

**Guidelines for Legislative Language**  
**for**  
**State Program Evaluation**

Final Report

December 18, 1998

David A. Dowell, Ph.D.

California State University Long Beach

## CONTENTS

ACKNOWLEDGMENTS .....	2
EXECUTIVE SUMMARY .....	3
Summary of the Guidelines .....	4
INTRODUCTION: WHY GUIDELINES? .....	5
Project Methods .....	5
Perspectives from Other States.....	6
Local Perspectives .....	6
Problems with Evaluation.....	6
Evaluation with Impact.....	7
The Role of the Legislature .....	7
THE GUIDELINES .....	8
1. Is evaluation of this program an important investment of state resources? .....	8
2. What questions does the Legislature need to have answered about this program?.....	8
3. What will it take to answer the Legislature’s questions – and can adequate resources be provided? .....	8
4. What will it take to ensure credible evaluation findings? .....	9
5. Who should be involved in this evaluation -- from inception to results?.....	9
6. When should evaluation findings be expected from this program? .....	9
7. What is the role of state agencies in this evaluation?.....	10
8. What information needs to be available for statewide evaluation? .....	10
CASE STUDIES .....	11
Charter School Effectiveness .....	11
Preventing Parolee Failure Program: An Evaluation .....	13
California Work Pays Demonstration Project .....	16
Discussion of the Case Studies.....	18
SAMPLE LEGISLATIVE LANGUAGE.....	21
Example 1: The “Youth Achievement Program” .....	21
Example 2: The “Community Safety Program” .....	23
FOR MORE INFORMATION .....	25
World Wide Web.....	25
Journals and Books.....	25
REFERENCES .....	28
APPENDIX: ONE-PAGE VERSION OF THE GUIDELINES .....	29

## ACKNOWLEDGMENTS

A large number of individuals supported the preparation of this report. Toby Ewing and Jay Schenirer provided important guidance. Elisabeth Kersten provided vision and leadership. Too many individuals contributed to the ideas contained in the project to mention all. Here are some of those who helped: David Illig, Lois VanBeers, David Panush, David Maxwell Jolly, Paul Warren, Mac Taylor, Charlene Simmons, Bill Padia, Phil Isenberg, Peter Hansel, Kim Connor, Patti Quate, Diane Cummins, Jeannine English, James Mayer, Joel Schwartz, Michael Jett, Leonor Ehling, Judy Chynoweth, Ginny Puddefoot, Melissa Brown, Lynn DeLapp, Cathy George, Peg Gerould, Toni Hafey, Anne Just, Allan Lammers, Vince Mandella, Anne McKinney, Frederick Morawcznski, Mary Noble, Mary Jo O'Brien, Ron Perry, Keith Prior, Carol Smith, Linda Steinmann, William White, Kevin Prior, Chuck Anders, Michael Barber, and John Berner. My appreciation extends to all.

The project was funded through the Faculty Fellows Program administered by the Center for California Studies at California State University Sacramento.

## EXECUTIVE SUMMARY

Citizens and policy makers want to see results from state-supported programs. *Program evaluation* can demonstrate results for successful programs and identify problems in programs that are not succeeding. Evaluation findings can have important impacts on state policies. Examples of significant California policy impacts include evaluation of GAIN (Greater Avenues for Independence) welfare reforms, evaluation of Proposition 99 smoking prevention programs, and evaluation of drug treatment inside Donovan state prison.

However, many – some would say most – evaluation reports on California state-funded programs are *useless to policy makers*. Why?

Discussions with policy makers, legislative staff and evaluation specialists associated with state government suggest several reasons:

- X Evaluations may fail to stay focused on key policy questions.
- X Evaluators may fail to involve key stakeholders in evaluation planning.
- X Too little time may be provided for programs to generate results.
- X Too little funding may be provided for quality evaluation.
- X Program staff may neglect evaluation planning at program start-up.
- X Findings may lack credibility because the evaluator is not independent.
- X Necessary data may not be available or may not be usable.

Evaluations that make these mistakes usually fail to have significant impacts upon state policy. Even more importantly, such evaluations undermine the credibility of the evaluation enterprise in the minds of policy makers.

Evaluations that do have significant impacts upon policy tend to have these features:

- X The focus on key policy questions remains sharp.
- X Adequate time and realistic funding are available.
- X Evaluation design is rigorous.
- X Evaluation direction is expert.
- X Key stakeholders are active participants -- including the Legislature.
- X Evaluation planning is integrated into program planning.
- X Strong oversight from a state agency is provided.

Achieving evaluation with impact requires Legislative leadership. The most important step the Legislature can take is to *focus significant evaluation attention on a limited number of high priority projects that receive adequate planning, time and resources*.

These guidelines identify issues to consider as legislation is crafted in order to ensure that evaluations will have useful impacts upon state policies. No set of guidelines alone will lead to more useful evaluation. Leadership from policy makers will be essential. These guidelines provide interested policy-makers with key questions to ask as legislation is drafted to nudge the process toward strategic investment of limited evaluation resources.

## Summary of the Guidelines

Evaluation is a complex activity and no simple recipe is possible. Acknowledging that complexity, these guidelines are written as a set of questions that should be considered when drafting legislative language calling for evaluation of state-sponsored programs.

- 1. *Is evaluation of this program an important investment of state resources?*** Setting priorities for evaluation is a key role for the Legislature. It is better to focus scarce attention and resources on a limited number of high priority projects than to try to evaluate everything.
- 2. *What questions does the Legislature need to have answered about this program?*** A clearly framed set of questions that identify what the Legislature wants to know helps greatly to focus evaluation efforts.
- 3. *What will it take to answer the Legislature's questions – and can adequate resources be provided?*** Evaluation that can significantly impact policy takes planning, expertise, time, and resources.
- 4. *What will it take to ensure credible evaluation findings?*** Evaluators must be independent from the program under review in order for findings to be credible with the Legislature in most cases.
- 5. *Who should be involved in this evaluation -- from inception to results?*** Representatives from all key stakeholder groups -- especially the Legislature -- should participate in evaluation conceptualization and planning.
- 6. *When should evaluation findings be expected from this program?*** Useful evaluation findings cannot be expected before programs have time to get organized, deliver services, and generate results. These steps take several years for state programs.
- 7. *What is the role of state agencies in this evaluation?*** State agencies can play a key role in overseeing third-party and locally conducted evaluation projects and in linking evaluators with policy makers.
- 8. *What information needs to be available for statewide evaluation?*** Standardizing key data elements from local programs with similar goals can provide a valuable resource for policy study but requires legislative mandate and state leadership.

## INTRODUCTION: WHY GUIDELINES?

Citizens want to see results from state-supported programs. *Program evaluation* can demonstrate results for successful programs and identify problems in programs that are not succeeding. Evaluation findings can have important impacts on state policies. In California, evaluation of GAIN (Greater Avenues for Independence) significantly shaped later state welfare reforms. Findings from evaluation of Proposition 99 smoking prevention programs were instrumental in preserving state support. Evidence from evaluation of a drug treatment program inside Donovan State Prison led to expansion of that approach.



However, many – some would say most – evaluation reports on California state-funded programs are *useless to policy makers*. Why?

---

A very knowledgeable, senior legislative staff member was asked the innocent question, “How do you use evaluation reports in preparing legislation?” The answer was a pantomime, dropping an imaginary evaluation report in an imaginary circular file. This turned out to be a slight exaggeration of the staff person’s actual opinion but a very meaningful gesture!

---

### Project Methods

To address concerns about the impact of evaluation, the Senate Office of Research of the California Legislature sponsored the project leading to this report. A request for proposals was issued through the Faculty Fellows program of the California State University resulting in the selection of a consultant -- the author of this report -- to prepare guidelines for legislative language for program evaluation.

Discussions were held with more than fifty individuals associated with California state policy making and evaluation in the Legislature, the executive branch, several state agencies, and the Foundation Consortium. Telephone discussions were held with individuals involved in evaluation in several other key states. Telephone discussions were held with Syracuse University researchers engaged in a national study of state government effectiveness. Evaluation reports sponsored or authored by California State agencies were reviewed. Three state evaluation reports were compared to the guidelines as case studies. A draft report was distributed to more than forty individuals. A web site version of the report was created.

## **Perspectives from Other States**

Discussions with individuals in a few other states suggest a range of approaches to measuring results of state policies. These approaches may be roughly placed in four categories although in practice, each state is a mixture of approaches.

- X Evaluation of specific programs based upon legislative priorities (example: Minnesota)
- X Comprehensive evaluation of performance of all state agencies (Florida)
- X Evaluation of state agency performance linked to statewide strategic planning (Washington)
- X Measurement of state-wide social indicators (“benchmarks”) not linked to specific agencies (Oregon)

California is, of course, larger and more complex than any other state. California’s current approach is mainly based upon evaluation of specific state programs although a few California state agencies link evaluation to strategic planning. Each of these approaches has both strengths and weaknesses.

## **Local Perspectives**

Local, county and non-profit service providers have a perspective on evaluation issues which differs in some key ways from the perspective of state government. Local providers are interested in using evaluation to improve services, tailoring information to local needs, improving the capacity of local programs to conduct and use evaluation, and minimizing the burden on services. In contrast, state government is interested in using evaluation to gain credible, independent information about program results and costs that can serve as the basis for state policy decisions. These perspectives are quite different and lead to important tensions between state government and local programs. Evaluation activities that serve one perspective may not serve the other. It is very useful to keep this tension in mind when planning an evaluation.

## **Problems with Evaluation**

Discussions with policy makers, legislative staff and evaluation specialists associated with state government suggest several reasons why evaluation may not always be useful:

- X Evaluation reports may fail to focus on key policy questions. The Legislature often wants to know *results* while some evaluations only describe clients and activities.
- X If evaluation planning does not involve all key stakeholders, findings may be misunderstood and unused.
- X The Legislature may require evaluation reports too early before programs have a realistic chance to generate results.
- X Funding earmarked for evaluation may be too little to support rigorous work.

- X Service program staff members may see evaluation as an “add-on” and neglect evaluation planning at program start-up.
- X Findings may lack credibility because the evaluator is not independent from the program under review.
- X Information necessary to answer key evaluation questions may not be available because local programs have not been required to collect standardize data.

Evaluations that fail to impact policy may be worse than merely useless. They undermine the credibility of the evaluation enterprise in the minds of policy makers. This may lead to a downward spiral in the utilization of evaluation by policy makers.

### **Evaluation with Impact**

Evaluations that do have significant impacts upon policy tend to be characterized by:

- X The focus on key policy questions remains sharp.
- X Adequate time and realistic funding are available.
- X Evaluation design is rigorous.
- X Evaluation direction is expert.
- X Key stakeholders are active participants -- including the Legislature.
- X Evaluation planning is integrated into program planning.
- X Strong oversight from a state agency is provided.

### **The Role of the Legislature**

Many priorities compete for the attention of policy makers. Evaluation’s claim for attention is based upon this premise: *Program evaluation is essential to effective government* because it can demonstrate results for successful programs and identify problems in programs that are not succeeding. Legislative leadership is essential. The Legislature can identify key policy issues and focus scarce evaluation resources on priority issues.

No set of guidelines by themselves will cause more effective evaluation to emerge. These guidelines are intended to provide interested policy-makers with the right issues to consider as legislation is drafted in order to nudge the process toward a wiser investment of limited evaluation resources. Evaluation planning is a complex activity and no simple recipe is possible. Acknowledging that complexity, these guidelines are written as a set of questions that should be considered when drafting legislative language calling for evaluation of state-sponsored programs.

## THE GUIDELINES

### 1. Is evaluation of this program an important investment of state resources?

Setting priorities for evaluation attention and resources is a key role for the Legislature. It is better to focus scarce attention and resources on a limited number of high priority evaluation projects than to try to evaluate everything.

Substantial evaluation requirements are warranted when:

- X Decision-makers are interested and evaluation findings can influence policy.
- X Programs affect the lives of citizens to an important degree.
- X Little is already known about program results.
- X Programs are costly or highly visible.
- X Adequate resources for evaluation can be made available.

Absent at least some of these circumstances, it is unlikely that evaluation will have any significant impact upon state policies. Substantial legislative evaluation requirements may *not be warranted*. It might be better to retain scarce state evaluation resources for higher priority policy questions.

### 2. What questions does the Legislature need to have answered about this program?

Clearly stated questions included in legislative language provide importance guidance for evaluation. Evaluation can answer questions like these:

- X *Results:* What were the results of the program? Were results better or worse than alternative programs with similar goals?
- X *Costs:* What did the program cost? Did the program cost more or less than alternative programs with similar goals?
- X *Services and Clients:* What services did the program provide? Whom did the program serve?
- X *Compliance:* Did the program comply with applicable laws and regulations?
- X *Management:* Was the program effectively managed?

### 3. What will it take to answer the Legislature's questions – and can adequate resources be provided?

Can the Legislature's questions be answered? Some evaluation questions cannot be answered. Others may require more time and resources than can be provided. Questions that focus on program results and costs usually have the greatest potential to impact state policies. Those require the most time-consuming and costly evaluation methods: rigorous designs, careful measurement, expert direction from an independent evaluator, and adequate time to

time to produce useful findings.

In contrast, questions about services, clients, compliance and management are usually easier to answer and can be addressed with less costly and less time-consuming methods. However, these will less often impact policy.

Consultation with qualified evaluators about contemplated evaluation requirements can help determine if the Legislature's questions can be answered and what it might cost. Individuals qualified to provide this consultation can be found in state agencies, in private consulting firms and in universities.

#### **4. What will it take to ensure credible evaluation findings?**

If an evaluation is conducted to judge program success or determine funding, it will usually be necessary to obtain an evaluator from outside the program in order to produce findings which can gain credibility.

Program staff members have an understandable motivation for program success. While this motivation is desirable, perhaps essential, for programs to succeed, it often interferes with the actual or the perceived objectivity of the staff -- especially if potential consequences of evaluation are significant.

If program staff members are asked to evaluate their own program, independent review of the evaluation design before the study is conducted and of the report after findings are obtained may bolster credibility. If a state agency conducts an evaluation on a program that it administers, similar independent review of the evaluation can bolster the credibility of the evaluation report.

#### **5. Who should be involved in this evaluation -- from inception to results?**

State-funded programs have many "stakeholders" -- people and groups who are interested in the programs. Identifying key stakeholders in legislative language can help ensure that:

- X all key points of view are included in evaluation planning,
- X evaluation stays focused upon policy concerns,
- X evaluation methods are feasible, and
- X evaluation findings are understood and used by stakeholders.

#### **6. When should evaluation findings be expected from this program?**

It takes time for programs to get organized, hire staff, deliver services and demonstrate results. Evaluation findings required too soon do not provide useful information about results. Due dates must balance the needs of policy-makers for timely information against the time

the time realistically needed for programs to get organized and produce meaningful impacts -- often two to four years. In formulating legislative requirements, consultation with individuals very knowledgeable about service delivery can help determine realistic due dates.

### **7. What is the role of state agencies in this evaluation?**

State agencies have a key, value-adding role to play in enhancing the quality and impact of evaluations. State agencies can:

- X provide guidance for evaluation in requests for proposals,
- X provide technical assistance to local programs regarding evaluation,
- X assist with access to and standardization of key data,
- X oversee evaluations conducted by third-party consultants and local programs, and
- X serve as liaisons between evaluators and policy-makers.

### **8. What information needs to be available for statewide evaluation?**

State funds commonly support multiple local programs with similar goals. If each local program collects information in a unique format, programs usually cannot be compared. However, if key elements are standardized, local data can be aggregated to assess important statewide policy issues. Limited standardization need not pose a significant burden to local programs. Limited standardization does not preclude local programs from gathering unique data in addition.

Standardization does require Legislative direction when creating programs as well as support and oversight by a state agency. Legislative direction for limited standardization of data elements can significantly “leverage” the state’s investment in services and evaluation to develop better information about the results of state policies.

## CASE STUDIES

This section provides three case studies that apply the guidelines to fairly recent evaluations of state programs. High profile case studies in three different policy areas were selected for review: charter schools, parole relapse prevention, and welfare reform. In each case, the suggested guidelines are used as the framework for the assessment. Following description of the case studies, lessons that can be drawn from these cases are discussed.

### Charter School Effectiveness

This is an analysis of the report to the Legislative Analyst's Office: *Evaluation of Charter School Effectiveness* prepared by SRI International (1997).

*Is evaluation of this program an important investment of state resources?* The California Legislature enacted the Charter Schools Act of 1992 to establish schools that would be free from most state and district regulations (Senate Bill 1448). Charter school laws are based upon the proposition that removing bureaucratic regulation will provide freedom to innovate and achieve higher educational quality. One hundred and twenty-four charter schools were authorized by the end of 1997.

Charter schools are a major, hotly-debated policy alternative in the field of education and extensive evaluation was clearly warranted.

*What questions does the Legislature need to have answered about this program?* This evaluation sought to address both descriptive questions about services and clients along with questions about results of charter schools:

- X What is the educational performance of charter schools when compared to noncharter public schools?
- X What are characteristics of charter schools in terms of educational programs, finances, and governance?
- X What practices or features are associated with particular educational outcomes?
- X What practices of sponsoring districts or other entities help or hinder the effectiveness of charter schools?

All of these are important policy questions, but some turned out to be unrealistic, given the information that could be developed.

*What will it take to answer the Legislature's questions – and can adequate resources be provided?* Descriptive questions about services and clients of charter schools could be answered by conducting interviews and surveys and gathering existing information on clients and programs. Questions about results required careful design and high quality measures.

The SRI report provides excellent answers to the descriptive questions. The report contains a wealth of information about educational practices of charter schools. High response rates to telephone interviews lend credibility to statistical summaries of program characteristics.

Charter schools turned out to be quite different from one another in their educational approaches. At the same time charter schools broke little new ground educationally; most of their educational practices can also be found in noncharter school settings. While these are descriptive rather than evaluative findings, they provide important background to the key policy issues, especially in a type of program as new to public policy as charter schools.

The SRI study is much less satisfactory in terms of answering questions about results. The key evaluation questions proved *unanswerable*. To their credit, the authors were straightforward about why these questions just could not be answered within the scope of the study. Charter schools had not been operating long enough to have clear achievement records. Data that would permit comparisons with noncharter schools did not exist. An evaluation design that would have permitted linking specific practices or features of charter schools with particular educational outcomes was far beyond the scope or resources of this evaluation project. These were serious obstacles to credible evaluation of the results of charter schools, and the authors rightly stated that the answers could not be provided within the scope of this study.

Relatively little money was spend on this evaluation project, under \$180,000. Given the policy significance of charter schools, this is much less than was warranted. However, the key questions about charter schools would likely have been unanswerable even with more money.

*What will it take to ensure credible evaluation findings?* This report is highly credible. Despite finding that key questions could not be answered, the authors resisted any temptation to paper over this shortcoming. This provides an example of why independent evaluation is so essential. Had the evaluators had an interest in the outcome of charter schools, there might have been a temptation to over-interpret inadequate information and provide misleading conclusions.

*Who should be involved in the evaluation process -- from inception to results?* It is clear from the report that the Legislative Analyst's Office was an active participant in the evaluation process, participating in some site visits. This likely helped the project stay focused on the important policy questions.

It is equally clear that the key evaluation questions about results were unrealistic, given the information and the resources available to this project. Had a group of charter school representatives been consulted early in the evaluation design phase -- even before issuing a request for proposals -- it is likely that they would have spotted the potential problem with comparable data quickly. This is a good example of why time and resources spent involving key stakeholders and planning evaluation can be very well spent. Involvement by individuals knowledgeable about charter schools in the initial conceptualization of the evaluation would have led to more realistic questions.

*When should evaluation findings be expected from this program?* Although this report was provided five years after Legislative action creating charter schools, project activities were conducted over a short seven-month period. Charter schools are highly varied and complex

programs. Charter schools are likely to vary in effectiveness. Measuring results of such a deliberately complex policy shift is very difficult. Understanding which succeed and which fail will require time-consuming and expensive evaluation methods and in particular will depend upon information available for statewide evaluation.

*What is the role of state agencies in this evaluation?* The Legislative Analyst's Office played an effective and key role in keeping this project focused on key policy questions.

*What information needs to be available for statewide evaluation?* Do we want to know which schools -- including charter schools -- are more effective than others? If we do, then a necessary step will be to gather information from schools that can be meaningfully compared. This information does not have to be based on norm-referenced achievement tests. Definitions and procedures for gathering the information just have to be the same in all schools. It is complex to design such a system. For example, students' language abilities must be carefully considered. Nevertheless, if we want to be able to eventually understand the impact on student achievement of charter schools, it will be necessary for charter schools to participate in some state-wide assessment program.

The principle of standardizing key information for statewide policy study goes well beyond the education arena. In almost every policy area, there are key pieces of data that, if collected in a standardized manner, would provide the basis for analysis of key policy questions. Only the Legislature can establish requirements for standardized information gathering.

### **Preventing Parolee Failure Program: An Evaluation**

This is an analysis of the report to the Legislature: *Preventing Parolee Failure Program: An Evaluation* provided by the California Department of Corrections (April 1997).

*Is evaluation of this program an important investment of state resources?* The overall goals of the Preventing Parolee Failure Program were to reduce parolee failures and subsequent returns to prison without increasing risk to society. The program was composed of five separate components operating independent of each other but with the same goal. One component trained parole unit supervisors in consistent decision making, with an emphasis upon encouraging the use of community resources. Three components provided substance abuse treatment and rehabilitation. A final component provided parolees with computer-assisted learning.

Thorough evaluation was clearly warranted for this program. With the cost of prisons skyrocketing, it was critical to evaluate carefully a program which promised to contain those costs.

*What questions does the Legislature need to have answered about this program?* This report examined recidivism rates for the Parolee Partnership, Prison Parole Network and residential multi-service centers. It also examined consistency in parole decision making and learning gains in the computer-assisted centers. The report attempted to provide cost savings estimates

estimates based upon the measured outcomes.

Program results and costs were key policy issues likely to interest the Legislature. It is less clear that consistency in decision making or that parolee learning outcomes were of keen interest to the Legislature.

*What will it take to answer the Legislature's questions – and can adequate resources be provided?* To answer these questions, a rigorous evaluation design was necessary. The study attempted to compare parolees who received services to similar parolees who did not receive services, a design often called a “matched comparison.” Matched comparisons are notoriously difficult to implement carefully enough to provide convincing findings. This report provided no exception to these difficulties.

Recidivism was examined for parolees in three drug treatment components: the Parolee Partnership Program, the Prison Parole Network, and the residential multi-service center. In all three studies, parolees had to “stick with” services for a period of time to be counted in the treatment group. In the Parolee Partnership Program, 700 offenders were serviced yet only 357 were included in the evaluation. The report did not describe what happened to the other parolees. Some -- we are not told how many -- were eliminated by the participation criterion. In the Prison Parole Network, 600 offenders were served but again only 357 (a coincidence of numbers?) were included in the evaluation. The report did not describe what happened to the other 243 parolees except to suggest that some were eliminated by the participation criterion. Reporting on the multi-service center study is similar. None of the studies provide critical information on how similar the treatment and comparison groups were on background and risk factors.

The practice of including only subjects who pass screening criteria is often called “creaming” to suggest that programs take the most qualified, most motivated participants. Of course, with the most qualified, most motivated participants, results are better. Program staff members sometimes assert that it is necessary to “fit clients to services” and to “deliver an adequate dose of services.” These arguments may apply to service delivery but they do not apply to evaluation design.

In the comparison groups for these studies, no subjects were eliminated by a participation criterion. Thus, treatment and control groups in these studies were different, to an unknown degree, in unknown ways. Thus, they were not comparable. This flaw unfortunately undermines the key conclusions of the report and findings cannot credibly be interpreted as supporting the success of these program components.

Is the report just simplifying a more complex set of findings? In another section, *Cost Effectiveness of the Total Program*, authors did not hesitate to present fairly complex time series graphs. In this section, the initial presentation of “return-to-prison” rates did not support a conclusion of program effectiveness, so the authors provide a complex analysis in an attempt to show that the program is really effective after all. Both the simple and the complex analyses show that “return-to-prison” rates actually went up after implementing the program. Nevertheless, the report argues, rates went up even more for parole units not

participating in the program and so the program is effective after all. This argument deserves serious consideration but evidence provided is not compelling.

A more convincing design would have included data from all parolees eligible for treatment services in the treatment comparison group just as data from all parolees in the unserved areas were included in the comparison group. In addition, the comparability of the treatment and control groups on demographic characteristics and risk factors should be examined carefully. These analyses appear to have been possible with the data in hand. Additionally, a specialized evaluation design known as “multiple baseline” might be ideal for analyzing the return-to-prison. These analyses are so important that someone should carry them out on the data which appear to have already been collected.

The cost-benefit analyses provided in the report are the right kinds of analyses. However, since they depend upon the flawed outcome findings, the cost-benefit estimates cannot be viewed as convincing.

The report does not indicate the budget for the evaluation but the identified problems appear unrelated to funding. The right kinds of data were collected and the sizes of the samples were adequate to conduct evaluation. In addition, the report shows evidence that the evaluators were skilled and knowledgeable.

This report fell short of demonstrating the effectiveness of parole relapse prevention programs. Note that the report cannot be interpreted as having demonstrated failure of the program. Based on this report, we just don't know. Policy-makers have not gained from this evaluation project information credible enough to serve as a basis for state policy.

*What will it take to ensure credible evaluation findings?* This report appears to be a work of advocacy rather than objectivity designed to convince the Legislature that the program should be refunded. Undoubtedly, the report authors and the program staff strongly believed in the program which may have made it difficult to conduct an objective evaluation.

This is a good example of why independent evaluation is so essential to effective government. A qualified independent evaluator would likely have provided a more compelling report, although that report might or might not have been as favorable to the program.

*Who should be involved in the evaluation process -- from inception to results?* The report says little about participation in the evaluation process. Broader input into the evaluation design might have addressed some of the concerns raised above. The report appears to be on-target in terms of addressing key legislative issues but no information is provided regarding how that laudable focus on key policy questions was attained.

*When should evaluation findings be expected from this program?* The program operated from 1991 forward and the report was provided in April, 1997. Six or seven years elapsed from the authorization of the program to the completion of this evaluative study. That appears to be about the necessary amount of time for a program of this type to generate

findings about results.

*What is the role of state agencies in this evaluation?* In this study, a state agency sought to fill the role of evaluator. The state agency might have played a stronger role as overseer of a third-party evaluation project. In other policy areas, such as employment (with the GAIN evaluation), state agencies have successfully played a key role as overseer ensuring that high quality evaluation is conducted by a third-party evaluator.

*What information needs to be available for statewide evaluation?* To carry out this evaluation, “recidivism” had to be defined and measured in the same way across different probation units. With any social indicator such as recidivism, the problems of ensuring quality data must be considered. The report does not describe how those problems were addressed in this evaluation. It may be that the Department of Corrections has sufficiently ensured standardization of this indicator that it does not pose a problem for this study. In many other policy areas, standardization cannot be presumed.

### **California Work Pays Demonstration Project**

This is an analysis of the report: *California Work Pays Demonstration Project* (School of Public Policy & Social Research, University of California, Los Angeles, December, 1996).

*Should you evaluate?* The Work Pays program was a California welfare policy reform. It was initiated under federal waivers that provided for exceptions to the Aid to Families with Dependent Children (AFDC) program regulations. Work Pays changed welfare policy in several ways. It reduced AFDC cash grants. It also waived time limits on two rules about earned income that determined the amount of earned income recipients could retain and the number of hours per month they could work.

Thorough evaluation was clearly warranted for this policy change. The Work Pays Demonstration Project was a major policy innovation with potentially very large impacts upon state costs and the welfare of citizens. Given that California spent 6.5 billion dollars on welfare in 1994-95, intensive evaluation was clearly warranted. Further, the federal government later implemented similar reforms of welfare making evaluation of the California experience of even greater significance. It is highly appropriate that this program received extensive evaluation attention.

*What questions does the Legislature need to have answered about this program?* This evaluation included key questions about results:

- X Do AFDC cases increase their work activity?
- X Do AFDC cases show increased earnings?
- X Do AFDC cases spend less time on aid?
- X Do AFDC recipients have more total income at their disposal?
- X Are there differences based upon county and/or aid category?

Evaluation also gathered descriptive from a language survey and a process evaluation. These provided important information. The wisdom of including the process evaluation component

in particular became apparent when findings showed that implementation of the program was uneven across sites. These were the right questions to evaluate.

*What will it take to answer the Legislature's questions – and can adequate resources be provided?* It took a great deal to answer these questions. This evaluation was very sophisticated and large in scale with thousands of subjects, randomized design, and extensive data collection. Care was taken to examine the possible impacts of attrition -- loss of research subjects. This was an extremely high quality evaluation design.

At the same time, two key weaknesses are apparent in the evaluation design. As the authors note, the “point in time” sampling methodology probably over-represented long-term welfare cases in the sample. This was problematic because the intervention arguably might be most effective with “new to aid” cases.

Second, the intervention was implemented unevenly in the various study sites. This creates some ambiguity about whether the evaluation tested a full implementation of the program. This circumstance was probably not under the control of the evaluators. It shows the wisdom of the evaluators in carrying out a process study along with the impact studies to determine the extent to which the policy was actually implemented.

Unfortunately, both of these design problems, if they have an effect, would likely lead to underestimation of program effects. This leads to the admonition appropriately provided by the authors that “Policy makers and administrators should be cautious in drawing any conclusions from these preliminary results or about the effectiveness of Work Pays.” While these issues were only partially under the potential control of the evaluators, it is unfortunate that such an extensive and well-funded evaluation was plagued by these difficulties.

*What will it take to ensure credible evaluation findings?* The credibility of this report is extremely high. The report demonstrates sophisticated evaluation planning and execution. Evaluators provided a highly objective evaluation. Had the evaluators been associated with the program, interpretation of the findings could have been shaded less objectively. This is a good example of the value of independence in evaluation.

*Who should be involved in the evaluation process -- from inception to results?* The report does not describe participation in evaluation planning. It is possible that involvement of program service staff in planning the evaluation design could have spotted the sampling problem which occurred and led to a stronger design.

*When should evaluation findings be expected from this program?* California's welfare reform dates from fall of 1992 and this report was issued in December 1996 covering the program period January 1993 to June 1995. Thus, about four years elapsed between the policy change and the first interim evaluation report. Four years is a very rapid time line for assessment of a policy change of this magnitude. It is likely to take another two years or so of evaluation to address the key design problems noted above.

*What is the role of state agencies in this evaluation?* The Department of Social Services,

with its extensive expertise in evaluation, played a valuable role in oversight of this project, helping to ensure the quality of the final product. This illustrates how a state agency can play that valuable oversight role in evaluations conducted by third parties.

*What information needs to be available for state-wide evaluation?* This study used data from counties, the California Employment Development Department, and the Medi-Cal system. The report covers important technical issues of linking these disparate data sources. Legislative encouragement to ensure that such state-wide data systems contain the key elements which enable them to be linked would facilitate the type of state-wide policy study represented by the Work Pays evaluation. This is true in a range of policy areas beyond welfare reform.

### **Discussion of the Case Studies**

*Should you evaluate?* Because high profile case studies were selected for review, all were likely ones that the Legislature would regard as warranting careful evaluation. In fact, two of the three policy changes may have warranted significantly more investment than was provided -- especially in evaluation planning.

*What does the Legislature need to know?* These studies mostly addressed questions that were likely of strong interest to the Legislature, results and costs. The parole study included some issues which may have been of less interest to the Legislature. The charter schools study and especially the Work Pays study are marred by complex reporting which prompted by the difficulty these two projects had in answering key evaluation questions.

*What will it take to answer the Legislature's questions – and can adequate resources be provided?* All three of these case studies suffered from weaknesses which impair the credibility of their potentially most important conclusions. Unrealistic evaluation questions might have been avoided by more inclusive planning. Design weaknesses might also have been avoided with more thorough planning. The lack of objectivity apparent in one study calls for independence in the evaluator. In one case (charter schools), the key evaluation questions proved *unanswerable* because data were unavailable. These cases demonstrate how difficult it is to generate convincing findings in the messy real world of policies and programs.

*How can credible findings be ensured?* Two of these evaluations were independently conducted while staff associated with the intervention program (the parole study) authored one. The independent evaluations are more objectively written documents. For example, compared to the parole evaluation, the Work Pays study had a stronger design, more subjects, and multiple measures. Yet, the Work Pays study clearly identified its design weaknesses and cautions policy makers against over-interpreting findings. The internally prepared parole study glosses over its weaknesses and makes strong claims for program success. Evaluator independence is clearly important.

*Who should be involved in the evaluation process -- from inception to results?* These reports don't provide much information about stakeholder participation in the evaluation process.

Had legislation required broader participation in the evaluation conceptualization and planning phases, some key problems might have been avoided. The Work Pays sample problem might have been anticipated. The difficulty of answering key questions about charter schools would have been discovered quickly. The “creaming” bias in the parole study might have been questioned.

*When should evaluation findings be available?* These evaluation reports were provided four to seven years after Legislative action authorizing these programs. For policy changes of the magnitude represented by two of these (welfare reform and charter schools), this time frame is perhaps minimally adequate. Most state programs will require similar amounts of time.

*What information needs to be available for statewide evaluation?* Charter schools illustrate the need for statewide standardization of key data elements. This problem is most obvious in public education because most people are familiar with the standardized tests common in that area. However, the principle is similar in other policy areas.

Standardizing key measures of results for all programs with similar goals can make it possible to compare programs. This need not impose a heavy burden on local programs and could even reduce their evaluation burden if it means they do not have to locally develop some data collection procedures.

It is a reality that the expertise and resources needed to conduct high quality evaluations are often simply unavailable to local programs. It might often be desirable to limit requirements imposed upon local programs to collection of key standardized data elements under the oversight of a state agency. These elements could then be used in statewide analyses that are adequately planned and supported. Standardizing measures will not be a simple undertaking in any policy area. It will require a legislative mandate and statewide leadership. The challenge is illustrated by the controversy over the Governor’s mandate for statewide public school testing. Nevertheless, selective use of this strategy might be a cost-effective means of acquiring better information about policy impacts.

*How can state agencies help enhance the quality and impact of evaluation?* A key role for state agencies is suggested by the foregoing discussion of standardization of key measures of results. In two of these case studies, state agencies played a useful role in keeping third-party evaluators focused on key policy questions. State agencies can play this role effectively in many instances. Evaluation findings are more credible when a state agency plays an oversight role than when the agency attempts to evaluate a program which it is also responsible for administering. In providing oversight, state agencies can give useful guidance to local programs on evaluation planning and serve as liaisons between evaluators and policy makers.

Three case studies is too small a sample from which to draw any generalizations. Each of these evaluations had significant problems that limited their impact upon state policy. Evaluations that fail to provide credible answers to key policy questions not only fail to help policy makers, they undermine the credibility of the entire evaluation enterprise. These case studies illustrate that useful evaluation requires careful evaluation planning. This

planning takes time, broad stakeholder participation and strong evaluation expertise. The most important step the Legislature can take is probably to:

*Focus evaluation attention on a limited number of high priority projects that receive adequate planning, time and resources.*

## SAMPLE LEGISLATIVE LANGUAGE

This section contains examples of hypothetical legislative language which illustrate the guidelines. Guidelines addressed by each element of the samples are noted in parentheses.

### **Example 1: The “Youth Achievement Program”**

*The “Youth Achievement Program” is an educational innovation to be piloted in multiple sites on a competitive grant basis. Eight local sites will receive an average of \$200,000 each for a total program budget of \$1.6 million. Language below exemplifies the evaluation component of the legislation. The goal of the “Youth Achievement Program” is to improve the academic achievement of middle-school youth.*

Because this program is new and innovative and poses the potential for significant benefits to youth, moderately extensive evaluation is warranted (Guidelines 1). The California Department of Education shall develop a plan for evaluating the program (Guideline 7). The sum of \$200,000 is appropriated (Guideline 3) for purposes of contracting with an independent qualified evaluator (Guideline 4) to conduct the evaluation.

The following questions should be addressed by program evaluation (Guideline 2):

- X How many children were served by local projects funded under this program and what were the services provided?
- X What proportion of children in the geographic target areas of funded projects received services?
- X Were the services provided by local projects aligned with the legislative goals of the program?
- X Did the academic achievement of children who were served by the Youth Achievement Program improve over the period of their participation?
- X Were improvements in achievement of children related to participation in the Youth Achievement Program?
- X What was the cost of delivering services for the program as a whole, for each local project, and for each child served?
- X Did some local projects demonstrate more success in improving the academic achievement of youth served than others?
- X What lessons were learned that would be useful in planning future programs?

Key audiences (Guideline 5) for evaluation findings include local communities served by funded projects, the Executive Branch Office of Child Development, the California Department of Education and the California Department of Social Services as well as the state Legislature. The evaluation plan shall incorporate a plan for involving key audiences (Guideline 5) including, but not limited to the following provisions:

- X an advisory committee representing key audiences and including legislative representation,
- X means of involving key audiences in planning evaluation activities,

- X means to apprise key audiences of emerging findings, and
- X means to communicate final results clearly to key audiences.

A final evaluation report shall be due no later than three years from point in time at which funding is made available to local programs (Guideline 6). An interim report shall be provided to key audiences no later than 14 months following the point in time at which funding is made available to local programs (Guideline 6).

Local projects shall cooperate with statewide evaluation activities from the beginning of program planning (Guideline 7,8). Local projects shall incorporate procedures for collecting key common program data elements standardized across all programs (Guideline 8). The program evaluator in consultation with local projects and the Department of Education shall develop these data elements (Guideline 8).

The Department of Education shall (a) provide technical assistance to local projects in designing and implementing evaluation activities (b) oversee the statewide evaluation, and (c) facilitate collaboration between the evaluator and local projects (Guideline 7).

## Example 2: The “Community Safety Program”

*The “Community Safety Program” is a safety program to be administered by local cities or counties in several California sites and funded on a request-for-proposal basis. Twenty local sites will receive an average of \$500,000 each for a total program budget of \$10 million. Language below exemplifies the evaluation component of the legislation. The goal of the “Community Safety Program” is to reduce serious crimes against property and persons.*

Because reducing crime is of utmost importance, extensive evaluation is warranted (Guidelines 1). The Office of Criminal Justice Planning shall develop a plan for evaluating the program (Guideline 7). The sum of \$600,000 is appropriated (Guideline 3) for purposes of contracting with an independent qualified evaluator to conduct the evaluation (Guideline 4).

The following questions should be addressed by program evaluation (Guideline 2):

- X How many citizens were served by local projects funded under this program and what services were provided?
- X What proportion of citizens in the geographic target areas of funded projects received services?
- X Were the services provided by local projects aligned with the legislative goals of the program?
- X Was there a reduction in the incidence of serious crimes against persons in areas served by local projects?
- X Was there a reduction in the incidence of serious crimes against property in areas served by local projects?
- X Can any observed reductions in crime be confidently attributed to the Community Safety Program?
- X What was the cost of delivering services for the program as a whole and for each local project?
- X Did some local projects demonstrate more success in reducing crime than others?
- X What was the estimated value of crimes prevented compared to the cost of the program?
- X What lessons were learned that would be useful in planning future programs?

Key audiences for evaluation findings include local communities served by funded projects, the Office of Criminal Justice Planning, and law enforcement agencies and professional organizations as well as the state Legislature (Guideline 5). The evaluation plan shall incorporate a plan for communication with key audiences including, but not limited to the following provisions (Guideline 5):

- X an advisory committee representing key audiences and including legislative representation,
- X means to involve key audiences in planning evaluation projects,
- X means to apprise key audiences of emerging findings,

X means to communicate final results clearly to key audiences.

A final evaluation report shall be due no later than four years from point in time at which funding is made available to local programs (Guideline 6). Interim reports shall be provided to key audiences no later than 14 and 36 months following the point in time at which funding is made available to local programs (Guideline 6).

Local projects shall cooperate with state-wide evaluation activities from the beginning of program planning (Guideline 7). Local projects shall incorporate procedures for collecting key common program data elements standardized across all programs (Guideline 8). The program evaluator in consultation with local projects and the Office of Criminal Justice Planning shall develop these data elements (Guideline 8).

The Office of Criminal Justice Planning shall (a) provide technical assistance to local projects in designing and implementing evaluation activities (b) oversee the statewide evaluation, and (c) facilitate collaboration between the evaluator and local projects (Guideline 7).

## FOR MORE INFORMATION

### World Wide Web

On the World Wide Web, these guidelines are located at

<http://www.csulb.edu/~ddowell/goldbook.htm>

The web site includes links to a wealth of on-line evaluation resources.

The most comprehensive on-line link to evaluation resources is Electronic Resources for Evaluators:

<http://www.itrs.usu.edu/AEA/resources.html>

### Journals and Books

The journals below are devoted to evaluation.

*Assessment & Evaluation in Higher Education*  
*Educational Evaluation and Policy Analysis*  
*Education Policy Analysis Archives*  
*Education Review: A Journal of Book Reviews*  
*Evaluation and Program Planning*  
*Evaluation and the Health Professions*  
*Evaluation Exchange*  
*Evaluation: International Journal of Theory, Research & Practice*  
*Evaluation Review*  
*Evaluation Review: Journal of Applied Social Research*  
*Journal of Evaluation in Clinical Practice*  
*Studies in Educational Evaluation*  
*Educational Evaluation and Policy Analysis*

There is an extensive literature on program evaluation. Below are some books in print which provide guidance to program evaluation. This does not include books in specialized policy areas such as health care, corrections, and so on.

Abma, Tineke A.; ed. (1999) *Advances in Program Evaluation; Telling Tales: On Narrative & Evaluation*, Jai Press.

Berk, Richard A. (1998) *Thinking about Program Evaluation*. Sage Publications.

Bickman, Leonard; ed. (1990) *Advances in Program Theory*. Jossey-Bass, Incorporated Publishers.

- Brainard, Edward A. (1996) *A Hands-On Guide to School Program Evaluation*. Phi Delta Kappa Educational Foundation.
- Calder, Judith (1993) *Program Evaluation & Quality; A Comprehensive Guide to Setting up an Evaluation System*. Nichols Publishing Company.
- Chen, Huey-Tsyh; ed. (1992) *Using Theory to Improve Program & Policy Evaluations*. Greenwood Publishing Group, Incorporated.
- Gredler, M. E. (1995) *Program Evaluation*. Prentice Hall.
- Gunn, Elizabeth M. (1991) *Program Planning & Evaluation for the Public Manager*. Waveland Press, Incorporated.
- Jaeger, Richard M.; ed. (1994) *Essential Tools for Educators; The Program Evaluation Guides for Schools*. Corwin Press, Incorporated.
- Joint Committee on Standards for Educational Programs Staff (1994) *The Program Evaluation Standards; How to Assess Evaluations of Educational Programs*. Sage Publications, Incorporated.
- Krause, Daniel (1995) *Effective Program Evaluation. An Introduction*. Nelson-Hall, Incorporated.
- Mika, Kristine L. (1996) *Program Outcome Evaluation; A Step-by-Step Handbook*. Families International, Incorporated.
- Mohr, Lawrence B. (1995) *Impact Analysis for Program Evaluation*. Sage Publications, Incorporated.
- Newman, Dianna L. (1995) *Applied Ethics for Program Evaluation*. Sage Publications, Incorporated.
- Owen, John M. (1994) *Program Evaluation; Forms & Approaches*. Paul & Company Publishers Consortium, Incorporated.
- Owen, John M. (1998) *Program Evaluation*, Paul & Company Publishers Consortium.
- Payne, David A. (1994) *Designing Educational Project & Program Evaluations; A Practical Overview Based on Research & Experience*. Kluwer Academic Publishers.
- Posavac, Emil (1996) *Program Evaluation; Methods & Case Studies*. Prentice Hall.
- Royse, David (1996) *Program Evaluation; An Introduction*. Nelson-Hall, Incorporated.
- Scriven, Michael; ed. (1993) *Hard-Won Lessons in Program Evaluation*. Jossey-Bass,

Incorporated Publishers.

Sechrest, Lee; ed. (1993) *Program Evaluation; A Pluralistic Enterprise*. Jossey-Bass, Incorporated Publishers.

Serow, Robert C.(1998) *Program Evaluation Handbook*. Ginn Press.

Shadish, William R. (1993) *Foundations of Program Evaluation; Theories of Practice*. Sage Publications, Incorporated.

Smith, Michael (1990) *Program Evaluation in the Human Services*. Springer Publishing Company, Incorporated.

Stake, Robert E.; ed. (1996) *Advances in Program Evaluation* (monograph series). Jai Press, Incorporated.

Stark, Joan S.; ed. (1994) *Assessment & Program Evaluation*. Ginn Press.

Sylvia, Ronald D. (1996) *Program Planning & Evaluation for the Public Manager*. Waveland Press, Incorporated.

Timmreck, Thomas S. (1995) *Planning Program Development & Evaluation*. Jones & Bartlett Publishers, Incorporated.

Vedung, Evert (1997) *Public Policy & Program Evaluation*. Transaction Publishers.

Worthen, Blaine (1997) *Program Evaluation*. Addison-Wesley Longman, Inc.

## REFERENCES

SRI International (1997) *Evaluation of Charter School Effectiveness*. Report to the California State Legislative Analyst's Office. Author: Menlo Park, California.

California Department of Corrections (April, 1997) *Preventing Parolee Failure Program: An Evaluation*. Report to the California State Legislature. Author: Sacramento, California.

School of Public Policy & Social Research, University of California, Los Angeles (December, 1996). *California Work Pays Demonstration Project: Interim Report of the First Thirty Months*. Author: Los Angeles, California.

## **APPENDIX: ONE-PAGE VERSION OF THE GUIDELINES**

The following one page version of the guidelines is suitable for reproduction on card stock as a handy reference.

# Guidelines for Legislative Language for State Program Evaluation

## *Quick Summary*

Evaluation is a complex activity and no simple recipe is possible. Acknowledging that complexity, these guidelines are written as a set of questions that should be considered when drafting legislative language calling for evaluation of state-sponsored programs.

- 1. *Is evaluation of this program an important investment of state resources?*** Setting priorities for evaluation is a key role for the Legislature. It is better to focus scarce attention and resources on a limited number of high priority projects than to try to evaluate everything.
- 2. *What questions does the Legislature need to have answered about this program?*** A clearly framed set of questions that identify what the Legislature wants to know helps greatly to focus evaluation efforts.
- 3. *What will it take to answer the Legislature's questions – and can adequate resources be provided?*** Evaluation that can significantly impact policy takes planning, expertise, time, and resources.
- 4. *What will it take to ensure credible evaluation findings?*** Evaluators must be independent from the program under review in order for findings to be credible with the Legislature in most cases.
- 5. *Who should be involved in this evaluation -- from inception to results?*** Representatives from all key stakeholder groups -- especially the Legislature -- should participate in evaluation conceptualization and planning.
- 6. *When should evaluation findings be expected from this program?*** Useful evaluation findings cannot be expected before programs have time to get organized, deliver services, and generate results. These steps take several years for state programs.
- 7. *What is the role of state agencies in this evaluation?*** State agencies can play a key role in overseeing third-party and locally conducted evaluation projects and in linking evaluators with policy makers.
- 8. *What information needs to be available for statewide evaluation?*** Standardizing key data elements from local programs with similar goals can provide a valuable resource for policy study but requires legislative mandate and state leadership.