

Ninety-Five Theses

1. Program evaluation is a process by which society learns about itself.
- X 2. Program evaluations should contribute to enlightened discussion of alternative plans for social action.
3. Evaluation is a handmaiden to gradualism; it is both conservative and committed to change.
- X 4. An evaluation of a particular program is only an episode in the continuing evolution of thought about a problem area.
5. The better and the more widely the workings of social programs are understood, the more rapidly policy will evolve

and the more the programs will contribute to a better quality of life.

6. Observations of social programs require a closer analysis than a lay interpreter can make, for unassisted judgment leads all too easily to false interpretations.
7. In debates over controversial programs, liars figure and figures often lie; the evaluator has a responsibility to protect his clients from both types of deception.

* * *

8. Ideally, every evaluation will inform the social system and improve its operations, but everyone agrees that evaluation is not rendering the service it should.
9. Commissioners of evaluations complain that the messages from evaluations are not useful, while evaluators complain that the messages are not used.

* * *

10. The evaluator has political influence even when he does not aspire to it.
- X 11. A theory of evaluation must be as much a theory of political interaction as it is a theory of how to determine facts.
- X 12. The hope that an evaluation will provide unequivocal answers, convincing enough to extinguish controversy about the merits of a social program, is certain to be disappointed.
- X 13. The evaluators' professional conclusions cannot substitute for the political process.
14. The distinction between evaluation and policy research is disappearing.

* * *

- X 15. Accountability emphasizes looking back in order to assign praise or blame; evaluation is better used to understand events and processes for the sake of guiding future activities.
16. Social renovations disappoint even their architects.
17. Time and again, political passion has been a driving spirit behind a call for rational analysis.
- X 18. A demand for accountability is a sign of pathology in the political system.

* * *

- X 19. An open society becomes a closed society when only the officials know what is going on. Insofar as information is a source of power, evaluations carried out to inform a policy maker have a disenfranchising effect.
- X 20. The ideal of efficiency in government is in tension with the ideal of democratic participation; rationalism is dangerously close to totalitarianism.

- X 21. The notion of the evaluator as a superman who will make all social choices easy and all programs efficient, turning public management into a technology, is a pipe dream.
- X 22. A context of command, with a manager in firm control, has been assumed in nearly all previous theories of evaluation.
- 23. An image of pluralistic accommodation more truly represents how policy and programs are shaped than does the Platonic image of concentrated power and responsibility.
- 24. The evaluator must learn to serve in contexts of accommodation and not dream idly of serving a Platonic guardian.
- X 25. In a context of accommodation, the evaluator cannot expect a "go/no-go" decision to turn on his assessment of outcomes.

* * *

- X 26. What is needed is information that supports negotiation rather than information calculated to point out the "correct" decision.

-
- 27. Events move forward by piecemeal adaptations.
 - X 28. It can scarcely be said that decisions about typical programs are "made"; rather, they emerge.
 - X 29. The policy-shaping community does not wait for a sure winner; it must act in the face of uncertainty, settling on plausible actions that are politically acceptable.

* * *

- 30. It is unwise for evaluation to focus on whether a project has "attained its goals."
- X 31. Goals are a necessary part of political rhetoric, but all social programs, even supposedly targeted ones, have broad aims.
- 32. Legislators who have sophisticated reasons for keeping goal statements lofty and nebulous unblushingly ask program administrators to state explicit goals.
- X 33. Unfortunately, whatever the evaluator decides to measure tends to become a primary goal of program operators.

* * *

34. Evaluators are not encouraged to ask the most trenchant questions about entrenched programs.
35. "Evaluate this program" is often a vague charge because a program or a system frequently has no clear boundaries.
- X 36. Before the evaluator can plan data collection, he must find out a great deal about the project as it exists and as it is conceived.
- X 37. A good evaluative question invites a differentiated answer instead of leaving the program plan, the delivery of the program, and the response of clients as unexamined elements within a closed black box.
- X 38. Strictly honest data collection can generate a misleading picture unless questions are framed to expose both the facts useful to partisans of the program and the facts useful to its critics.

* * *

- X 39. Before laying out a design, the evaluator should do considerable homework. Pertinent questions should be identified by examining the history of similar programs, the related social theory, and the expectations of program advocates, critics, and prospective clients.
40. Precise assessment of outcomes is sensible only after thorough pilot work has pinned down a highly appropriate form for an innovation under test.
- X 41. When a prototype program is evaluated, the full range of realizations likely to occur in practice should be observed.
42. Flexibility and diversity are preferable to the rigidity written into many evaluation contracts.

* * *

43. The evaluator who does not press for productive assignments and the freedom to carry them out takes the King's shilling for selfish reasons.
44. The evaluator's aspiration to benefit the larger community has to be reconciled—sometimes painfully—with commitments to a sponsor and to informants, with the evaluator's political convictions, and with his desire to stay in business.
45. Managers have many reasons for wishing to maintain control over evaluative information; the evaluator can respect all such reasons that fall within the sphere of management.
46. The crucial ethical problem appears to be freedom to communicate during and after the study, subject to legitimate concerns for privacy, national security, and faithfulness to contractual commitments.
47. With some hesitation, we advise the evaluator to release findings piecemeal and informally to the audiences that need them. The impotence that comes with delay may be a greater risk than the possibility that early returns will be misread.

* * *

48. Nothing makes a larger difference in the use of evaluations than the personal factor—the interest of officials in learning from the evaluation and the desire of the evaluator to get attention for what he knows.
49. Communication overload is a common fault; many an evaluation is reported with self-defeating thoroughness.
-

- X 50. Much of the most significant communication of findings is informal, and not all of it is deliberate; some of the most significant effects are indirect, affecting audiences far removed from the program under investigation.
51. An evaluation of a particular project has its greatest implications for projects that will be put in place in the future.
52. A program evaluation that gets attention is likely to affect the prevailing view of social purposes, whether or not it immediately affects the fate of the program studied.
- X 53. Advice on evaluation typically speaks of an investigation as a stand-alone study that will draw its conclusions about a program in complete isolation from other sources of information.
- X 54. It is better for an evaluative inquiry to launch a small fleet of studies than to put all its resources into a single approach.
- X 55. Much that is written on evaluation recommends some one “scientifically rigorous” plan. Evaluations should, however, take many forms, and less rigorous approaches have value in many circumstances.
- X 56. Results of a program evaluation are so dependent on the setting that replication is only a figure of speech; the evaluator is essentially an historian.

* * *

- X 57. An elegant study provides dangerously convincing evidence when it seems to answer a question that it did not in fact squarely address.
- X 58. Merit lies not in form of inquiry but in relevance of information. The context of command or accommodation, the stage of program maturity, and the closeness of the evaluator to the probable users should all affect the style of an evaluation.
- X 59. The evaluator will be wise not to declare allegiance to either a quantitative-scientific-summative methodology or a qualitative-naturalistic-descriptive methodology.
60. External validity—that is, the validity of inferences that go beyond the data—is the crux; increasing internal validity by elegant design often reduces relevance.

* * *

- X 61. Adding a control costs something in dollars, in attention, and perhaps in quality of data; a control that fortifies the study in one respect is likely to weaken it in another.
- 62. A strictly representative sample may provide less information than a sample that overrepresents exceptional cases and deliberately varies realizations.
- X 63. The symmetric, nonsequential designs familiar from laboratory research and survey research are rarely appropriate for evaluations.
- 64. Multiple indicators of outcomes reinforce one another logically as well as statistically. This is true for measures of adequacy of program implementation as well as for measures of changes in client behavior.

* * *

- 65. In project-by-project evaluation, each study analyzes a spoonful dipped from a sea of uncertainties.
- X 66. In any primary statistical investigation, analyses by independent teams should be made before the report is distributed.
- X 67. Evaluations of a program conducted in parallel by different teams can capitalize on disparate perspectives and technical skills.
- X 68. The evaluator should allocate investigative resources by considering four criteria simultaneously: prior uncertainty about a question, costs of information, anticipated information yield, and leverage of the information on subsequent thinking and action.
- 69. A particular control is warranted if it can be installed at reasonable costs and if, in the absence of that control, a positive effect could be persuasively explained away.
- 70. The importance of comparative data depends on the nature of the comparison proposed and on the stage of program maturity.

- X 71. When programs have multiple and perhaps dissimilar outcomes, comparison is invariably judgmental. No technology for comparing benefits will silence partisan discord.

* * *

72. Present institutional arrangements for evaluation make it difficult or impossible to carry on the most useful kinds of evaluation.
73. In typical federal contracting, many basic research decisions are made without consulting the evaluators who will do the work.
- X 74. The personal scientific responsibility found in ordinary research grants is lacking in contract evaluation: the "principal investigator" is a firm with interchangeable personnel.
75. Though the information from an evaluation is typically not used at a foreseeable moment to make a foreseen choice, in many evaluations a deadline set at the start of the study dominates the effort.
76. Evaluation contracts are increasing in size, but tying many strands into a single knot is rarely the best way to get useful information.
- X 77. Large-scale evaluations are not necessarily better than smaller ones.
- X 78. Major evaluations should have multiple sponsorship by agencies with different perspectives.
- X 79. Decentralizing much evaluation to the state level would be a healthy development.

* * *

80. Society will obtain the assistance that evaluations can give only when there is a strong evaluation profession, clear about its social role and the nature of its work.
81. There is a boom town excitement in the evaluation community, but in constant dollars federal funding for evaluation research has regressed in the last few years.
82. It is inconceivable that evaluators will win their battle for appropriate responsibilities if they remain unacquainted with one another, insensitive to their common interests, and fractionated intellectually.

* * *

83. For any suitably broad social problem, a "social problem study group" should be set up. It would be charged to inform itself by weighing, digesting, and interpreting what is known. It would foster needed investigations and make the policy-shaping community aware of what is and is not known.

* * *

84. Honesty and balance in program evaluation will be increased by critical review of the performance of evaluators and sponsors.
85. Oversight by peers is the most promising means of upholding professional standards and of precipitating debate about strategic and tactical issues.
86. The best safeguard against prematurely frozen standards for evaluative practice is multiple, independent sources of criticism.
87. There is need for exchanges more energetic than the typical academic discussion and more responsible than debate among partisans.
88. Reviews of evaluation should be far more frequent than at present, and reviews from diverse perspectives should appear together.

* * *

89. For the prospective evaluator, basic training at the doctoral level in a specific social science is preferable to training restricted to evaluation methods.
90. Training in evaluation is too often the stepchild of a department chiefly engaged in training academicians or providers of service.
91. Case-study seminars scrutinizing diverse evaluative studies provide a needed interdisciplinary perspective.

-
92. Internships with policy agencies that use evaluation sensitize future evaluators to the realities of evaluation use and nonuse. These realities are hard to convey in a classroom.

* * *

93. The evaluator is an educator; his success is to be judged by what others learn.
- X 94. Those who shape policy should reach decisions with their eyes open; it is the evaluator's task to illuminate the situation, not to dictate the decision.
- X 95. Scientific quality is not the principal standard; an evaluation should aim to be comprehensible, correct and complete, and credible to partisans on all sides.