I13-F0017
Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London: Sage.
Petersilia, J. (1987). *The influence of criminal justice research*. Santa Monica, CA: Rand.
Petrosino, A. (2003). Standards for evidence and evidence for standards: The case of school-based drug prevention. *Annals of the American Academy of Political and Social Science*, *587*, 180-207.
Petrosino, A., Birkeland, S., Hacsi, T. A., Murphy-Graham, E., & Weiss, C. H. (2006). US state government and D.A.R.E.: The story in four states. *Evidence and Policy*, *2*(3), 291-319.
Petrosino, A., & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, *1*(4), 435-450.
Pogrow, S. (1996). Reforming the wannabe reformers: Why education reforms almost always end up making things worse. *Phi Delta Kappan*, *77*(10), 656-663.
Preskill, H., & Torres, R. (1999). *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage.
Pyra, K. (2003). *Knowledge translation: A review of the literature*. Nova Scotia, Canada: Nova Scotia Health Research Foundation.
Rich, A. (2004). *Think tanks, public policy, and the politics of expertise*. Cambridge, UK: Cambridge University Press.
Rigby, E. (2005). Linking research and policy on Capitol Hill: Insights from research brokers. *Evidence and Policy*, *1*(2), 195-213.
Rosenbaum, D. P., Gordon, S., & Hanson, S. (1998). Assessing the effects of school-based drug education: A six-year multilevel analysis of project D.A.R.E. *Journal of Research in Crime and Delinquency*, *35*, 381-412.
Rosenstock, L., & Lee, L. J. (2002). Attacks on science: The risks to evidence-based policy. *American Journal of Public Health*, *92*(1), 14-18.
Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, *52*, 59-82.
Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration*, *80*(1), 1-22.
Saunders, L. (2005). Research and policy: Reflections on their relationship. *Evidence and Policy*, *1*(3), 383-390.
Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful review of studies of mathematics curricula. *Educational Researcher*, *35*(2), 13-21.
Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27*. Ypsilanti, MI: High/Scope Press.

Sherman, L. W. (2003). Misleading evidence and evidence-led policy: Making social science more experimental. *Annals of the American Academy of Political and Social Science*, *589*, 6-19.

Shonkoff, J. P. (2000). Science, policy, and practice: Three cultures in search of a shared mission. *Child Development*, *71*(1), 181-187.

Shulha, L., & Cousins, B. (1997). Evaluation use: Theory, research and practice since 1986. *Evaluation Practice*, *18*(3), 195-208.

Simmons, W. (2005). Evidence-based practice: Building capacity for informed professional judgment. *Voices in Urban Education* (Annenberg Institute for School Reform), *6*, 5-13.

Sowell, T. (1980). *Knowledge and decisions*. New York: Basic Books.

St. Pierre, R. G., & Puma, M. J. (2000). Toward the dream of the experimenting society. In L. Bickman (Ed.), *Validity and social experimentation: Donald Campbell's legacy* (pp. 169-192). Thousand Oaks, CA: Sage.

Taylor, D. (2005). Governing through evidence: Participation and power in policy evaluation. *Journal of Social Policy*, *34*(4), 601-618.

Tobler, N. (1986). Meta-analysis of 143 adolescent drug prevention programs: Quantitative outcome results compared to a control or comparison group. *Journal of Drug Issues*, *16*, 537-556.

Walter, I., Nutley, S., & Davies, H. (2005). What works to promote evidence-based practice? A cross-sector review. *Evidence and Policy*, *1*(3), 335-363.

Walter, I., Nutley, S., Percy-Smith, J., McNeish, D., & Frost, S. (2004). *Improving the use of research in social care practice* (Knowledge Review 7). London: Social Care Institute for Excellence.

Weiss, C., Murphy-Graham, E., & Birkeland, S. (2005). An alternate route to policy influence: Evidence from a study of the Drug Abuse Resistance Education (D.A.R.E.) program. *American Journal of Evaluation*, *26*, 12-31.

Weiss, C. H. (Ed.). (1977). *Using research in public policy making*. Lexington, MA: Lexington Books/D.C. Heath and Co.

Weiss, C. H. (1989). Congressional committees as users of analysis. *Journal of Policy Analysis and Management*, *8*(3), 411-431.

Weiss, C. H. (1995). The haphazard connection: Social science and public policy. *International Journal of Educational Research*, *23*(2), 137-150.