

# Influential Evaluations<sup>☆</sup>

GARY T. HENRY

## ABSTRACT

Evaluations can influence perceptions about social problems, the selection of social policies, and adapting policy implementation. In this paper, exemplars of influential evaluations are described. These evaluations and the qualities which characterize them give us evidence to consider as we develop a clearer picture of what evaluation should look like in the future.

Evaluations are powerful instruments in democracies, although evaluators may have underplayed the potential for influence that evaluations have. A diminished sense of the potential for influence may have come from the focus on immediate utilization that characterized early research on evaluation use (Caplan, 1977; Knorr, 1977; Patton et al., 1977; Weiss & Bucuvalas, 1977). Moreover, recent conceptual work on process use limits the focus of changes that occur during an evaluation to those who directly participate in evaluation (Cousins, 2003; Patton, 1997; Shulha & Cousins, 1997). Recent work on developing a better understanding of and evidence about the influence of evaluation (Henry & Mark, 2003; Mark & Henry, 2004; Kirkhart, 2000) encourages evaluators and sponsors of evaluation to rethink the outcomes influenced by evaluations and the change processes that evaluations can, and frequently do, unleash.

In this article, I begin to supply evidence of influential evaluations from case examples that stand as exemplars of evaluations' impacts. The term influence, long a mainstay of political science, refers to changes that occur which would not have occurred in the absence of the putative source of the change (Dahl, 1956; Lowi, 1969). For our purposes, a change between a pre-evaluation state and a post-evaluation state, provides an indication that the evaluation was influential, if rival hypotheses for explaining the difference can be rendered implausible. The pre-evaluation and post-evaluation states could be individual attitudes, behaviors of organization members who serve as change agents, or government policies (Henry & Mark, 2003). But it is also possible that evaluations can be influential in stemming the momentum for change

---

<sup>☆</sup> A previous version of this paper was presented at the Eastern Evaluation Research Society, Absecon, NJ, April 29, 2003. Helpful comments on the original paper were provided by members of the panel, the audience, and Melvin M. Mark.

---

**Gary T. Henry** • Georgia State University, P.O. Box 4039, Atlanta, GA 30302-4039, USA; Tel: (1) 404-651-2343; Fax: (1) 404-651-3525; E-mail: gthenry@gsu.edu.

---

**American Journal of Evaluation**, Vol. 24, No. 4, 2003, pp. 515-524. All rights of reproduction in any form reserved.  
ISSN: 1098-2140 © 2003 by American Evaluation Association. Published by Elsevier Inc. All rights reserved.

---

by preventing ineffective policies or programs from being funded and implemented, as a later example will show. Beginning from the premise and growing recognition that evaluation as well as evaluation theories should be centered on social betterment (Covert, 1995; Lipsey, 2001; Mark, Henry, & Julnes, 2000; Smith, 2001), I will lay out a trichotomy of contributions that evaluations can make in the pursuit of social betterment and then supply some examples from the literature for each of them.

### THE SHINING PATH TO SOCIAL BETTERMENT

The ultimate purpose of evaluation is social betterment (Henry, 2000; Mark et al., 2000) or improving social conditions, especially for the intended beneficiaries of social programs. Undeniably, this relationship is usually complex and often a long time coming. When considered in this light, evaluations can be seen as social interventions which produce outcomes (Mark & Henry, 2004). The outcomes or consequences that are most germane to the study of evaluation influence are “the *subset of outcomes that could plausibly lead toward or away from social betterment*” (Henry & Mark, 2003, emphasis in original). Henry (2000) presented three functions through which evaluation can contribute to the pursuit of social betterment:

1. Determining the common good,
2. Selecting a course of action, and
3. Adapting the course of action.

Evaluations have influenced the path toward social betterment by providing evidence and insights in all three of these areas. Table 1 presents these three functions, a brief description of each, a list of some ways evaluation findings can contribute to each function, and some exemplar evaluations and/or types of evaluations that are likely to fulfill the function. Specific examples from the list in Table 1 will be described, albeit briefly, in the next three sections to illustrate how evaluations influenced attitudes and actions that plausibly led toward social betterment.

### RAISING A SOCIAL PROBLEM—THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Unflinching implementation of an extremely controversial change in a federal assessment program, the National Assessment of Educational Progress (NAEP), has provided the public, journalists, and policymakers with comparative information about educational achievement in the states for over a decade. Prior to 1990, SAT scores were the only source of state-to-state comparisons about educational achievement: however, interpreting these scores was confounded by differences in the samples of students taking the exam from state to state and by the lack of a clear, real world referent about what the test scores represented. Because education is largely a state-level function in the US, the absence of technically accurate and meaningful information about students' skills left the question about how well or poorly each state's public schools were serving the students largely a matter of controversy and guesswork.

Since NAEP scores have been available for individual states, reports about SAT scores of a self-selected group of graduating seniors are balanced by reports on reading, math, science, and writing achievement, generally for fourth and eighth graders, using consistent standards and assessment methods over time. For example, as shown in Table 2, in 2002, 28% of Georgia's

**TABLE 1.**

<i>Functions Performed in Pursuit of Social Betterment</i>	<i>Description</i>	<i>Relevant Evaluation Findings</i>	<i>Examples of Specific Evaluations and Types of Evaluation</i>
Determining the common good	Raising a social problem	Measuring the extent of a “problem”; Finding that current policy does not produce desired outcomes; Finding minimal outcomes for specific target populations that are intended beneficiaries	National Assessment of Educational Progress; Surveillance studies; Outcomes performance indicators; Needs analysis; Social indicators
Selecting a course of action	Choosing a policy or program	Validating a proposed solution or social intervention; Comparing the merit and worth of alternative solutions	Project High Scope; Fort Bragg Continuum of Care for Mental Health Services; Tennessee STAR experiments; Chicago Child-Parent Center; DARE evaluations
Adapting the course of action	Improving a policy, program, or organization; Modifying a program for local circumstances	Assessing policy implementation by comparing across sites; Assessing the impacts of differences in program implementation	Pre-K Longitudinal Study; Cost, Quality and Outcomes Study; Impact of Resource Utilization on Educational Outcomes

**TABLE 2.**  
**Percent Proficient on NAEP Assessments for Georgia**

<i>Georgia</i>	<i>1990</i>	<i>1992</i>	<i>1994</i>	<i>1996</i>	<i>1998</i>	<i>2000</i>	<i>2002</i>	<i>Comparing Latest Score to National Average</i>
Math (4)		15		13		18		Below
Math (8)	17	13		16		19		Below
Reading (4)		25	26		24		28	Meets
Reading (8)					25		26	Below
Science (4)						23		Below
Science (8)				21		23		Below
Writing(4)							23	Below
Writing (8)					23		25	Below

fourth graders were estimated to be proficient in reading compared to 30% nationally (NCES, 2002). Along with those from 10 other states, Georgia’s fourth graders’ reading improved significantly over their counterparts who were in the fourth grade in 1998, when 24% were judged proficient. However, the increase represents a bump up from a dip in 1998, because the fourth graders in 2002 performed no better than their peers attending the fourth grade in 1992 and 1994 (NCES, 2002).

In some quarters, measures that show progress or lack of progress in realizing the common good have been viewed as separate and distinct from evaluation, rather than as one of the important descriptive functions that evaluation can fulfill (but see Mark et al., 2000). Along with needs assessments, NAEP, state educational assessments, surveillance studies conducted by federal and state public health agencies, social indicator reports, and outcomes monitoring reports by local, state, and federal agencies can keep persistent social problems in the eyes of the public. Moreover, these reports can expose emerging social problems. Objective data that show the extent of existing social problems will not by themselves lead to social betterment, but having social problems recognized is a necessary pre-condition for consideration of social policies to address the problem (Kingdon, 1995). For example, Mintrom (1997) showed that declining test scores influenced state decisions to adopt educational reforms. In addition, Melkers and Willoughby (2001) show that state agency performance indicators are routinely included in annual budget requests and, that these requests influence the agencies’ requests for funds and the nature of the dialog and negotiations over appropriations requests by Governors. Keeping social problems on the agenda often requires presenting bad news, reporting lack of measurable progress in reducing the problems, and, sometimes, releasing negative findings about program impacts. Paradoxically, however, without bad news, there is no reason to undertake the challenge of adopting new social policies that may improve the outcomes of the program intended beneficiaries.

**SELECTING A COURSE OF ACTION—THREE INFLUENTIAL EVALUATIONS**

The evaluations of the Perry Preschool Project, the Fort Bragg Demonstration of Continuum of Care for Mental Health Services, and the Drug Abuse Resistance Education (DARE) program have been broadly recognized for their influence on public policies and programs. Each

presents an interesting case from which we can extract working hypotheses about what makes evaluations influential. I have selected these three evaluations from among numerous others to illustrate the potency of tightly designed efficacy studies that yield positive results; thorough and comprehensive quasi-experiments that yield null findings; and experiments that compare a prevalent program model to an alternative model.

### **Perry Preschool Project Evaluation**

In 1962, the evaluation of an intensive early childhood education program began in Ypsilanti, Michigan children from high poverty households were randomly assigned in matched groups to treatment ( $n = 58$ ) and control ( $n = 65$ ) over five waves (Barnett, 1992). The three- and four-year-olds in the treatment group received a half-day, developmentally oriented program 5 days per week during the regular school year and their families received home visits by the teachers once a week, which involved the children and their mothers in home-based educational activities. The treatment group fared significantly better than the control group on achievement tests and measures of school success (e.g., years spent in special education and high school graduation rates), economic success (e.g., employment), and social adjustment (e.g., arrests) (Berruta-Clement, Schweinhart, Barnett, Epstein, & Weikart, 1984; Schweinhart & Weikart, 1980, 1997). Barnett (1992) calculated the monetary benefit to taxpayers after subtracting program costs to be \$10,541 in 1988 dollars. The net benefits remained positive in a number of sensitivity analyses, indicating they were robust to alternative assumptions.

While Barnett cautions about extrapolating the net benefits to large-scale compensatory education programs, he states, “. . . compensatory education can be an economically efficient and politically palatable public investment” (1992, p. 305). The Perry Preschool evaluation, arguably, has done more to justify the investment in early childhood education, including Head Start, than any other. In 1997, the General Accounting Office (GAO) was able to find only 22 articles that assessed the impacts of Head Start, which had received an appropriation of \$4.2 billion for FY 1998 and now has increased to \$6.5 billion for FY 2003. Only four covered Head Start during the 1990s, all were relatively small scale, and none were national in scope. While many other excellent evaluations of demonstration programs exist, the Perry Preschool Program evaluation provides the best available evidence to justify public expenditures in early childhood education, shaping public perceptions and motivating policy-makers to pursue early education programs in 42 states (Meyer et al., 2002). It is interesting to note that in the response of the Head Start Bureau to the GAO report, June Gibbs Brown (p. 49, emphasis in original) notes that, the agency’s funded research program “. . . was designed to provide information necessary to *improve* the effectiveness of Head Start . . .,” but none of the evaluations or research projects that she cites were designed to provide estimates of the effects of the major types of regular Head Start programs for three- and four-year-olds. An attempt to remedy the deficiency of information about the impacts of Head Start was mandated by Congress in 1998 and got underway in 2002. Until this study is completed, many contemporary attitudes about the efficacy of early childhood education rest on the findings and cost-effectiveness calculations from the Perry Preschool Project undertaken during the 1960s.

### **Fort Bragg Demonstration of Continuum of Care Evaluation**

In 1990, an innovative system of delivering mental health services to children and adolescents was initiated in Fort Bragg, North Carolina. The demonstration featured a continuum

of coordinated mental health services that were expected to provide better access to services, more effective treatment regimes, and ultimately, improved mental health outcomes, lower costs, better client satisfaction, and quicker recovery (Bickman, 1996). Results showed that the differences in outcomes between the demonstration site and the comparison site were small and for the most part nonsignificant and the costs per child were 59% higher in the demonstration site. As Sechrest and Walsh (1997) point out, when evaluations yield null (no difference) findings, controversies often arise about the course of action that the evaluation supports and thus, the influence of the evaluation is often diluted.

In this case, the evaluators effectively ruled out implementation failure as an explanation of the null findings. Bickman (1996, p. 692) reports two coordinated studies of program quality, including an implementation study that assessed the demonstration site's service as of sufficient quality and, in addition, "The two most significant nontherapeutic activities, case management and intake-assessment central to a continuum of care, also were judged to be of high quality." By including the assessment of the quality of program implementation in the evaluation design, the evaluation increased the potential for the findings to influence policy. Other threats to the validity of the findings were addressed both by the principal evaluator (Bickman, 1996) and subsequent reviewers (Saxe & Cross, 1997; Sechrest & Walsh, 1997), leaving the latter reviewers to conclude that "the Fort Bragg study is the definitive word" (p. 540).

When an evaluation concludes that a program produced positive effects on important outcomes, it is easy to see that the findings could influence more positive attitudes about the intervention and that the findings could justify expanding the intervention, as illustrated in the Perry Preschool Project evaluation. Negative results could produce opposite reactions to avoid harming those that the program was supposed to benefit. But it seems more difficult to see how null findings could be influential. Implementation flaws or methodological problems could be the source of null findings, making them more difficult to act upon or to change pre-existing beliefs. The Fort Bragg Demonstration appeared to stall the enthusiasm for the "continuum of care" approach to managing mental health services. Instead, attention turned to managing costs (Bahar, 1997) and evaluating the effectiveness of individual mental health treatments regimes, rather than management systems (Weisz, Han, & Valeri, 1997). The Fort Bragg Demonstration Evaluation indicated that social betterment, in this case in the form of improved mental health outcomes for children and adolescents, was not to be found by moving toward a continuum of care system for mental health services, thus, forestalling costly changes in service delivery that would have left children no better off. If debunking bad ideas is the cognitive contribution that is a defining characteristic of democracies (Hurley, 1989), then null findings, when supported by rigorous methods and careful assessments of program fidelity, can contribute.

## DARE Evaluations

Perhaps nowhere has the perceived potency of evaluation been as well documented as in the case of the Drug Abuse Resistance Education program. In the popular press, Stephen Glass (1997) documented the harassment of several evaluators and suppression of evidence from the evaluations that, for the most part indicated that DARE had no effect and some even indicating negative effects. Wyson, Aniskiewicz, and Wright (1994) included an assessment of the "symbolic politics" of DARE in a scholarly article that included a report of their longitudinal evaluation of the program. They describe the "social construction" of the "drug crisis" and the threats to the direct and indirect stakeholders that arise from their findings that the program

has no effects on drug use or attitudes about drugs (1994, p. 457). As evaluations which found no effects of DARE on drug use began to pile up, support for DARE and the whole approach to reducing drug use began to dry up (Glass, 1997). A meta-analysis of evaluations of DARE and two other types of drug abuse education programs, which appeared in a highly regarded peer-reviewed journal, added a crucial piece of information when the evaluators reported, "Across four outcome domains, DARE's effect sizes are smaller than those for interactive programs" (Ennett, Tobler, Ringwalt, & Fleweling, 1994). These findings led Glass to summarize that, "DARE is not merely a failure, but crowds out money for programs that actually keep kids off drugs" (1997, p. 23). The previous findings of no effect were strengthened and, therefore, more influential, when the evaluation comparing effect sizes of competing alternatives was added to studies comparing DARE outcomes to outcomes of children who had not participated in drug abuse education programs. The DARE organization, which was an effective mobilizing, recruiting, and fundraising organization began to completely revamp the program, including changes to the curriculum and instructional methods based on those used in the more effective programs ([www.dare.com](http://www.dare.com)). A Robert Wood Johnson Foundation grant is assisting to fund the transformation and a team of university-based researchers are evaluating the impacts of the new program.

### ADAPTING THE COURSE OF ACTION

Evaluators constantly struggle with the choice between an approach that assesses the quality of program implementation and an approach that assesses program effectiveness. As the DARE example points out, choosing the latter increases the risk that tension will develop between the evaluators, on one hand, and the direct and indirect stakeholders who have a vested interest in the program, on the other hand. However, as the Fort Bragg Demonstration Evaluation showed, it is possible to have a high quality program that has a high degree of fidelity to the implementation plan that does not improve the desired outcomes. Therefore, reviewing programs without measuring outcomes may lead to better-run programs, but nonetheless move society no further along the road to social betterment. Of course, if the program's effectiveness has been previously validated, then improving the fidelity of its processes may indeed make the beneficiaries better off (Mark et al., 2000).

A study of the effects of child care on children's development provided evidence that two measures of program quality affect children's outcomes (Peisner-Feinberg & Burchinal, 1997). The measures of quality that they analyzed in the study were both based on well-known and established instruments, thus allowing replication and use in other evaluation and monitoring studies. For the first time in a large-scale study, the researchers tested and established that these measures of program quality were related to children's language development, pre-academic skills, and social behaviors in a large-scale study. While the results also indicated that the composite measure of program quality was slightly but significantly related to more problem behaviors, the effects were very positive overall. This study has influenced a number of early childhood education programs to define quality based on the characteristics of the measures used in this study and to monitor quality using similar measures. It has influenced other evaluations of other preschool program by providing evidence that has been further supported in more recent studies (NICHD Early Child Care Research Network, 2002; NICHD Early Child Care Research Network & Duncan, 2003) that these measures of quality make a positive difference across a wide number of child care sites, after controlling for a large number of



family characteristics and selection bias. Admittedly, these studies are correlational rather than experimental, however, along with efficacy studies of high quality demonstration programs, such as the Perry Preschool Project, they provide justification for raising the quality of preschool programs and have persuaded many states to pursue high quality programs.

## CONCLUSIONS

The evaluations described in this paper have influenced attitudes and actions that may lead to improved social conditions, especially for those who are recognized as the intended beneficiaries of public programs. These and similar evaluations invite us to reconsider how we think about the outcomes of evaluation and how we think about the attributes of evaluations that make them likely to enhance social betterment. To wit, one attribute of these examples is that all of them are of high technical quality as evidenced by the fact that their write-ups appear in peer-reviewed journals. The processes that influential evaluations can set off include changing attitudes or behaviors, persuading others, justifying policies and public expenditures, empowering change agents, and placing an item on the public agenda, among others (Henry & Mark, 2003; Mark & Henry, 2004). If evaluation findings are dismissed in these processes because of technical flaws or mired down by squabbles over their accuracy, it is unlikely that they will be influential. From the preceding examples of influential evaluations, we can glean evidence that high technical quality may be very important for influencing attitudes, actions, and interactions concerning social problems, the course of action that may ameliorate those problems, as well as how to adapt the programs to improve program outcomes.

Moreover, the importance of technical quality may be an issue which sets influential evaluations apart from responsive, use-focused evaluations. Use or utilization has been the frame through which the outcomes of evaluations have been viewed since the inception of modern evaluation. Responsiveness to a particular stakeholder or stakeholders has been viewed as an important attribute to increase utilization and more highly valued than technical quality for utilization-focused evaluation (Patton, 1997). Several evaluation scholars have concluded that technical quality and responsiveness are likely to come into conflict (Bryk & Raudenbush, 1983; Greene, 1990; Patton, 1997). As Greene (1990) noted, those who advocate responding to particular stakeholders “turn the ‘problem’ of threats to technical quality invoked by use-oriented responsiveness into a nonproblem by not defining technical quality” (p. 272). She goes on to add, “And the designs and methods of use-oriented evaluators, particularly with their methodological eclecticism, may undermine or threaten the technical standards of most if not all compatriot social scientists” (1990, p. 273).

Greene (1990) is undoubtedly correct in pointing out that the choices about to whom or what to be responsive and how to meet or sidestep technical standards of quality is ultimately about values and the political decision about whose interests an evaluation is to serve. But there are questions, vital to the future of evaluation raised by bringing these issues to the forefront again and engaging the field in reflection about evaluation practice and its purposes. Weiss (1998) has suggested that a question of particular moment is how we *improve* the influence of evaluation, rather than concerning ourselves with how to *increase* the influence of evaluation. The influential evaluations described in this paper and the qualities which characterize them give us data points that all can view as we develop a better understanding of how evaluation should be conducted to improve their influence.



## REFERENCES

- Bahar, L. B. (1997). The Fort Bragg evaluation: A snapshot in time. *American Psychologist*, 52, 557–559.
- Barnett, W. S. (1992). Benefits of compensatory preschool education. *Journal of Human Resources*, 27, 279–312.
- Berruta-Clement, J. R., Schweinhart, L. J., Barnett, W. S., Epstein, A. S., & Weikart, D. P. (1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19*. Ypsilanti, MI: High/Scope Press.
- Bickman, L. (1996). A Continuum of care: More is not always better. *American Psychologist*, 51, 698–701.
- Bryk, A. S., & Raudenbush, S. W. (1983). The potential contributions of program evaluation to social problem solving: A view based on the CIS and Push-Excel experiences. In A. S. Bryk (Ed.), *Stakeholder-based evaluation. New Directions for Program Evaluation* (Vol. 17, pp. 97–108). San Francisco: Jossey-Bass.
- Caplan, N. (1977). A minimal set of condition necessary for the utilization of social science knowledge in the policy formulation at the national level. In C. H. Weiss (Ed.), *Using social research in public policy making*. Lexington, MA: Lexington Books.
- Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kelligan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 245–266). Dordrecht: Kluwer.
- Covert, R. W. (1995). A twenty-year veteran's reflection on the Guiding Principles for Evaluators. In W. R. Shadish, D. L. Newman, M. A. Scheirer, & C. Wye (Eds.), *Guiding principles for evaluators. New directions for evaluation* (Vol. 66, pp. 35–46). San Francisco: Jossey-Bass.
- Dahl, R.A. (1956). *A preface to democratic theory*. Chicago: University of Chicago Press.
- Ennett, S. T., Tobler, N. S., Ringwalt, C. L., & Flewelling, R. L. (1994). How effective is Drug Abuse Resistance Education? A meta-analysis of Project DARE outcome evaluations. *American Journal of Public Health*, 84, 1394–1401.
- Glass, S. (1997). Don't You D.A.R.E. *The New Republic*, March 3, 18–28.
- Greene, J. C. (1990). Technical quality versus user responsiveness in evaluation practice. *Evaluation and Program Planning*, 13, 267–274.
- Henry, G. T. (2000). Why not use? In V. Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use. New directions for evaluation* (Vol. 88, pp. 85–98). San Francisco, Jossey-Bass.
- Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, 24, .
- Hurley, S. (1989). *Natural reasons: Personality and polity*. New York: Oxford University Press.
- Kingdon, J. (1995). *Agendas, alternatives, and public policies* (2nd ed.). New York: HarperCollins.
- Kirkhart, K. (2000). Reconceptualizing evaluation use: An integrated theory of influence. In V. Caracelli & H. Preskill (Eds.), *The expanding scope of evaluation use. New directions for evaluation* (Vol. 88, pp. 5–24). San Francisco: Jossey-Bass.
- Knorr, K. D. (1977). Policymakers' use of social science knowledge: Symbolic or instrumental? In C. H. Weiss (Ed.), *Using social research in public policy making*. Lexington, MA: Lexington Books.
- Lipsey, M. W. (2001). Re: Unsolved problems and unfinished business. *American Journal of Evaluation*, 22, 325–328.
- Lowi, T. J. (1969). *The end of liberalism: Ideology, policy and the crisis of public authority*. New York, NY: Norton.
- Mark, M. M. & Henry, G. T. (2004). *The mechanisms and outcomes of evaluation influence: Evaluation*, 10(1).
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.
- Melkers, J. E., & Willoughby, K. G. (2001). Budgeters views of state performance-budgeting systems: Distinctions across branches. *Public Administration Review*, 61, 54–64.

- Meyer, L., Orlofsky, G.F., Skinner, R.A. & Spicer, S. (2002). Quality counts. *Education Week*, January 10: <http://www.edweek.org/sreports/qc02/templates/article.cfm?slug=17sos.h21>, downloaded September 2003.
- Mintrom, M. (1997). Policy entrepreneurs and the diffusion of innovation. *American Journal of Political Science*, 41, 738–770.
- National Center for Educational Statistics. (2002). Retrieved September 2003, from NCES website <http://nces.ed.gov/nationsreportcard/reading/results2002/stateachieve-g4-compare.asp>
- NICHD Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD study of early child care. *American Educational Research Journal*, 39, 133–164.
- NICHD Early Child Care Research Network, & Duncan, G. J. (2003). *Impacts of child care quality on children's preschool cognitive development* (pp. 23–24).
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new-century text*. Thousand Oaks, CA: Sage.
- Patton, M. Q., Grimes, P. S., Guthrie, K. M., Brennan N. J., French, B. D., & Blyth, D. A. (1977). In search of impact: An analysis of the utilization of federal health evaluation research. In C. H. Weiss (Ed.), *Using social research in public policy making*. Lexington, MA: Lexington Books.
- Peisner-Feinberg, E. S., & Burchinal, M. R. (1997). Relations between preschool children's child care experiences and concurrent development: The cost quality, and outcomes study. *Merrill-Palmer Quarterly*, 43, 451–477.
- Saxe, L., & Cross, T. P. (1997). Interpreting the Fort Bragg children's mental health demonstration project: The cup is half full. *American Psychologist*, 52, 553–556.
- Schweinhart, L. J., & Weikart, D. P. (1980). *Young children grow up: The effects of the Perry Preschool Program on youths through age 15*. Ypsilanti, MI: High/Scope Press.
- Schweinhart, L. J., & Weikart, D. P. (1997). The high/scope preschool curriculum comparison study through age 23. *Early Childhood Research Quarterly*, 12, 117–143.
- Sechrest, L., & Walsh, M. (1997). Dogma or data: Bragging rights. *American Psychologist*, 52, 536–540.
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *Evaluation Practice*, 18, 195–208.
- Smith, M. F. (2001). Evaluation: Preview of the future #2. *American Journal of Evaluation*, 22, 281–300.
- Weiss, C. H. (1998). Improving the use of evaluations: Whose job is it anyway? *Advances in Educational Productivity*, 7, 263–276.
- Weiss, C. H., & Bucuvalas, M. J. (1977). The challenge of social research to decision making, In C. H. Weiss (Ed.), *Using social research in public policy making*. Lexington Books: Lexington, MA.
- Weisz, J. R., Han, S. S., & Valeri, S. M. (1997). More of what? Issues raised by the Fort Bragg Study. *American Psychologist*, 52, 541–545.
- Wysong, E., Aniskiewicz, R., & Wright, D. (1994). Truth and DARE: Tracking drug education to graduation and as symbolic politics. *Social Problems*, 41, 448–472.